Facultad de Ciencias Agronómicas - Universidad de Chile

ESTADISTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA

Una visión conceptual y aplicada

$$S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{n-1}}$$

Antonio Rustom J. | 2012

Antonio Rustom J.

ESTADÍSTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA. Una visión conceptual y aplicada.

Responsable edición: Pedro Calandra B.

Diseño de portada: Claudia Rustom S.

Compilación: Denisse Espinoza A.

Inscripción Nº: 223.022
del Registro de Propiedad Intelectual

ISBN: 978-956-19-0790-4

Departamento de Economía Agraria
Facultad de Ciencias Agronómicas
Universidad de Chile
Avda. Santa Rosa 11315, La Pintana, Santiago, Chile.

Versión digital disponible en: http://www.agren.cl/estadistica

Santiago de Chile 2012

ESTADISTICA DESCRIPTIVA, PROBABILIDAD E INFERENCIA
Una visión conceptual y aplicada

ANTONIO RUSTOM J.

REVISORES DE CONTENIDO

CLAUDIO FERNÁNDEZ L.
ALBERTO MANSILLA M.

2012

INDICE

PROLOGO

Este libro va dirigido a alumnos que estudian agronomía y es el resultado de las experiencias en mi docencia en las carreras de Ingeniería Agronómica, Ingeniería Forestal y Medicina Veterinaria principalmente en la Universidad de Chile y en la Universidad Santo Tomás, y fundamentalmente por mi labor como profesor consultor de alumnos tesistas y de mi interrelación con investigadores en aspectos metodológicos estadísticos de sus anteproyectos y proyectos.

El desarrollo de los contenidos hace mucho énfasis en lo conceptual con ejemplos y problemas orientados a las áreas mencionados. En éste, las demostraciones de teoremas o propiedades se han limitado a aquellas que cumplan con ser un reforzamiento de lo conceptual para que no sean un distractor de lo esencial que es el concepto.

El libro sigue un orden lógico, en el cual primero se hace una revisión de los elementos de estadística descriptiva que, a parte de servir sus propios fines de describir datos, permite introducir aquellos conceptos fundamentales de la estadística como son la media aritmética, la varianza, la desviación estándar y el coeficiente de variación, amén de otros, como los relacionados a los percentiles, con gran importancia estadística y cultural.

Las unidades de probabilidad cumplen con ser un respaldo para la fundamentación en el desarrollo de las unidades posteriores, principalmente de las distribuciones de probabilidad notables y comportamiento de las muestras aleatorias.

Las unidades esenciales del libro, para aquellos que manejan las nociones ya mencionadas, son las de distribución Normal, distribuciones en el muestreo de poblaciones, la estimación y pruebas de hipótesis para los parámetros: media aritmética, varianza y proporción.

El libro incluye, además, un conjunto de ejercicios y problemas propuestos, con temática orientada a las ciencias silvoagropecuarias, la mayoría de los cuales se resuelven utilizando como referencia los ejemplos desarrollados en el texto.

Con frecuencia algunos alumnos consultan por qué los problemas no incluyen las respuestas, pregunta que considero que refleja que tales alumnos todavía no se compenetran con que la estadística es una metodología al servicio de las ciencias. Así, en un problema de prueba de hipótesis, el resultado es **todo** el desarrollo bien **conceptualizado** y en un **orden lógico**. En cambio una respuesta simplista como " se acepta la hipótesis nula" o "se rechaza la hipótesis nula" carece totalmente de sentido sin el contexto previo. No es casualidad que ningún libro de estadística incluya respuesta a problemas propuestos de tal naturaleza. Sin embargo, hay problemas, especialmente de probabilidades o tamaño de muestra, en los cuales es posible dar una respuesta que resuma el desarrollo pertinente. En casos como éste se han incluido las respuestas.

Debo agradecer a todos los académicos de la facultad con los cuales me he interrelacionado y que sin saberlo han aportado a que este libro se haya escrito, al igual que a todos aquellos que aparecen en la bibliografía. Al profesor Marcos Mora quien, como director del Departamento de Economía Agraria, apoyó y gestionó para que la Facultad patrocinara su publicación.

Mi mayor muestra de gratitud y amistad al Profesor Claudio Fernández por su disposición para leer el libro y aportar con sus sugerencias para mejorar el original.

Al Profesor Alberto Mansilla, mi entrañable amigo, por su importante influencia para despertar en mí el interés por la Estadística, y en relación a este libro, por mostrarme una forma didáctica de presentación de la teoría de probabilidades.

A la Facultad de Ciencias Agronómicas por hacer posible la publicación de este libro, al Jefe de Biblioteca, Profesor Pedro Calandra, por su responsabilidad en la edición, y especialmente a Denisse Espinoza por su paciencia y dedicación para llevarla a cabo.

Principalmente mis agradecimientos a Eliana, mi esposa, cuya paciencia para soportarme sentado durante horas frente al computador, me sirvieron de estímulo para seguir adelante y concluir el texto.

Antonio Rustom J

Santiago, 2012

1. ESTADISTICA DESCRIPTIVA

1.1 Introducción.

Se postula que *"quien tiene la información tiene el poder"*. Posiblemente de ahí las grandes inversiones de los países, principalmente los desarrollados, en generar conocimientos a través de investigaciones de las más diferentes disciplinas.

Hoy en día la generación de información y su recopilación ha adquirido gran volumen y se requiere de instrumentos que sean capaces de procesarla en volumen y rapidez.

La información siempre, y con mayor razón hoy en día, es importante para la **toma de decisiones** las que deben ser oportunas y óptimas. Con mala o insuficiente información posiblemente la decisión sea mala , por muy bueno que sea el procesamiento de ésta. Por el contrario, por muy buena que sea la información si el procesamiento es malo seguramente también la decisión sea equivocada. En consecuencia, un sólido respaldo para una **acertada toma de decisiones**, contempla ambas aspectos: ***información buena y suficiente, procesamiento correcto.***

La Estadística es una disciplina que proporciona la metodología, fundada en la Matemática, para obtener, recopilar, procesar, resumir y presentar datos referentes a un estudio de interés, transformándolos en *estadísticas* con el fin de interpretarlas para obtener conclusiones, dando garantía de idoneidad en los procedimientos. También propone metodologías que permita deducir características poblacionales a partir de muestras de ella.

Actualmente la Estadística está tan difundida y sus méritos tan aceptados que prácticamente no existe actividad que no la utilice de una u otra manera, a tal punto que cualquier investigación que genere *datos* y no la utilice en la forma adecuada para su análisis, corre el riesgo que sus conclusiones no sean consideradas *científicamente válidas*. Por **dato** se entenderá un *valor* que mida en *un* individuo una característica, que puede ser una *cualidad* o una *cantidad*. Por ejemplo: color de pelo "rubio" ; calificación "regular" ; rendimiento "72 qq/ha" . Cada uno de ellos, rubio, regular, 72 es un dato.

<u>Abuso y mal uso de la estadística.</u>

A pesar de la evidente utilidad de la estadística, su **uso** se presta para **mal uso** e incluso para **abusos**, lo que ha permitido que surjan detractores que basan sus opiniones en estos últimos sin reconocer sus grandes ventajas. A continuación un par de estas opiniones:

1) Benjamín Disraeli hizo la siguiente aseveración "Existen tres tipos de mentiras, las mentiras ordinarias, las grandes mentiras y las mentiras estadísticas".

Darrel Huff en su libro *Cómo mentir con la Estadística*, anotó al respecto "los bribones ya conocen tales trucos; los hombres honrados deben aprenderlos para defenderse" (tomado del texto *Estadística para administradores* de Levin, R.. & Rubin, D.)

2) Hace años, una escritora humorística chilena, Eliana Simon, publicó en una revista nacional un aforismo que decía: "Todo se puede probar con pruebas y lo que no se prueba con pruebas, se prueba con estadísticas". Sin embargo la misma escritora escribió también "Por lo general, el que no cree en las estadísticas, creería en ellas si las entendiera" (tomado del libro *Estadística Elemental* de Horacio D'Ottone).

Es cierto, como se expresó más arriba, que personas sin escrúpulos se sirven de ella para sus propios fines cuando no tienen otros argumentos para respaldar sus posiciones. A continuación algunos ejemplos.

1) La atención hospitalaria es mala y como prueba está que el porcentaje de enfermos fallecidos en los hospitales es muy superior al porcentaje de enfermos fallecidos en sus casas.
Es obvio que el porcentaje de fallecidos sea más alto en los hospitales, independiente de la calidad de la atención.

2) El 33% de las alumnas de un curso de ingeniería se casó con profesores de la universidad.
Lo cual resulta cierto, pero no se dijo que el curso tenía solamente **tres** alumnas.

3) Según una estadística se producen más accidentes en el centro de Santiago a 35 km/h que a 65 km/h.
La razón es que en el centro la causa de los accidentes es por la congestión vehicular, causa también de la baja velocidad.

En otros casos se debe a un mal uso o interpretación de ella, como lo ilustran los siguientes ejemplos.

1) La producción industrial en el año 1963 está al mismo nivel que en 1950, ya que como se puede apreciar entre 1950 y 1958 ésta disminuyó un 30%, mientras que entre 1958 y 1963 aumentó un 30%.
La razón de esta mala conclusión está en que las bases de cálculo de ambos porcentajes es distinta. Así, si en 1950 la producción es 100, en 1958 será 70 y por tanto en 1963 será 91, es decir, 9% menor que en 1950.
2) Un diario publicaba "los compositores encuentran inconcebible que más del 100% de lo recaudado por el Departamento de Derecho de Autor se destine a pagar al personal que trabaja en el servicio, y el resto a cancelar derechos a los autores del país".
Aquí está muy expresada la idea, porque si lo recaudado es 100% *no hay resto* para cancelar a los autores.
3) Un estudio reveló una alta correlación entre el peso de un niño de básica y su rapidez de lectura, deduciéndose que los niños gordos tienen mayor rapidez de lectura que los flacos.
En este caso la alta correlación es verdadera , pero la deducción es mala, por que, en primer lugar asocian peso con "gordura", en circunstancia que el peso está altamente correlacionado a la edad y por lo tanto a la estatura. En segundo lugar, los alumnos de mayor peso están asociados a mayor edad y por lo tanto a alumnos de los últimos cursos de básica.

Uso de la Estadística.

La Estadística es **necesaria** cuando existe *variabilidad* entre los datos. Sin variabilidad en las observaciones la Estadística carece de valor. Se puede decir, entonces, que la Estadística es en general el *estudio de la variabilidad.* Dos aspectos importantes de ésta son:

1º Describir información.
Esto es válido **sólo** para el conjunto de datos descritos y se realiza mediante:
i) tablas de frecuencias y/o porcentajes
ii) gráficos

iii) medidas que resumen la información, como media o promedio, moda, mediana, desviación estándar, coeficiente de variación , etc.

De esta manera una gran cantidad de datos pueden ser mostrados en forma "resumida" y susceptibles de ser interpretados.

2º Hacer inferencias.

Corresponde a la obtención de conclusiones acerca de las características de una *población* a partir de una muestra de ésta.

1.2. Términos estadísticos básicos.

Por ***Universo*** se entenderá el conjunto de individuos objeto de nuestro interés o estudio. La especificación del universo, en general, no es trivial, pues es necesario que no haya ambigüedad respecto a quien forma parte o *no forma parte* de este conjunto.

Por ***Población*** se entenderá el **conjunto de datos** de una característica medida en cada individuo del universo. Así, asociado a un mismo universo se podrán tener varias poblaciones. Para distinguir una población de otra denominaremos ***variable*** a cada una de estas características, por ejemplo, la variable peso, la variable altura, la variable sexo, la variable estado civil, etc. En consecuencia, los *diferentes* valores que toma una característica se denomina variable.

Por ***muestra*** se entiende cualquier *subconjunto de la población*.

Existen distintas formas de elegir una muestra. Las dos más opuestas son: las muestras *dirigidas* donde la selección de los individuos de la población se efectúa al gusto del investigador ; las muestras *aleatorias*, que son las que tienen *validez estadística* y son aquellas donde los individuos son seleccionados mediante un procedimiento regido por el azar, por ejemplo, a través de *números aleatorios*.

Por **parámetro** se entenderá cualquier valor característico de una *población*, por ejemplo, el peso promedio, la altura máxima o el estado civil más frecuente. Este valor es *constante*.

Por ***estadígrafo*** o *estadístico* se entenderá un valor característico obtenido a partir de una *muestra* . Esta cantidad es *variable* , puesto que depende de la muestra, ya que de una población se puede elegir un conjunto "muy grande" de muestras cada una con un valor característico distinto.

1.3 Tipos de variables.

Para representar adecuadamente poblaciones es necesario reconocer el tipo de variable que se necesita describir. Se puede distinguir dos tipos de variables, las que a su vez se pueden subdividir en otros dos tipos.

Tipos de variables
- Cualitativas
  - Nominales
  - Ordinales
- Cuantitativas
  - Discretas
  - Continuas

Variable cualitativa, es aquella que mide una *cualidad* . ***Variable cuantitativa***, es aquella que mide una *cantidad*.

Variable nominal, es aquella cuyos valores son nombres o códigos sin una relación de orden intrínseco entre ellos. Ejemplos son: sexo ; estado civil ; nacionalidad ; religión ; raza o color de piel.

Variable ordinal, corresponde a aquella cuyos valores son nombres o códigos , pero con una relación de orden intrínseco entre ellos, es decir, sus valores conllevan un ordenamiento de mejor a peor o de mayor a menor. Por ejemplo: la calificación (excelente , bueno , regular , malo); el grado en las F.F.A.A.(General , Coronel , Capitán , ...) ; la calidad (extra , primera , segunda , ...) o nivel de infestación (sana , leve , moderada , ...).

Variable discreta, usualmente es aquella que solo toma valores enteros. Por ejemplo: número de hijos por familia ; número de elementos defectuosos en una partida de repuestos o número de insectos por hoja.

Variable continua, son las de mayor jerarquía matemática, y corresponden a aquellas que pueden asumir cualquier valor *real* dentro de un cierto rango. Por ejemplo:estatura ; peso ; edad ; rendimiento de un cultivo o el tiempo que demora un corredor en los 100 m.

1.4 Descripción de variables.

En general, cualquiera sea el tipo de la variable a resumir, existen tres formas de realizarla:
1° Por medio de **tablas de frecuencias**, que corresponde a una tabla forma da por columnas, donde en la primera columna se anotan los diferentes valores de la variable (*clases o categorías*) y en las siguientes columnas los diversos tipos de frecuencia. Por ***frecuencia absoluta*** se entiende el número de individuos que pertenece a una misma *clase*.
2° Mediante **gráficos**, que son recursos pictóricos que permi ten ilustrar mediante un dibujo *ad hoc* lo que aparece en la tabla de frecuencias. Existen diversos tipos de gráficos y el uso de cada uno depende del tipo de variable a representar.
3° Con **medidas resúmenes**, que corresponden a parámetros o *estadígrafos*, según se trate de una población o una muestra, y que sirven para mostrar posicionamiento de los datos, *medidas de posición*, o el grado de concentración de estos, *medidas de dispersión*.

Estas posibilidades de presentación de datos pueden ser elegidas en forma excluyente o complementarias, incluso las tres simultáneamente. A continuación se explicará la manera en que es posible resumir cada tipo de variable.

<u>Descripción de Variables nominales.</u>

1º Mediante tablas de frecuencia cuya estructura es la siguiente:

VALOR	**f_i**	**h_i (%)**
n_1	f_1	h_1
n_2	f_2	h_2
n_3	f_3	h_3
...	...	...
...	...	...
n_k	f_k	h_k
TOTAL	**N**	**100,0%**

donde $\mathbf{f_i}$: es la frecuencia absoluta ; **N**: tamaño de la población y la frecuencia relativa, expresada en porcentaje, $\mathbf{h_i = 100 * f_i / N}$.

En el cuadro 4.1 se muestra un ejemplo de este tipo de variable.

Raza	f_i	h_i(%)
Pastor Alemán	38	31,7
Doberman	12	10,0
Labrador	3	2,5
Pekinés	44	36,7
Poodle	23	19,1
TOTAL	120	100,0

Cuadro 4.1. Perros atendidos en una clínica Veterinaria, por raza.

2º A través de gráficos de los cuales los más conocidos y utilizados son:

Los de **barra simple** que se usan para representar tanto frecuencias absolutas , como frecuencias relativas. Se dibujan como barras rectangulares de altura proporcional a la frecuencia y todos de igual base. Las barras van separadas porque representan categorías y no valores numéricos en el eje X.

La figura 4.1 es la representación gráfica del cuadro 4.1.

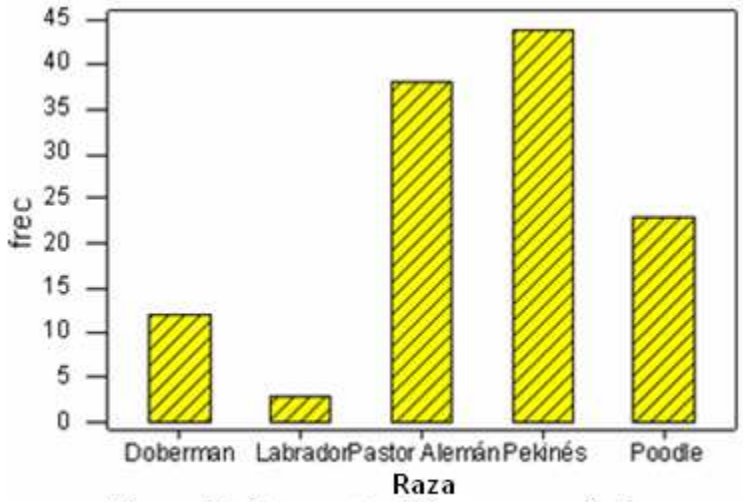


Figura 4.1. Perros atendidos en una clínica Veterinaria, por raza.

Los **circulares** son gráficos simulando una torta con porciones de diferentes tamaño, que sirven para expresar la frecuencia relativa o porcentaje de cada categoría, donde los tamaños de los sectores circulares son proporcional al porcentaje que representa cada categoría.

La figura 4.2 representa la misma información anterior en términos porcentuales.

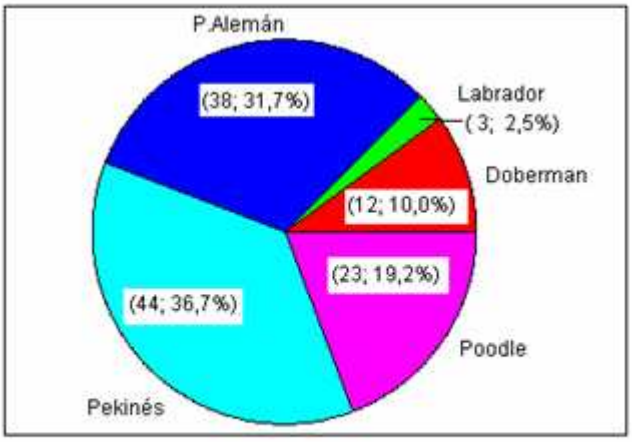


Figura 4.2. Perros atendidos en una clínica Veterinaria, por raza.

Los de **barras agrupadas** sirven para representar frecuencias absolutas o relativas, cuando existen subdivisiones dentro de cada categoría, como se ilustra en el cuadro 4.2.

Raza	f_i	h_i(%)	<1	1-2	3-4
Pastor Alemán	38	31,7	14	10	14
Doberman	12	10,0	1	7	4
Labrador	3	2,5	2	0	1
Pekinés	44	36,7	28	9	7
Poodle	23	19,1	12	8	3
TOTAL	120	100,0	57	34	29

Cuadro 4.2. Perros atendidos en una clínica Veterinaria, por raza y grupo de edad.

Por ejemplo si la clasificación de perros atendidos en la Clínica Veterinaria se subdividiera por grupos de edad, el gráfico para su representación puede ser el de barras agrupadas, como el de la figura 4.3.

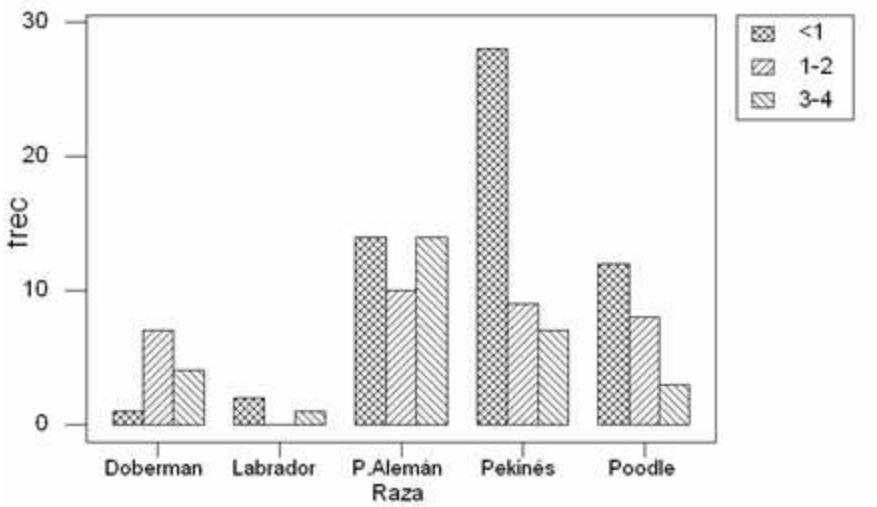


Figura 4.3. Perros atendidos en una clínica Veterinaria, clasificados por raza y edad.

Los gráficos de **barras compuestas o subdivididas** en los cuales cada barra corresponde al 100% de una clase y cada subdivisión es proporcional al porcentaje que representa cada subcategoría.

La misma información de la figura 4.3 se presenta en forma de barras subdivididas en la figura 4.4.

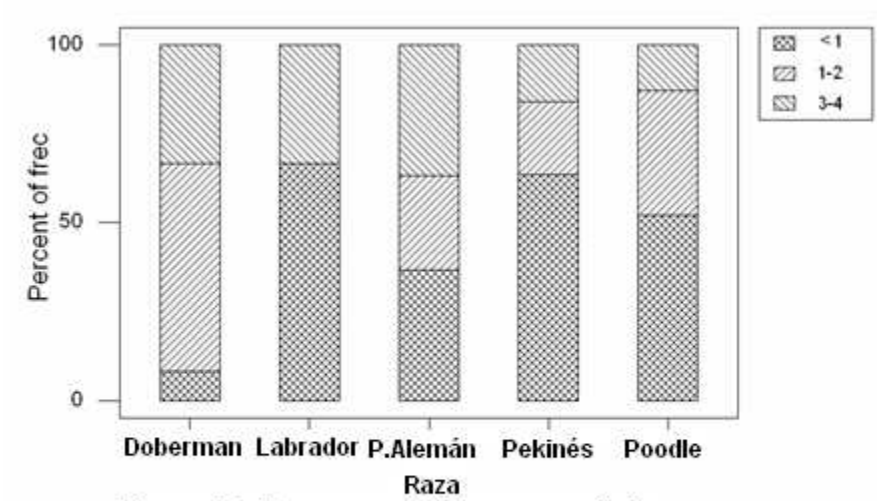


Figura 4.4. Perros atendidos en una clínica Veterinaria, clasificados por raza y edad.

Nótese que cada barra tiene la misma altura, independiente de la frecuencia que ella represente, pues cada barra muestra el particionamiento de cada categoría. Este tipo de gráfico no es de utilidad cuando el número de subdivisiones es mayor a 4, ya que la comparación entre las categorias se hace más confusa.

Los **gráficos de línea** casi siempre están vinculados a la variable *tiempo*, asociada al eje de abscisas. Como su nombre lo indica estos se forman al unir los diferentes puntos en el tiempo

por medio de segmentos rectilíneos. Tienen la ventaja de permitir la superposición en paralelo de dos o más líneas lo que facilita la comparación de otros fenómenos asociados al mismo período. En la realidad es una representación de una variable continua como el tiempo.

Un ejemplo se muestra en la figura 4.5 donde se representa la evolución del Indice Bursátil Agroindustrial en los años 2004, 2005 y 2006.

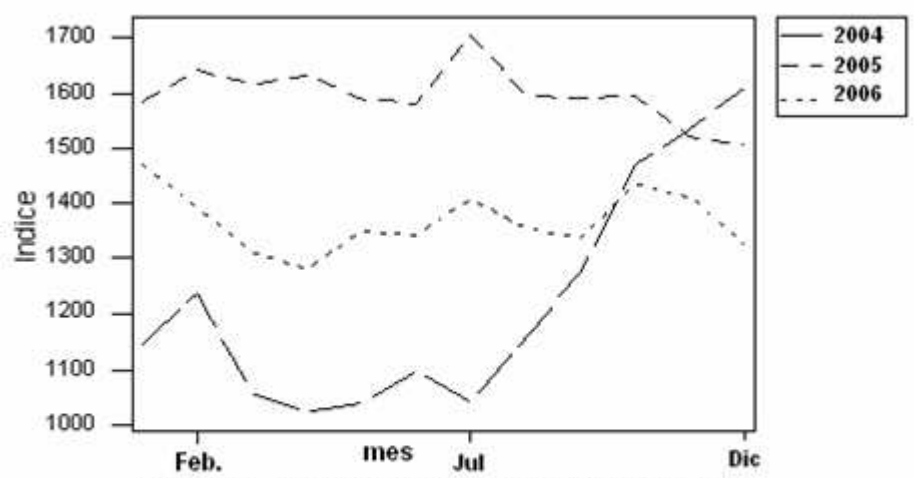


Figura 4.5. Variacion mensual del Indice Bursatil Agroindustrial, años 2004,2005 y 2006.

Los **pictogramas** son dibujos cuyas figuras se relacionan al fenómeno que se está representando, por ejemplo, "barriles" para representar producción de petróleo ; "vacas" para representar masa ganadera o "personas" para representar poblaciones. Son de poco valor académico, porque están orientados a la divulgación.

3º Utilizando medidas resúmenes, que en el caso de las variables nominales la única posible es la *moda*. Se llama ***Moda (Mo)***, al valor de la variable que tiene mayor frecuencia, o sea, el valor que más se repite en la población o muestra.

Según el ejemplo del cuadro 4.1 la moda es Pekinés, *Mo = Pekinés*, pues de las razas atendidas fue la más frecuente con 44 ejemplares.

Descripción de variables ordinales.

En general utiliza el mismo tipo de tablas de frecuencia y de gráficos que el tipo anterior, la diferencia radica en que los valores llevan un ordenamiento tanto en la tabla de frecuencia como en el gráfico.

Como medidas resúmenes, para este tipo de variables, además de la *moda* se puede utilizar la *mediana.* Se llama ***Mediana (Me)*** o ***valor mediano*** , al valor de la variable que ocupa la **posición central** o las **dos posiciones centrales** de los datos **ordenados**. Así la mediana es un valor o dos valores que separa a los datos ordenados en dos grupos con igual número de observaciones, uno con valores *mayores o iguales* a la *mediana* y el otro con valores *menores o iguales* a la *mediana.*

Ejemplos 4.1

a) En una evaluación por nivel de daño por pudrición en racimos de uva estos se calificaron como *sano (S), leve (L), moderado (M) y grave (G).* Esta es una escala ordinal, porque sano es el menor nivel de daño y grave el mayor. En la inspección de 7 racimos se determinaron los siguientes niveles para cada uno: S , S , L , M , G , L , S. Para encontrar la mediana es necesario ordenar los datos en uno de los dos sentidos, sea: S S S L L M G. El valor que ocupa la posición central es L que se ubica en el cuarto lugar, por lo tanto *Me = leve*. Nótese que a la izquierda hay 3 valores S, menores a L, y a la derecha hay 3 valores, una L igual a la mediana y los otros M y G mayores a la mediana L. En este mismo ejemplo la moda es S.

b) Si en la misma situación anterior el número de racimos evaluados fuera un número par, entonces, resultarían dos valores medianos, iguales o distintos. Por ejemplo en 10 racimos los niveles, ya ordenados, resultaron: S S S S S L L L M G. Los dos valores que ocupan las posiciones centrales, 5ª y 6ª ubicación, son S y L respectivamente, por lo tanto una mediana es S y la otra es L. A la izquierda de S hay 4 valores iguales a S y a la derecha de L hay 4 valores, dos iguales a L y otros dos mayores.

Descripción de Variables cuantitativas para datos no agrupados.

Si el número de datos, N , no es grande estos, pueden ser tratados en forma individual como cantidades X_1 , X_2 , X_3 , ..., X_N. En esta situación no se tabulan y tampoco es posible mostrarlos en un gráfico, pero si se pueden resumir en términos de dos tipos de medidas: medidas de *posición* y medidas de *dispersión.*

Las *medidas de posición* de tendencia central , cumplen el propósito de indicar el valor alrededor del cual se distribuyen los datos, es decir, una especie de *centro de gravedad* de estos. En general se pretende informar del orden de magnitud de los datos. Algo equivalente a decir, por ejemplo, "los honorarios son del orden de los $ 20.000 diarios". Existen, también, otros tipos de medidas de posición que no son de tendencia central y que se presentarán posteriormente.

Las *medidas de dispersión*, tienen por finalidad cuantificar la *variabilidad* de los datos, es decir, que tan separados o disímiles son uno de otro. Se puede decir que es una medida del "grado de concentración o de densidad" de los datos en torno a su centro de gravedad.

Medidas de posición de tendencia central.

Entre las medidas de posición más relevantes se mencionan la *Moda* y la *Mediana* , definidas anteriormente, y la *Media aritmética* que es la más importante de todas para variables cuantitativas, debido a su amplia utilización, a sus propiedades matemáticas y a su vinculación a la *distribución normal*.

La *moda* es importante, principalmente, en variables cualitativas o cuando el interés es la mayoría. La *mediana*, también es más importante para variables cualitativas ordinales y en ciertas situaciones especiales de variables cuantitativas.

La ***media aritmética*** , designada y definida como $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$, tiene un uso muy difundido y conlleva una serie de propiedades muy importantes.

A continuación se listan una serie de propiedades de la media aritmética, denominada comúnmente *promedio*, y ejemplos ilustrativos de ellas.

P_1: $\sum_{i=1}^{N} X_i = N*\mu$, esta propiedad es una consecuencia directa de la definición.

P_2: $\sum_{i=1}^{N}(X_i - \mu) = 0$, que establece que la suma de los desvíos, respecto a la media, de un conjunto N de datos es *siempre igual a cero*. Se llama desvío a la diferencia $(X_i - \mu)$ e indica cuantas unidades está el valor X_i por sobre o por bajo la media del grupo, dependiendo si es positiva o negativa respectivamente.

P_3: $Y_i = X_i + k \Rightarrow \mu_Y = \mu_X + k$, esta propiedad dice que si a cada uno de los datos de un grupo se le suma una cantidad constante k, entonces, el promedio de los nuevos datos es igual al promedio original aumentado en la cantidad k.

P_4: $Y_i = k*X_i \Rightarrow \mu_Y = k*\mu_X$, es decir, si cada dato de un conjunto es amplificado por una constante k, entonces el nuevo promedio es k veces el promedio original.

P_5 : $Y_i = c*X_i + k \Rightarrow \mu_Y = c*\mu_X + k$, es la expresión de las propiedades 3 y 4 en forma combinada.

P_6 : La media de una constante es la constante, propiedad bastante trivial e intuitiva.

Ejemplos 4.2

a) Si el ingreso per cápita de una familia compuesta por 5 personas es de $ 75.000, entonces, el ingreso familiar es de $ 375.000, independiente del ingreso de cada uno.

b) Si la edad promedio de un grupo familiar es actualmente 38 años, entonces la edad promedio de este mismo grupo familiar en 14 años más será de 52 años.

c) Si en la arveja el peso de su vaina vacía es siempre igual al peso de los granos que contiene, entonces, el peso promedio de las vainas completas es el doble del peso promedio de su contenido.

d) En una empresa donde el sueldo promedio de sus empleados es de $ 220.000, el sindicato logra un reajuste de sueldos del 12% más una asignación fija de $ 20.000 por trabajador. Entonces, el sueldo promedio reajustado en la empresa será igual a :
220.000 más 12% de 220.000 más 20.000, o sea, $1{,}12*220.000 + 20.000$, es decir, de $ 266.400.

Observaciones.

1) Cuando los datos están "bien distribuidos" la media aritmética y la mediana tienen valores muy parecidos, por lo cual se puede utilizar cualquiera de las dos como medida de posición, pero debe preferirse la media aritmética por ser más familiar para la mayoría de las personas y por tener más propiedades vinculantes a otras medidas y a la distribución normal.
2) La media aritmética, sin embargo, es muy sensible a valores extremos y por lo tanto su valor deja de ser "representativo" del conjunto de datos. En casos como estos se puede utilizar la mediana o la media calculada excluyendo los datos extremos, haciendo la aclaración correspondiente.

Medidas de dispersión.

Estas tienen por objetivo dar una cuantificación de la heterogeneidad de los datos, es decir, dar una medida de qué tan parecido o disímiles son los datos de una población entre si.

El **Rango** es una manera sencilla de hacerlo midiendo cuán repartidos están los datos y se define por $\boldsymbol{R = X_{max} - X_{min}}$. Para calcular el rango es necesario, por tanto, identificar los valores extremos de los datos. Su desventaja es que al considerar sólo los valores extremos y no los datos restantes resulta una medida poco eficiente.

La **Varianza**, $\boldsymbol{\sigma^2}$, es otra forma de medir la variabilidad de los datos. Su construcción se realiza sobre la base de los desvíos respecto a la media aritmética y cuya definición es $\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$. Se puede demostrar que $\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu^2}$, la que resulta ser una forma más práctica para su cálculo. La varianza es una medida que se complementa muy bien con la media aritmética, en especial cuando se asocian a la distribución normal. Sin embargo la varianza tiene el gran inconveniente que sus unidades de medida están al cuadrado, por lo que no tiene interpretación en la realidad, por ejemplo sus unidades pueden ser "kg al cuadrado" o "años al cuadrado". Este inconveniente se subsana con la **Desviación Estándar o Desviación típica,** $\boldsymbol{\sigma}$, que se define como la raíz cuadrada de la varianza, cuya expresión es

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \mu^2}}$$

Las propiedades más importantes de la varianza y la desviación típica se explican a continuación.

P$_1$: $\boldsymbol{Y_i = X_i + k \Rightarrow \sigma_Y^2 = \sigma_X^2}$ y $\boldsymbol{\sigma_Y = \sigma_X}$, que establece que la varianza y la desviación estándar no se altera al sumar una constante a los datos.

P$_2$: $\boldsymbol{Y_i = k * X_i \Rightarrow \sigma_Y^2 = k^2 * \sigma_X^2}$ y $\boldsymbol{\sigma_Y = k * \sigma_X}$, que especifica que al multiplicar los datos por una constante, la varianza queda amplificada por la constante al cuadrado y la desviación estándar sólo por la constante.

P$_3$: $\boldsymbol{X_i = k \Rightarrow \sigma_X^2 = \sigma_X = 0}$, es decir, que la variabilidad de una constante es cero.

Ejemplo 4.3

Se mostrarán, numéricamente, las propiedades de la media y la varianza utilizando los datos de la siguiente tabla.

X_i	$Y_i = X_i + 4$	$V_i = 3*X_i$
5	9	15
8	12	24
12	16	36
20	24	60
22	26	66

$\sum X_i = 67$; $\sum X_i^2 = 1117$ $\Rightarrow$ $\mu_X = 13,4$; $\sigma_X^2 = 43,84$.

$\sum Y_i = 87$; $\sum Y_i^2 = 1733$ $\Rightarrow$ $\mu_Y = \mu_X + 4 = 17,4$; $\sigma_Y^2 = \sigma_X^2 = 43,84$.

$\sum V_i = 201$; $\sum V_i^2 = 10053$ $\Rightarrow$ $\mu_V = 3*\mu_X = 40,2$; $\sigma_V^2 = 9*\sigma_X^2 = 394,56$.

Observación.

Una notación utilizada universalmente consiste en resumir una información cuantitativa en la forma $\mu \pm \sigma$.

Medida de dispersión relativa.

Establecer la homegeneidad o heterogeneidad de los datos de una población mediante la desviación típica o la varianza, requiere conocimiento y principalmente experiencia del fenómeno en estudio para una correcta interpretación de ésta. Una medida útil porque mide la dispersión en forma relativa es el ***Coeficiente de Variación***, que permite una interpretación más objetiva de la variabilidad, definida por $CV = [\frac{\sigma}{\mu} * 100]\%$. Con la dispersión relativa es posible establecer rangos que determinen niveles de variabilidad poblacional de homogeneidad o heterogeneidad, así por ejemplo CV menores al 5% indican , por lo general , gran homogeneidad, CV de alrededor del 20% corresponden por lo general a una homogeneidad moderada , mientras que *CV* mayores al 50% indican gran heterogeneidad. Puede alcanzar, incluso porcentajes muy superiores a 100%.

Ejemplo 4.4

Se expresa que en una lechería A la producción por vaca es 15 ± 2 , entonces se entiende que la producción promedio por vaca es 15 litros, con una desviación estándar de 2 litros y un $CV = 13,3\%$.

Si en otra lechería B la producción por vaca es $14 \pm 0,5$, entonces en ésta la producción promedio por vaca es de 14 litros con una desviación estándar de 0,5 litros y un $CV = 3,6\%$.

En consecuencia, la producción en la lechería B es más ***homogénea*** que en la lechería A.

En una distribución normal o gaussiana, se establece, como se justificará cuando se estudie esta distribución, que aproximadamente el 68% de los individuos tienen valores en el rango dado por $\mu - \sigma$ y $\mu + \sigma$. Por experiencia se sabe que la producción sigue un comportamiento normal, luego en el caso de la lechería A se puede deducir que el 68% de las vacas se esperaría que tengan una producción entre 13 y 17 litros, mientras que en la lechería B se esperaría una producción entre 13,5 y 14,5 litros para el 68% de las vacas. Con esta otra presentación, también se evidencia que la producción en la lechería B es más homogénea que en la lechería A.

Descripción de Variables cuantitativas discretas para datos agrupados.

En este caso las tablas son similares a las de variables cualitativas, pero pueden incluir, además, frecuencias acumuladas: En la primera columna, ahora se indican los diferentes valores X_i que asume la variable en estudio y en las siguientes columnas la frecuencia $\mathbf{f_i}$ que representa las veces que se repite el valor X_i , la frecuencia acumulada $\mathbf{F_i}$ es la **suma parcial** de las f_i , por ejemplo $F_3 = f_1+f_2+f_3$, $F_i = f_1+f_2+f_3+f_4+...+f_i$, y $\mathbf{H_i}$ es la **expresión porcentual** de $\mathbf{F_i}$ o si se prefiere es la suma parcial de las h_i, como lo muestra el ejemplo del cuadro 4.3.

n°lesiones/hoja (X_i)	f_i	h_i(%)	F_i	H_i(%)
0	128	32,0	128	32,0
1	100	25,0	228	57,0
2	52	13,0	280	70,0
3	20	5,0	300	75,0
4	40	10,0	340	85,0
5	60	15,0	400	100,0
Total	400	100,0		

Cuadro 4.3. Número de lesiones causadas por virus en 400 hojas de tabaco.

En la tabla , la frecuencia 52 corresponde al número de hojas que presentaron 2 lesiones, cuyo valor porcentual es 13,0%; la frecuencia acumulada 300 indica que existen 300 hojas con 3 o menos lesiones y el 57% de la última columna dice que en el 57% de las hojas se encontró a lo más una lesión.

El gráfico a utilizar para representar estos datos se denomina **gráfico de varas** que consiste en ubicar sobre el eje horizontal X los valores X_i y trazar sobre este valor una línea perpendicular, vara, de altura proporcional a la frecuencia.

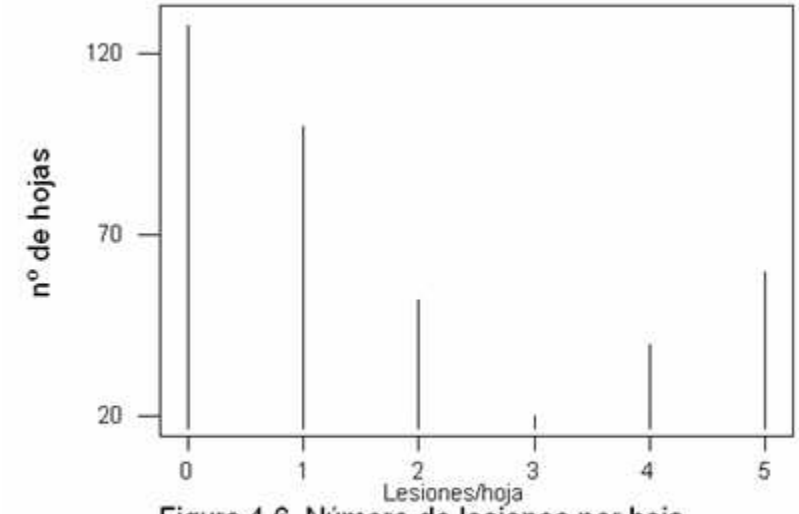


Figura 4.6. Número de lesiones por hoja causadas por virus en 400 hojas de tabaco.

Las medidas de posición, al igual que antes,incluye a la Moda que es el valor X_i de mayor frecuencia, la Mediana, ya definida anteriormente, que ahora se determina como el valor X_i tal que $H_i \geq 50\%$ y $H_{i-1} < 50\%$, es decir, "el valor en el cual se supera por **primera vez el 50%"** .

La Media aritmética se calcula utilizando la frecuencia f_i , ya que este número indica las veces que se repite el valor X_i , como lo indica la siguiente expresión $\mu = \frac{\sum_{i=1}^{I} f_i X_i}{N}$.

Entre las medidas de dispersión, la **Varianza** se obtiene igualmente que la media, ponderando los desvíos de los datos por la frecuencia f_i . Su expresión es $\sigma^2 = \frac{\sum_{i=1}^{I} f_i (X_i - \mu)^2}{N}$ y su fórmula práctica de cálculo es $\sigma^2 = \frac{\sum_{i=1}^{I} f_i X_i^2}{N} - \mu^2$. La desviación típica es por definición la raíz positiva de la varianza y el CV la razón porcentual entre la desviación típica y la media.

Ejemplo 4.5

Con los datos del cuadro 4.3, se obtiene que la Moda es 0, que la Mediana es 1 y que μ = ($0*128 + 1*100 + 2*52 + ... + 5*60)/400 = 1,81$ lesiones/hoja.
Observe que este promedio no es un valor entero, pero igual tiene interpretación y es una forma útil para comparar situaciones. Hay que comprender que el promedio es un valor referencial, de mucha utilidad, pero no necesariamente debe coincidir con algún valor observado. Es posible leer que un futbolista M es más goleador que otro P, porque M tiene un promedio de goles por partido de 1,6, mientras que el promedio de goles de P es de 1,2.
Para los mismos datos la varianza se calcula σ^2= ($0^2*128 + ... + 5^2*60)/400 - (1,81)^2$, lo que da $3,2939$, por lo tanto σ = $\sqrt{3,2939}$ = 1,8149 y $CV = 100,3\%$.

Descripción de variables continuas para datos agrupados.

Si la variable es **continua** los datos se clasifican en clases que son intervalos, denominándose **tabla de frecuencias de intervalos**.
La frecuencia **f_i** representa ahora el **número de datos comprendido en el intervalo** y el resto de la tabla se confecciona en la misma forma que en la tabulación de variables discretas, pero incluyendo, además, una columna con el valor marca de clase X_i. La tabla adquiere la estructura que se muestra a continuación.

Intervalo	**X_i**	**f_i**	**h_i(%)**	**F_i**	**H_i(%)**
$L_0 \leq X < L_1$	X_1	f_1	h_1	F_1	H_1
$L_1 \leq X < L_2$	X_2	f_2	h_2	F_2	H_2
$L_2 \leq X < L_3$	X_3	f_3	h_3	F_3	H_3
.....	...	...	.....	....	....
$L_{i-1} \leq X < L_i$	X_i	f_i	h_i	F_i	H_i
.....	...	...	.....	....	.....
$L_{k-1} \leq X \leq L_k$	X_k	f_k	h_k	N	100,0
Total		**N**	**100,0**		

donde: **L_{i-1}** e **L_i** : son los límites inferior y superior respectivamente del intervalo i-ésimo; **$X_i = \frac{L_{i-1} + L_i}{2}$** , recibe el nombre de valor clase del intervalo "i" , cuyo supuesto es que representa al promedio de los datos incluidos en el intervalo, lo que no necesariamente ocurre así y **$c_i = L_i - L_{i-1}$** , recibe el nombre de *amplitud* del intervalo "i", amplitud que puede ser distinta para cada intervalo. Por lo general, intervalos de igual amplitud facilita los cálculos.

Los gráficos utilizados en variables continuas son ***Histogramas*** y ***Polígonos de frecuencias***

La tabla corresponde a la distribución de la producción de 500 manzanos enanos

Producción(kg/árbol)	Frecuencia
$60 \leq X < 75$	45
$75 \leq X < 90$	60
$90 \leq X < 105$	70
$105 \leq X < 120$	110
$120 \leq X < 135$	90
$135 \leq X < 150$	70
$150 \leq X \leq 165$	55
TOTAL	**500**

Cuadro 4.3 Producción en kg de 500 manzanos enanos.

El **histograma y polígono de frecuencias no acumuladas** se muestra en la figura 4.7.

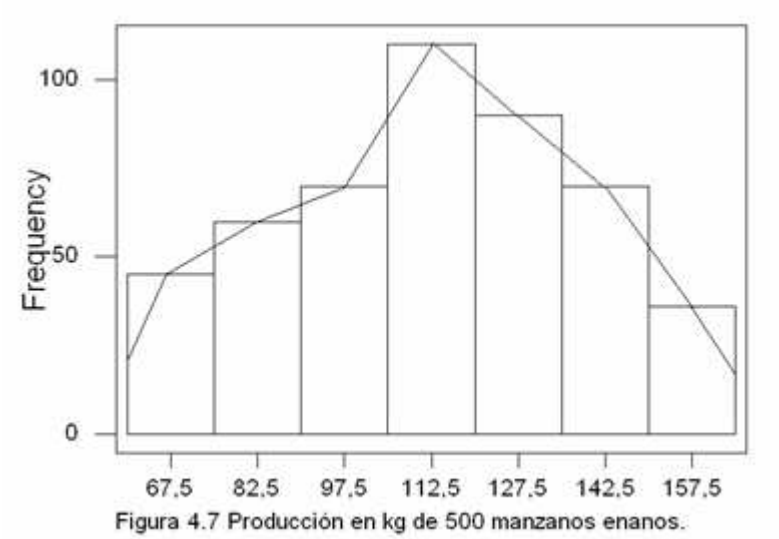


Figura 4.7 Producción en kg de 500 manzanos enanos.

La figura 4.8 ilustra la información anterior mediante un **histograma y polígono de frecuencia acumulada**.

Los histogramas de frecuencias acumuladas tienen altura $\mathbf{F_i}$ o $\mathbf{H_i}$. Los polígonos de frecuencias acumuladas *unen* los rectángulos en diagonal, empezando en **0** y terminando en **N** o **1** (100%), según sea el caso, *tendiendo* a la forma de la curva llamada *ojiva*.

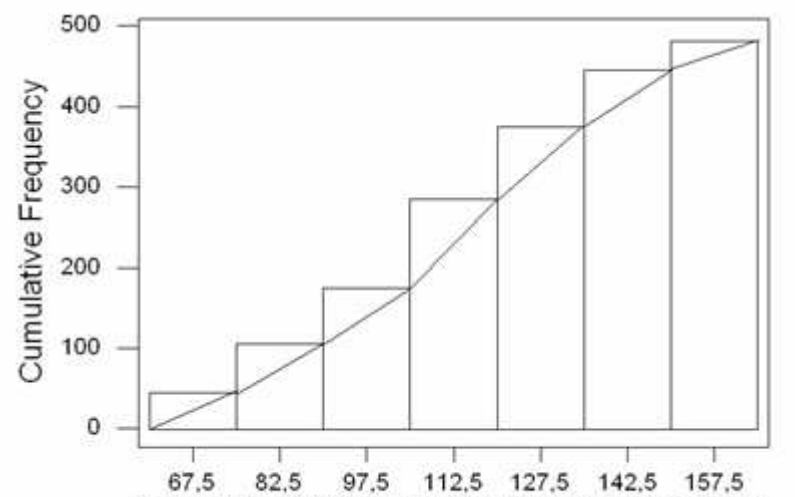


Figura 4.8 Distribución acumulativa de la producción de 500 manzanos enanos

En cuanto a las medidas resúmenes en este caso se da una gran variedad , las que se agrupan en *medidas de posición* , como son la media aritmética , la mediana , la moda (aunque esta última no tiene un gran sentido práctico) , las cuartilas , percentilas etc. ; y *medidas de dispersión* , como son la amplitud , la desviación típica , el coeficiente de variación, etc.

La *media aritmética* se calcula considerando la frecuencia f_i , pero como en este caso la frecuencia no representa a un **único valor** , sino a un intervalo , debe utilizarse para este cálculo el **valor clase** X_i , quedando la fórmula en forma similar a la de variable discreta: $\mu = \frac{\sum_{i=1}^{k} f_i X_i}{N}$.

La *Varianza* , σ^2, se obtiene , por la misma razón que la media , ponderando los **desvíos de los valores clase** X_i respecto a la media aritmética por la frecuencia f_i , quedando su expresión en la forma:

$\sigma^2 = \frac{\sum_{i=1}^{I} f_i (X_i - \mu)^2}{N}$ y su fórmula práctica de cálculo es $\sigma^2 = \frac{\sum_{i=1}^{I} f_i X_i^2}{N} - \mu^2$.

La *desviación típica* , σ , es como antes la raíz positiva de la varianza y el CV la razón porcentual entre la desviación típica y la media.

<u>Otras medidas de posición</u>.

Las cuartilas , quintilas , decilas y percentilas son otro tipo de medidas de posición , siendo la percentila la que involucra a todas las otras, incluyendo a la mediana.

Existen 99 percentilas: P_1 a P_{99} y corresponden a valores dentro del rango de los datos, de modo que **entre dos percentilas sucesivas , P_i y P_{i+1} , *siempre* queda comprendido *el 1% de los datos.***

Así , por ejemplo , entre la percentila P_{35} y la percentila P_{58} se encuentra un 23% de las observaciones , puesto que entre ellas existen (58 - 35) percentilas sucesivas.

Se llama *intervalo percentil k* al intervalo "i" tal que $H_i \geq$ k% y $H_{i-1} \leq$ k% o en palabras "el valor en el cual se supera por **primera vez el k%**" acumulado de las observaciones.

La fórmula para determinar la percentila k , está dada por: $\mathbf{P_k = L_{i-1} + \frac{\frac{k*N}{100} - F_{i-1}}{f_i} * c_i}$,donde

$\mathbf{L_{i-1}}$: límite inferior del intervalo percentil k
$\mathbf{F_{i-1}}$: frecuencia acumulada hasta el intervalo anterior al percentil k
$\mathbf{f_i}$: frecuencia del intervalo percentil k
$\mathbf{c_i}$: amplitud del intervalo percentil k

El percentil k , se debe interpretar en el sentido que el ***k% de las observaciones*** es menor a P_k y el otro ***(100 - k)% de observaciones*** tiene valores mayores.

La figura 4.9 explica como se determina la percentila k.

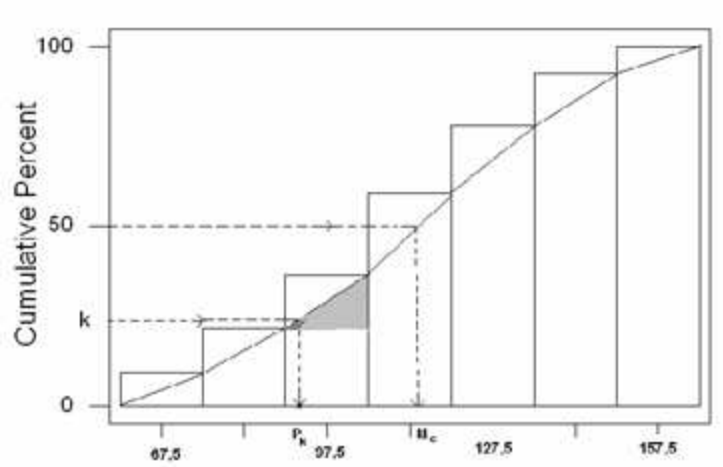


Figura 4.9. Gráfico ilustrativo determinación de la percentila k y de la mediana.

La figura muestra que el punto donde el porcentaje k, en el eje del porcentaje acumulado, intersecta al polígono de frecuencia acumulada determina en el eje de abscisa el valor P_k el que se calcula mediante interpolación o por proporcionalidad en triángulos rectángulos, área sombreada pequeña versus área sombreada mayor, lo que origina la fórmula dada. La figura, también muestra el caso de la mediana, cuya explicación es similar a la dada.

El cuadro siguiente muestra las relaciones de cuartilas , quintilas y decilas con las percentilas:

Cuartilas	Quintilas	Decilas
$Q_1 = P_{25}$	$C_1 = P_{20}$	$D_1 = P_{10}$
$Q_2 = P_{50}$	$C_2 = P_{40}$	$D_2 = P_{20}$
$Q_3 = P_{75}$	$C_3 = P_{60}$	$D_3 = P_{30}$
	$C_4 = P_{80}$	$D_4 = P_{40}$
		$D_5 = P_{50}$
		$D_6 = P_{60}$
		$D_7 = P_{70}$
		$D_8 = P_{80}$
		$D_9 = P_{90}$

Observe que de acuerdo a las relaciones anteriores y a la definición de mediana se deducen las siguientes equivalencias: $\boldsymbol{Me = Q_2 = D_5 = P_{50}}$.

Ejemplo 4.6

Se utilizarán los datos del cuadro 4.3, para lo cual será necesario completar la tabla en la forma siguiente

Producción(kg/árbol)	X_i	f_i	h_i(%)	F_i	H_i(%)
$60 \leq X < 75$	67,5	45	9,0	45	9,0
$75 \leq X < 90$	82,5	60	12,0	105	21,0
$90 \leq X < 105$	97,5	70	14,0	175	35,0
$105 \leq X < 120$	112,5	110	22,0	285	57,0
$120 \leq X < 135$	127,5	90	18,0	375	75,0
$135 \leq X < 150$	142,5	70	14,0	445	89,0
$150 \leq X \leq 165$	157,5	55	11,0	500	100,0
TOTAL		500	100,0		

Para caracterizar la información de la tabla las mejores medidas son la media aritmética y la desviación típica las que resultan de los siguientes cálculos.

$\mu = \frac{45*67,5 + ...+55*157,5}{500} = \frac{57300}{500} = 114,6$ kg ; $\sigma^2 = \frac{45*(67,5)^2 + ...+55*(157,5)^2}{500} - (114,6)^2 = 706,59$

$\sigma = \sqrt{706,59} = 26,58$ kg y Coeficiente de Variación C.V = 26,58/114,6 = 23,2%. Luego la variabilidad relativa de la producción de los árboles es de 23,2%, que se puede interpretar como una producción homogénea. La mediana, $Me = P_{50} = 105 + \frac{\frac{50}{100}*500 - 175}{110} * 15 = 115,2$ kg, es sólo un complemento a la información anterior y su interpretación es que el 50% de los árboles tienen una producción menor a 115,2 kg y el otro 50% una producción mayor a ese valor.

Otra información relevante se obtiene con aplicación de los percentiles, como por ejemplo si interesa saber el valor del percentil 82, $P_{82} = 135 + \frac{\frac{82}{100}*500 - 375}{70} * 15 = 142,5$ kg y su interpretación es que el 82% de los árboles produce menos de 142,5 kg y el otro 18% produce más de 142,5 kg.

Determinar qué porcentaje de los árboles tienen una producción menor a 100 kg. se realiza aplicando el concepto de percentil, $100 = 90 + \frac{\frac{k}{100}*500 - 105}{70} * 15$ de donde se despeja k = 30,3%. La respuesta es que el 30,3% de los árboles produce menos de 100 kg.

El mismo procedimiento se utiliza para saber cuántos árboles tienen una producción mayor a 130 kg , $130 = 120 + \frac{\frac{k}{100}*500 - 285}{90} * 15$, que da un valor para k de 69 %. Luego el 69% de los árboles produce menos de 130 kg y por lo tanto el 31% de 500 , igual a 155 árboles, tienen una producción mayor a los 130 kg.

Si se establece que el 20% de los árboles de menor producción serán sometidos a una poda especial, se necesita establecer cuál será la producción máxima de los árboles sometidos a esta poda. Esto requiere calcular el percentil 20, $P_{20} = 75 + \frac{\frac{20}{100}*500 - 45}{60} * 15 = 88,75$ kg, y por lo tanto deben ser seleccionados todos los árboles que tenga producción menor a 88,75 kg.

1.5 Otros tipos de gráficos.

En forma más reciente han surgido otras formas gráficas para representar información cuantitativa. Dos de ellos, de bastante interés, son el *diagrama de tallo y hoja* (Stem-and-Leaf) y el *diagrama de caja* (Boxplot).

Diagrama de tallo y hoja.

Una forma muy adecuada de organizar un número moderado de datos individuales consiste en dividir cada dato en dos parte, su *tallo* y su *hoja*. Si por ejemplo el conjunto de datos son números de dos dígitos, ya sea decenas y unidades o entero y decimal, entonces las decenas o el entero es el **tallo** y las unidades o el decimal es la **hoja.**

Ejemplo 5.1

Los valores 42; 32; 13; 18; 23; 44; 41;18; 15; 25; 35; 28; 17; 28; 42; 51; 50; 21; 27; 36 corresponden a las altura de 20 plantas regeneradas de coigüe medidas en una cuadrícula en un bosque nativo y cuya representación en un diagrama de tallo y hoja queda como sigue.

```
Stem-and-leaf of Altura    N  = 20
Leaf Unit = 1,0

    5     1 35788
   (6)    2 135788
    9     3 256
    6     4 1224
    2     5 01
```

El diagrama del ejemplo se obtuvo digitando los 20 datos en una columna con la siguiente secuencia de comandos:

Graph → *Steam-and-Leaf* (opcional *Trim outliers*) → Increment = 10, porque los datos corresponden a decenas.

En el cuadro la columna del centro, el *tallo*, indica la cifra de las decenas, y los de la derecha, las *hojas,* indica la cifra de las unidades. En la columna de la izquierda el ( ) indica la "moda"de las hojas y los números hacia arriba y abajo es el número de datos acumulados alrededor de la "moda". En este ejemplo la moda es (6) que indica que existen 6 valores entre 20 y 29. La primera fila indica que los valores entre 10 y 19 son 13 15 17 18 18; el 5 indica el número de datos acumulado hasta la moda. En la tercera fila el tallo es 3 que corresponde a los datos 32 35 36; el 9 indica cuantos datos hay acumulado desde abajo hasta la moda.

Diagrama de caja.

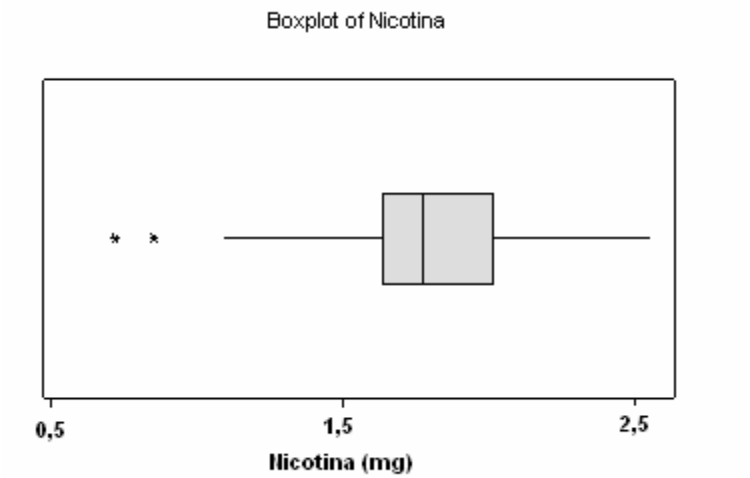
Se usa para graficar algunos estadísticos de orden y dispersión que describen un conjunto de datos. Consiste en dibujar en eje horizontal (o vertical) un segmento de línea que va del dato menor al mayor (*Rango de los datos*). Entre ellos se dibujan dos rectángulos adyacentes (caja) que empieza en el valor Q_1*, le sigue* una línea que indica la *mediana* (Q_2) de los datos y termina en el valor Q_3 . La longitud de la caja (Q_3 - Q_1*)* se llama *rango intercuartil* y es otra medida de dispersión de los datos.

Otra forma de este diagrama, lo que depende del programa estadístico utilizado, indican los valores que se alejan más de lo "razonable" de la masa de datos (Outliers), que pueden servir como diagnóstico de situaciones irregulares o anormales de los datos. MINITAB utiliza como

criterio un segmento de línea (bigote) cuyo límite inferior es $Q_1 - 1,5*(Q_3 - Q_1)$ y como límite superior $Q_3 + 1,5*(Q_3 - Q_1)$; los valores fuera de este rango; outliers, los indica con " * ".

Ejemplos 5.2

a) El gráfico corresponde a 40 datos de contenido de **nicotina en cigarrillos** cuyos estadísticos son: Min = 0,72 ; Max = 2,55 ; Q_1 = 1,63 ; Q_2 = 1,770 ; Q_3 = 2,02 ; μ = 1,774.



Los límites del segmento de línea son: $1,63 - 1,5*(2,02 - 1,63)$ y $2,02 + 1,5*(2,02 - 1,63)$, es decir, 1,05 y 2,61. Los asteriscos indican los dos valores inusuales, "outliers", que corresponden al valor mínimo 0,72 y al valor que le sigue 0,85. Los 38 valores restantes quedan comprendidos entre los los límites 1,05 y 2,61.

Los estadísticos y el gráfico del ejemplo se obtuvo digitando los datos de nicotina en una columna de la planilla de MINITAB y la siguiente secuencia de comandos:

Stat $\rightarrow$ *BasicStatistics* $\rightarrow$ *DisplayDescriptiveStatistics*
$\rightarrow$ *Graphs* $\rightarrow$ *Boxplot of data.*

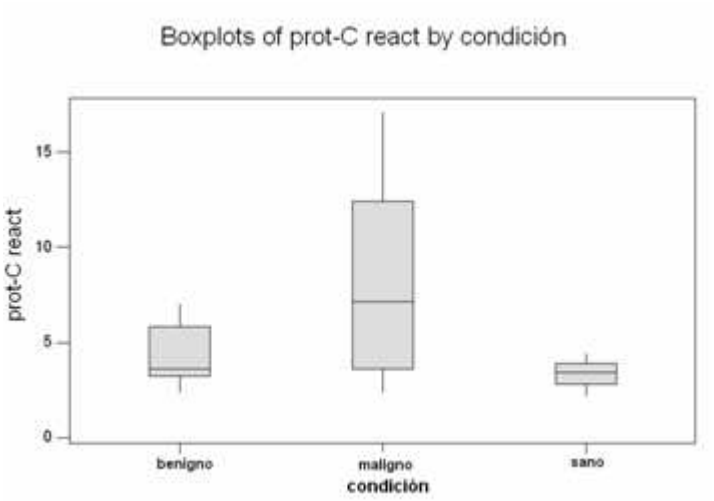
b) Los datos de " Determinación de proteina C-reactiva en hembras caninas con tumores mamarios benignos y malignos" [1] fueron procesados con MINITAB, siguiendo la secuencia de comandos indicados más arriba, obteniéndose los estadísticos y el gráfico que se muestran a continuación.

[1] R. Crossley, et al, Escuela Medicina Veterinaria, Univ.Santo Tomás

Descriptive Statistics: prot-C react

Variable	condición	N	Mean	StDev	Minimum	Q1	Median	Q3
prot-C react	benigno	10	4,290	1,543	2,400	3,225	3,600	5,850
	maligno	10	8,20	5,21	2,40	3,62	7,15	12,40
	sano	10	3,370	0,685	2,200	2,800	3,450	3,925

Variable	condición	Maximum
prot-C react	benigno	7,000
	maligno	17,10
	sano	4,400



El cuadro muestra diferencias de promedios (Mean) de proteina C entre las tres condiciones de las hembras caninas, con un valor claramente superior entre las hembras con tumores malignos. Analizando los valores de la mediana (Median) se verifica que estos son muy similares entre los grupos *sano* y *benigno*, pero con un valor muy superior para el grupo de los *malignos*, lo que se ilustra en el gráfico de caja (boxplot), en el cual se aprecia, además, la gran dispersión en contenido de proteina-C entre las hembras con tumores malignos, al punto que sus valores menores se confunden con los de los otros dos grupos, lo que se constata en la coincidencia de los valores mínimos de los tres grupos. Esto significa que, si se desea utilizar esta técnica para determinar tumores malignos, valores bajos de proteina-C no son discriminatorios, por lo que un valor bajo de proteina-C no permite descartar tumores malignos.

La búsqueda de valores que permitan diferenciar tumores malignos de benignos hay que centrarla, entonces, en los valores altos, donde la **mediana** del grupo de tumores malignos se ve, en el gráfico, que supera a todos los de los otros dos grupos, razón por la cual se podría adoptar la mediana 7,15, como valor límite inferior para decidir cuando un tumor es maligno.

Este caso puede ser un claro ejemplo en que la mediana se comporta mejor que la media aritmética para comparar grupos, debido a la gran diferencia de dispersión entre estos.

2. PROBABILIDAD

2.1 Modelos Matemáticos.

En el desarrollo histórico de los esfuerzos por conocer la realidad han habido tres ideas creativas que han sido fundamentales a las ciencias, cada una en su época: la idea del orden, la idea de la causa mecánica y la de la probabilidad. Para los antiguos la ciencia consistía principalmente en ordenar las cosas. A partir de Galileo y Newton la ciencia pasó a ser la búsqueda de las causas de los fenómenos observables. Actualmente una buena parte de la ciencia moderna tiene como concepto primordial la probabilidad de ocurrencia de ciertos comportamientos. (Extractado de "La ciencia su método y su historia", Silvia Bravo, 1991).

Todo modelo es una representación aproximada de la realidad y no es sensato intentar desarrollar un modelo que la represente en forma exacta. El modelo debe ser adecuado, pero simple, luego no debe incluir técnicas sofisticadas que aporten una mayor precisión innecesaria o que requieran información difícil de obtener o cara. En la elaboración de un modelo se hacen algunos supuestos básicos cuya validez debe ser probada. La validación de un modelo exige deducir un cierto número de consecuencias y corroborarlas con las observaciones.

Por lo tanto un buen modelo es aquel que une la simplicidad con una razonable aproximación a la realidad, sin omisiones importantes en el desarrollo del fenómeno.

Los fenómenos naturales se clasifican en dos tipos.

Fenómenos determinísticos.

Son aquellos en los que el resultado esperado queda determinado por las condiciones bajo las cuales se realiza, es decir, son predecibles.

Muchos de los fenómenos de la física o de la química, que se estudian en la enseñanza media o en un primer año universitario, satisfacen esta condición y por lo tanto el modelo matemático que los describe corresponde a una ecuación. Así, la ley de Boyle-Mariotte que relaciona la presión y volumen de un gas a temperatura constante; la fórmula $d = v*t$ que relaciona la distancia recorrida por un móvil que mantiene cierta rapidez media v durante un tiempo t, o $2H_2 + O_2 \rightarrow 2H_2O$, son ejemplos de este tipo de fenómenos.

Fenómenos no determinísticos o aleatorios.

Son aquellos en los cuales el azar tiene una participación importante y por lo tanto los modelos determinísticos no son adecuados, pues el resultado de estos fenómenos no son predecibles con exactitud y por lo tanto se utilizan modelos matemáticos estocásticos para describirlos, los cuales llevan incorporados una componente que representa la *incertidumbre*. Así, el resultado del lanzamiento de una dado; de una moneda; la cantidad de agua lluvia que cae en una estación meteorológica durante un año; cantidad de partículas emitidas en un intervalo de tiempo por una fuente radiactiva; producción en qq/ha de una variedad de trigo o el tiempo de espera en un paradero por un bus, son algunas de las innumerables situaciones de este tipo de fenómenos.

En resumen se puede decir que un **modelo determinístico** supone que el resultado está determinado por las condiciones iniciales, mientras que en un **modelo estocástico** las condiciones experimentales determinan solamente el comportamiento probabilístico de los resultados posibles.

Características de los experimentos aleatorios.

En lo sucesivo se utilizará el término experimento, pues es necesario poder realizarlos a voluntad. Sus características son:

1º Es posible repetirlo indefinidamente sin cambiar esencialmente las condiciones en que se realiza.
2º No es posible predecir un resultado en particular.
3º Es posible describir el conjunto de todos los resultados posibles.
4º A medida que el experimento se repite los resultados parecen ocurrir en forma caprichosa, pero cuando el experimento se repite un número grande de veces se observa un comportamiento de regularidad que lo caracteriza.

2.2 Espacio muestral y eventos.

Estos son los conceptos a base de los cuales se formaliza toda la teoría de las probabilidades, cuyas definiciones y ejemplos se dan a continuación.

Definición.

Se llama **espacio muestral** al conjunto S de todos los resultados posibles de un experimento o fenómeno aleatorio ε.

Es el símil al concepto de población y puede haber más de un espacio muestral para un mismo experimento. Ejemplos de experimentos aleatorios con sus posibles espacios muestrales se listan a continuación:

ε_1: lanzamiento de una moneda ; S $= \{c\,,\, s\}$.
ε_2: lanzamiento de dos monedas ; $S_1 = \{(c\,,c),\, (c,\, s),\, (s,\, c),\, (s,\, s)\}$ que corresponde al espacio muestral *más detallado* o $S_2 = \{0,\, 1,\, 2\}$ si lo que interesa es indicar el número de caras obtenidas en cada lanzamiento. Hay que diferenciar entre el resultado (c, s) y (s, c), lo que se puede explicar utilizando el artificio de que las dos monedas están pintadas de color diferente, supóngase rojo y blanco, entonces (c, s) corresponde a obtener *cara* con la moneda roja y *sello* con la moneda blanca, mientras que (s, c) corresponde a la situación inversa. También puede razonarse haciendo la consideración que la moneda es la misma y que se lanza dos veces.
ε_3: lanzamiento de un dado ; S $= \{1, 2,\, 3,\, 4,\, 5,\, 6\}$.
ε_4: lanzamiento de dos dados ; en este caso el espacio muestral más detallado es el producto cruz $\{1, 2, 3, 4, 5, 6\}$x$\{1,\, 2,\, 3,\, 4,\, 5,\, 6\}$, es decir, S $= \{(1,1), (1,2), ..., (3,4), ...(6,6)\}$.
ε_5: medición del agua lluvia diaria caida en una estación de monitoreo ; S $= \{$h/ $0 \leq$ h $\leq 100\}$, asumiéndose que el agua caída en ese lugar es **imposible** que supere los 100 mm.

ε_6: medición del rendimiento, en qq/ha, de una variedad de trigo ; S $= \{$p/ $0 \leq$ p $\leq 80\}$. Aunque se piense que no se va a dar un rendimiento nulo no hay inconveniente en que el

espacio muestral los incluya, como se verá más adelante. Lo que no puede suceder es que el espacio muestral "quede corto".
ε_7: número de plantas enfermas al seleccionar 3 plantas de un vivero ; S $= \{0, 1, 2, 3\}$.

Definición.

Se llama **suceso o evento** a cualquier subconjunto del espacio muestral, incluídos el propio S y el conjunto vacío..

Para designar sucesos se utilizan las primeras letras del abecedario en mayúsculas: A, B, C,..., así A $= \{c\}$ es un suceso asociado a ε_1 ; B $= \{(c, s), (s, c)\}$ es un suceso asociado a ε_2 ; C $= \{1, 6\}$ y D $= \{2, 4, 6\}$ son sucesos asociados a ε_4 ; E $= \{$h/ $15 \leq$ h $\leq 30\}$ y F $= \{$p/ p $> 45\}$ son sucesos asociados a ε_5 y ε_6 respectivamente.

Notación de sucesos.

Con la finalidad de tener un lenguaje para la probabilidad exenta de ambigüedad es necesario establecer una notación precisa para expresar nuevos sucesos a partir de la combinación de dos o más de ellos. Esta notación se logra a través del uso de la teoría de conjuntos. El área sombreada de cada figura representa el sector en el cual se ubica el resultado del experimento.

Si $s \in S$ es el resultado de un experimento, entonces se dice que:

1) ocurre un suceso A si y solo si $s \in A$, que se denotará por $\boldsymbol{A}$

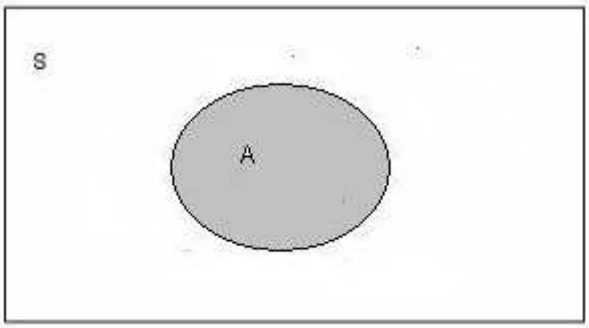


Figura 2.1. Ocurre A

2) no ocurre el suceso A si y solo si $s \in A'$, que se denotará por $\boldsymbol{A'}$

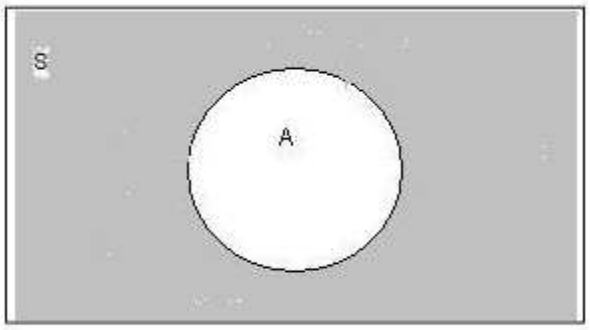


Figura 2.2. No ocurre A

3) ocurre A *o* B *o ambos* si y solo si $s \in (A \cup B)$, que se denotará por $\boldsymbol{A \cup B}$

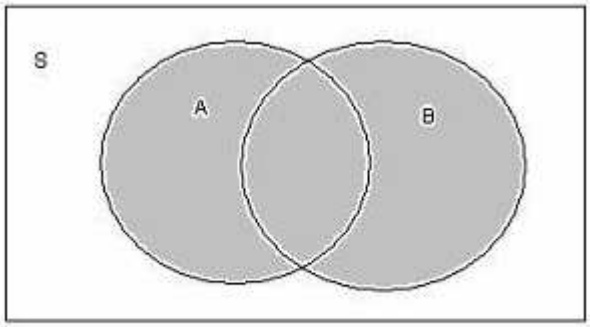


Figura 2.3. Ocurre A o B o ambos

4) ocurre A *y* B si y solo si $s \in (A \cap B)$, que se denotará por $\boldsymbol{A \cap B}$

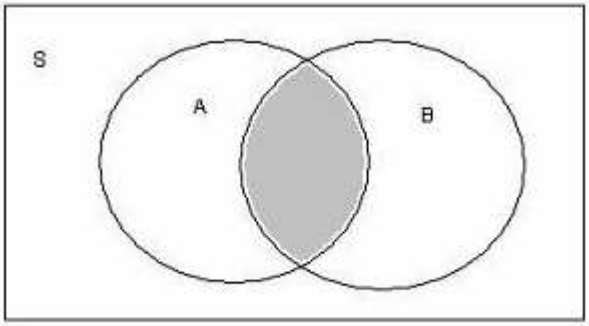


Figura 2.4. Ocurre A y B

5) ocurre A *y* no ocurre B, equivalente a decir *ocurre sólo A,* si y solo si $s \in (A \cap B')$, que se denotará por $\boldsymbol{A \cap B'}$

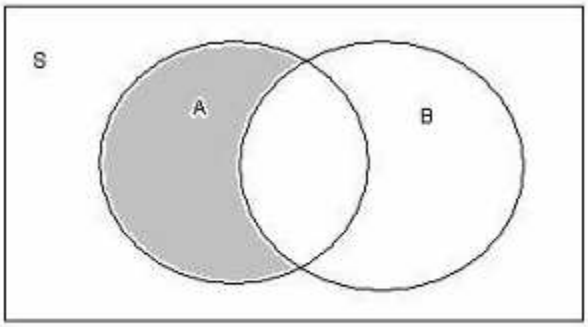


Figura 2.5. Ocurre A , pero no ocurre B

6) no ocurre A *ni* ocurre B, equivalente a decir *no ocurre ninguno de los sucesos* si y solo si $s \in (A' \cap B')$, que se denotará por $\boldsymbol{A' \cap B'}$**.**

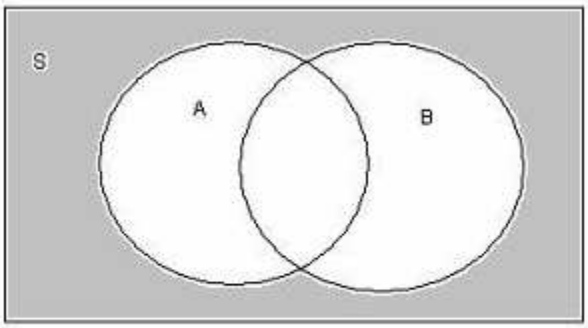


Figura 2.6. No ocurre A y no ocurre B

7) A *y* B no ocurren juntos si y solo si $A \cap B = \phi$

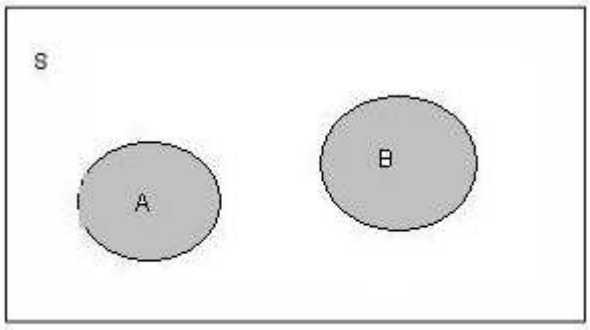


Figura 2.7. A y B no ocurren juntos, son sucesos mutuamente excluyentes.

<u>Definición</u>.

Se dice que los sucesos A y B son **mutuamente excluyentes** si no pueden ocurrir juntos.

La condición de exclusión es muy importante, porque permite establecer que si uno de los sucesos ocurre, entonces el otro no ocurre.

Definición.

Se llama **suceso elemental** a aquel suceso que está constituido por uno de los resultados de un experimento, es decir, es un conjunto unitario.

Así, si un experimento tiene asociado un espacio muestral de cardinalidad *n*, #S= n, entonces existen *n* sucesos elementales vinculados $A_i = \{s_i\}$, i = 1, 2, 3, ...,n.

Ejemplos 2.1.

a) Al considerar los sucesos $A = \{1, 6\}$, $B = \{2, 4, 6\}, C = \{1, 3, 5\}$, $D = \{6\}$ asociados al experimento ε_3 se establece que D es un suceso elemental, que D y C son sucesos mutuamente excluyentes y que B y C son sucesos complementarios, $B' = C$, y por lo tanto son también mutuamente excluyentes.

b) El espacio muestral asociado al experimento ε_4 se puede descomponer en 36 sucesos elementales $A_1 = \{(1,1)\}, A_2 = \{(1,2)\}, A_3 = \{(1,3)\}$, ..., $A_{36} = \{(6,6)\}$.

2.3 Frecuencia relativa, la probabilidad y sus propiedades.

Sea ε un experimento que se repite *n* veces, A un suceso cualquiera asociado a éste y f_A la frecuencia absoluta del suceso A, entonces la frecuencia relativa de A es $h_A = f_A/n$. La frecuencia relativa tiene las siguientes propiedades:
1º $0 \leq h_A \leq 1$
2º $h_A = 1$ si y solo si A ocurre en las *n* repeticiones, es decir, *ocurre siempre*.
3º $h_A = 0$ si y solo si A ocurre *nunca* en las *n* repeticiones.
4º Si A y B son dos sucesos mutuamente excluyentes, entonces $h_{(A \cup B)} = h_A + h_B$
5º Cuando $n \to \infty$, entonces la frecuencia relativa h_A tiende a la probabilidad del suceso A. De esta forma se puede considerar que h_A es la probabilidad *empírica* de A.

Tomando como modelo la frecuencia relativa y sus propiedades se establece la siguiente definición.

Definición.

Sea S un espacio muestral asociado a un experimento ε y P una función que le asocia a cada suceso de S un número real bajo las siguientes condiciones:
1º $0 \leq P(A) \leq 1$, para todo $A \subseteq S$
2º $P(S) = 1$
3º Si $A \cap B = \phi$ implica que $P(A \cup B) = P(A) + P(B)$, entonces P es una **probabilidad para S** y **(S,P)** se designa como un **espacio de probabilidad de S.**

Consecuencia.

Si en un espacio muestral finito S, de cardinalidad #S = n, se conoce la probabilidad p_i de cada suceso elemental de S, que satisfacen las condiciones,

i) $p_i \geq 0$, i = 1, 2, 3, ..., n y ii) $\sum_{i=1}^{n} p_i = 1$, entonces todo suceso A tiene asignada una probabilidad que se puede deducir a partir de los sucesos elementales, pues A siempre se

puede expresar como la unión de sucesos elementales y estos por definición son mutuamente excluyentes.
Por ejemplo $A = \{2, 4, 5\} = \{2\} \cup \{4\} \cup \{5\}$ y por lo tanto
$P(\{2, 4, 5\}) = P(\{2\}) + P(\{4\}) + P(\{5\})$, en virtud de la condición 3º de la probabilidad.

Ejemplos 3.1.

a) Sea $S = \{a, b, c, d\}$ y P tal que $P(\{a\}) = 1/6$, $P(\{b\}) = 1/5$, $P(\{c\}) = 1/3$, $P(\{d\}) = 3/10$ y el suceso $A = \{a, c, d\}$, entonces P es una probabilidad bien definida para S, porque i) $P\{s_i\} \geq 0$, para todo $s_i \in S$ y ii) $\sum_{i=1}^{4} P\{s_i\} = 1/6 + 1/5 + 1/3 + 3/10 = 1$, luego $P(A) = P(\{a\}) + P(\{c\}) + P(\{d\}) = 1/6 + 1/3 + 3/10 = 4/5$.

b) Sea $S = \{1, 2, 3\}$ y P tal que $P(\{1\}) = 1/10$, $P(\{1, 2\}) = 2/5$, $P(\{3\}) = 3/5$.
En este caso P es una probabilidad bien definida, porque se puede determinar $P(\{1\}) = 1/10$, $P(\{2\}) = P(\{1, 2\}) - P(\{1\}) = 3/10$ y $P(\{3\}) = 3/5$, positivos, y $P(\{1\}) + P(\{2\}) + P(\{3\}) = 1$.

c) Sea $S = \{1, 2, 3\}$ y P tal que $P(\{1, 2\}) = 2/5$, $P(\{3\}) = 3/5$.
En esta situación P no es una función de probabilidad , porque no se pueden determinar a partir de las condiciones dadas $P(\{1\})$, $P(\{2\})$, $P(\{1, 3\})$ y $P(\{2, 3\})$.

Las propiedades más importantes de la probabilidad se enuncian y demuestran a continuación.

Teorema 1.

Probabilidad que no ocurra el suceso A: $P(A') = 1 - P(A)$.

Demostración.

$S = A \cup A'$ y $A \cap A' = \phi$, luego $P(S) = P(A) + P(A') = 1$, de acuerdo a la tercera y segunda condición de la probabilidad. De la última igualdad, despejando se tiene $P(A') = 1 - P(A)$.

Teorema 2.

Probabilidad del *suceso imposible*, cuya notación es ϕ: $P(\phi) = 0$.

Demostración.

$P(\phi) = P(S') = 1 - P(S) = 1 - 1 = 0$, por teorema 1 y segunda condición de la probabilidad.

Teorema 3.

Probabilidad que ocurra *al menos uno de los sucesos* A *o* B : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Demostración.

$A \cup B = A \cup (B \cap A')$ y $B = (A \cap B) \cup (B \cap A')$, luego $P(A \cup B) = P(A) + P(B \cap A')$ por ser A y $(B \cap A')$ sucesos mutuamente excluyentes. $P(B) = P(A \cap B) + P(B \cap A')$, pues $(A \cap B)$ y $(B \cap A')$ son mutuamente excluyentes. Despejando P($B \cap A'$) de la última igualdad y sustituyéndola en la anterior se obtiene $P(A \cup B) = P(A) + (P(B) - P(A \cap B))$ que corresponde a la propiedad enunciada.

Teorema 4.

Probabilidad que ocurra al menos uno de los sucesos A, B o C: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Demostración.

La demostración se consigue aplicando recurrentemente el teorema 3.

Teorema 5.

Probabilidad que entre dos sucesos A y B ocurra **sólo** A : $P(A \cap B') = P(A) - P(A \cap B)$

Demostración.

$A = A \cap S = A \cap (B \cup B') = (A \cap B) \cup (A \cap B')$, usando propiedades de conjuntos. Además como $(A \cap B)$ y $(A \cap B')$ son sucesos mutuamente excluyentes $P(A) = P((A \cap B) \cup (A \cap B')) = P(A \cap B) + P(A \cap B')$. Despejando $P(A \cap B')$ de la igualdad se obtiene la propiedad buscada.

Teorema 6.

Probabilidad que no ocurra el suceso A ni ocurra el suceso B: $P(A' \cap B') = 1 - P(A \cup B)$.

Demostración.

Una propiedad en teoría de conjunto establece que $(A \cup B)' = (A' \cap B')$, luego $P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$, aplicando el teorema 1.

Consecuencia.

Una propiedad muy útil en probabilidad dice que "la probabilidad que ocurra **al menos uno** de entre varios sucesos es igual a 1 menos la probabilidad que **no ocurra ninguno de los sucesos**". Esta propiedad se deduce del teorema 6, que en el caso de dos sucesos se expresa

como $P(A \cup B) = 1 - P(A' \cap B')$ y en el caso de tres sucesos como $P(A \cup B \cup C) = 1 - P(A' \cap B' \cap C')$.

Teorema 7.

Si $A \subset B$, entonces $P(A) \leq P(B)$.

Demostración.

$B = A \cup (B \cap A')$, luego $P(B) = P(A) + P(B \cap A')$, por lo tanto $P(B) \geq P(A)$, pues $P(B \cap A') \geq 0$.

Ejemplos 3.2

a) Dada $P(A) = 1/2$, $P(B) = 1/3$ y $P(A \cap B) = 1/5$, se puede establecer que :

- $P(B') = 1 - P(B) = 1 - 1/3 = 2/3$, por teorema 1.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/3 - 1/5 = 19/30$, por teorema 3.
- $P(A' \cap B) = P(B) - P(A \cap B) = 1/3 - 1/5 = 2/15$, por teorema 5.
- $P(A' \cup B') = P(A \cap B)' = 1 - P(A \cap B) = 1 - 1/5 = 4/5$, por otra propiedad de conjuntos que establece que (A ∩ B)' = (A' ∪ B') y teorema 1.
- $P(A' \cup B) = P(A') + P(B) - P(A' \cap B) = (1 - 1/2) + 1/3 - 2/15 = 7/10$.

b) En un vivero una planta puede tener una enfermedad X con probabilidad $1/5$, otra enfermedad Y con probabilidad $2/7$ y la enfermedad X o la enfermedad Y o ambas con probabilidad $3/7$ ¿Cuál es la probabilidad de que una planta cualquiera tenga:

i) ambas enfermedades ? ; ii) sólo la enfermedad Y ? ; iii) no esté enferma ?
Del enunciado se establece $P(X) = 1/5$; $P(Y) = 2/7$ y $P(X \cup Y) = 3/7$, entonces

i) se debe determinar $P(X \cap Y)$. Al despejar la probabilidad de la intersección en el teorema 3, se establece que $P(X \cap Y) = P(X) + P(Y) - P(X \cup Y) = 1/5 + 2/7 - 3/7 = 2/35$.
ii) lo que se desea es $P(X' \cap Y)$, es decir, que no tenga la enfermedad X y tenga la enfermedad Y, por lo tanto $P(X' \cap Y) = P(Y) - P(X \cap Y) = 2/7 - 2/35 = 8/35$.
iii) que no esté enferma significa que no tenga la enfermedad X y no tenga la enfermedad Y, luego se debe calcular $P(X' \cap Y') = 1 - P(X \cup Y) = 1 - 3/7 = 4/7$.

2.4 Probabilidad en espacio muestral finito equiprobable.

Un espacio muestral S es finito si su cardinalidad es un número natural *n* y es equiprobable si todos los resultados de un experimento ε tienen la misma posibilidad de ocurrir. La condición de equiprobabilidad debe justificarse cuidadosamente.

Ejemplos 4.1

Considérense los siguientes experimentos y sus correspondientes espacios muestrales.

a) ε_1 : lanzamiento de un dado simétrico y $S = \{1, 2, 3, 4, 5, 6\}$, entonces S es un espacio muestral finito equiprobable.

b) ε_2 : lanzamiento de una moneda equilibrada y $S = \{c, s\}$, entonces S es un espacio muestral finito equiprobable.

c) ε_3 : dos lanzamientos de una moneda equilibrada y $S = \{(c, c), (c, s), (s, c); (s, s)\}$, entonces S es un espacio muestral finito equiprobable.

d) ε_4 : dos lanzamientos de una moneda equilibrada y $S = \{0, 1, 2\}$, donde 0, 1 o 2 indican el número de caras obtenidas en ambos lanzamientos. Entonces S **no** es un espacio equiprobable, porque $\{0\}$ es equivalente a $\{(s, s)\}$; $\{1\}$ es equivalente a $\{(c, s), (s, c)\}$ y $\{2\}$ es equivalente a $\{(c, c)\}$.

e) ε_5 : extracción de 3 fichas al azar, sin sustitución, de una bolsa que contiene 6 fichas rojas, 4 blancas y 5 azules. Entonces, si S es el conjunto de todas las combinaciones posibles de 15 fichas tomadas de a 3, éste es un espacio muestral finito equiprobable de $\binom{15}{3} = 455$ resultados.

f) Si en el mismo experimento anterior S representa el número de fichas rojas obtenidas, entonces S **no** es un espacio muestral equiprobable, pues el número de combinaciones que no contienen fichas rojas es distinto al número que contiene una roja y distinto al que contiene dos rojas y distinto al que contiene las tres rojas, luego sus posibilidades son distintas.

Asignación de probabilidades en espacios muestrales finitos equiprobables.

Si S es un espacio muestral finito equiprobable, entonces hay *n* resultados con igual probabilidad *p*, para los cuales se debe satisfacer que: $\sum_{i=1}^{n} P(\{s_i\}) = \sum_{i=1}^{n} p = n*p=1$, de donde resulta que $p = 1/n$. La consecuencia es que en todo espacio muestral equiprobable de cardinalidad *n*, cada suceso elemental tiene probabilidad $P(\{s_i\}) = 1/\#S = 1/n$ y por lo tanto cualquier suceso asociado a este espacio muestral tiene una probabilidad asociada directamente proporcional a su cardinalidad. A partir de esta condición se establece la *definición clásica* de probabilidad de sucesos en los siguiente términos $P(A) = \#A/\#S = \frac{\text{número de casos favorables}}{\text{número de casos posibles}}$.

Ejemplos 4.2

a) Si $S = \{(c,c),(c,s),(s,c);(s,s)\}$ es un espacio equiprobable correspondiente al lanzamiento de dos monedas legales, entonces

- la probabilidad de obtener 2 caras es $P((c,c)) = 1/4$, pues hay 1 resultado favorable entre 4 resultados posibles
- la probabilidad de obtener 1 cara es $P((c,s),(s,c)) = 2/4 = 1/2$.

b) Con una bolsa que contiene 6 fichas rojas, 4 blancas y 5 azules, se realiza el experimento:

i) ε: extraer **una** ficha al azar. En este caso el espacio muestral equiprobable es el conjunto de las 15 fichas, bajo el supuesto que la única diferencia entre las fichas es su color. Entonces la probabilidad de que la ficha obtenida sea de uno de los tres colores posibles es proporcional al número de fichas de ese color, o sea, $P(\text{azul}) = 5/15$, $P(\text{blanca}) = 4/15$ y $P(\text{roja}) = 6/15$.

ii) ε : extracción de 3 fichas al azar, sin sustitución. Este es el experimento ε_5 del ejemplo 4.1 y el espacio muestral equiprobable, cuyos elementos son conjuntos ternarios de la forma {r, b, r} o {a, a, a}, es muy amplio para expresarlo por extensión, que por lo demás no interesa, porque sólo es importante su cardinalidad, que como se explicó antes corresponde a las combinaciones entre 15 fichas tomadas de a 3, o sea, $\#S = \binom{15}{3} = 455$. Probabilidades tipo, asociadas a este experimento, se calculan a continuación:

- $P(\text{3 fichas blancas}) = \binom{4}{3}/\binom{15}{3} = 4/455$, pues hay 4 combinaciones para obtener 3 fichas blancas.

-$P(\text{una ficha de cada color}) = P(\text{1 roja, 1 azul y 1 blanca}) = \frac{\binom{6}{1}*\binom{5}{1}*\binom{4}{1}}{\binom{15}{3}} = \frac{6*5*4}{455} = 24/91$, esto se explica porque hay 6 formas de seleccionar una ficha roja, 5 para una ficha azul y 4 para blanca y 120 formas de que sea una de cada color.

-$P(\text{dos fichas rojas y una azul}) = \frac{\binom{6}{2}*\binom{5}{1}}{\binom{15}{3}} = 75/455$, pues dos fichas rojas se pueden obtener como combinación de dos fichas elegidas de entre las 6 rojas que hay.

- $P(\text{al menos una ficha roja}) = 1 - P(\text{ninguna roja}) = 1 - \frac{\binom{9}{3}}{\binom{15}{3}} = 1 - 84/455 = 371/455$, utilizando la consecuencia del teorema 6 y por qué 3 fichas no rojas se pueden elegir de entre las 9 fichas que son blancas o azules.

- $P(\text{a lo más 2 fichas rojas}) = P(\text{ninguna roja o 1 roja o 2 rojas})$

$$= P(\text{ninguna roja}) + P(\text{1 roja}) + P(\text{2 rojas})$$

$$= \frac{\binom{9}{3}}{\binom{15}{3}} + \frac{\binom{9}{2}*\binom{6}{1}}{\binom{15}{3}} + \frac{\binom{9}{1}*\binom{6}{2}}{\binom{15}{3}} = 87/91$$

Tanto en este caso como en el anterior el espacio muestral corresponde al número de fichas rojas obtenidas al seleccionar 3 fichas al azar, esto es, $S = \{0,1,2,3\}$ y por lo tanto $\{0\}$, $\{1\}$, $\{2\}$, $\{3\}$ son los sucesos elementales de S y en consecuencia $P(\{0\}) + P(\{1\}) + P(\{2\}) + P(\{3\}) = 1$.

Se puede observar que "a lo más 2 fichas rojas" es equivalente a 0 o 1 o 2 fichas rojas, por lo tanto, despejando $P(\{3\})$ en la igualdad anterior, se establece que

$P(\text{a lo más 2 fichas rojas}) = 1 - P(\{3\}) = 1 - \frac{\binom{6}{3}}{\binom{15}{3}} = 1 - 20/455 = 87/91.$

Por otra parte "al menos una ficha roja" es equivalente a 1 o 2 o 3 fichas rojas. Despejando $P(\{0\})$ de la misma igualdad anterior se tiene que $P(\text{al menos una ficha roja}) = 1 - P(\{0\})$, lo que es otra fundamentación para la importante propiedad utilizada en la probabilidad anterior.

c) Se realiza el experimento que consiste en lanzar un dado simétrico dos veces, luego el espacio muestral equiprobable está formado por los 36 pares ordenados que se obtienen con el producto {1, 2, 3, 4, 5, 6} x {1, 2, 3, 4, 5 ,6}. No es dificultoso expresar este espacio muestral por extensión en los siguientes términos $S = \{(1,1),(1,2),...,(1,6),(2,1),....(2,6),....,(6,5),(6,6)\}$ y a partir de éste calcular las probabilidades de obtener:

- **dos seis**, lo que se plantea $P((6,6)) = 1/36$, pues hay un resultado favorable entre 36 posibles.
- **un tres y cualquier otro número**, lo que equivale a los pares que tengan primer elemento 3 y segundo elemento distinto a tres o viceversa, luego hay 10 pares que cumplen con la condición, en consecuencia $P(\text{sólo un tres en ambos dados}) = 10/36$.
- **al menos un tres**, es equivalente a sólo una vez tres o dos veces tres, luego $P(\text{al menos un tres}) = P(\text{sólo un tres}) + P((3,3)) = 10/36 + 1/36 = 11/36$. Otra forma consiste en aplicar la propiedad $P(\text{al menos un tres}) = 1 - P(\text{ningún tres}) = 1 - 25/36 = 11/36$, pues con el primer y segundo dado habría que obtener {1,2,4,5,6}, cuyo producto cruz corresponde a 25 pares ordenados.
- **seis puntos en total**. Sea $A = \{(x,y)/\, x + y = 6\} = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$, entonces $P(A) = 5/36$.
- **un par**, o sea, el suceso $B = \{(x,y)/\, x = y\} = \{(1,1),(2,2),(3,3),.....;(6,6)\}$ y por lo tanto $P(\text{un par}) = P(B) = 6/36$.
- **un número menor en el primer lanzamiento que con el segundo**, que queda representado por el suceso $C = \{(x,y)/\, x < y\} = \{(1,2),(1,3),...,(1,6),(2,3),....,(5,6)\}$. Este suceso tiene cardinalidad 15 y por lo tanto $P(C) = 15/36$.

d) Los 25 huertos de una localidad se clasificaron en términos del sistema de riego en *tecnificado (T)* o *surco (S)* y de su tamaño en *mediano (M)* o *pequeño (P).* Se encontraron que 13 huertos son de tamaño pequeño; 10 riega por surco ; 5 de tamaño pequeño y riego tecnificado. Se necesita realizar una encuesta en la localidad para lo cual se deben seleccionar 5 huertos al azar. Interesa calcular la probabilidad de que los 5 huertos seleccionados
i) tengan riego tecnificado ; ii) sean de tamaño mediano ; iii) sean de tamaño pequeño y tengan riego tecnificado; iv) sean de tamaño mediano y rieguen por surco.

Lo primero es cruzar la información en una tabla 2 por 2 e ir ubicando la información entregada como se muestra en la primera tabla. Las siguientes celdas se rellenan por defecto como ocurre en la segunda tabla.

tipo riego \ tamaño	M	P	Total
T		**5**	
S			**10**
Total		**13**	**25**

→

tipo riego \ tamaño	M	P	Total
T	10	**5**	15
S	2	8	**10**
Total	12	**13**	**25**

A continuación se trata de identificar los valores adecuados para calcular las probabilidades de interés.

i) $P(5\,T) = \frac{\binom{15}{5}}{\binom{25}{5}} = 3003/53130 = 0,0565$, pues 15 son los huertos con riego tecnificado.

ii) $P(5\,M) = \frac{\binom{12}{5}}{\binom{25}{5}} = 792/53130 = 0,0149$, pues 12 son los huertos de tamaño mediano.

iii) $P(5 \text{ de } (T \cap P)) = \frac{\binom{5}{5}}{\binom{25}{5}} = 1/53130 = 0,00002$, pues son sólo 5 los huertos pequeños y con riego tecnificado. De acuerdo a la probabilidad obtenida es *muy difícil* que esta situación pueda ocurrir.

iv) $P(5 \text{ de } (M \cap S)) = 0$. Este suceso es *imposible* que ocurra, porque se deben elegir 5 de esa condición y existen sólo 2.

2.5 Probabilidad condicional.

Considérese la bolsa con 6 fichas rojas, 5 azules y 4 blancas de la cual se extraen fichas, una a una, definiéndose los sucesos $A = \{$la 1ª ficha obtenida es blanca$\}$ y $B = \{$la 2ª ficha obtenida es blanca$\}$, entonces la probabilidad de B dependerá de lo que ocurra antes de extraer la 2ª ficha lo que se puede realizar de dos formas.

i) con sustitución

En este caso después de cada extracción la bolsa se mantiene en las mismas condiciones iniciales cada vez, por lo tanto $P(A) = P(B) = 4/15$, es decir, la probabilidad en cada extracción es constante.

ii) sin sustitución

En esta situación después de extraer la 1ª ficha y no restituirla, la condición inicial de la bolsa ha sido modificada, por lo tanto $P(A) = 4/15$, pero para determinar $P(B)$ es necesario conocer la composición de la bolsa *después* de extraer la 1ª ficha y ello depende de si ocurrió o no el suceso A, o sea, la probabilidad de B está condicionada a la ocurrencia o no ocurrencia de A.

Este nuevo concepto necesita explicarse y para ello se debe tener una notación adecuada. P(B/A) designa la probabilidad de que ocurra B dado que ha ocurrido A, lo que se lee "probabilidad de B dado A". Para el caso de los dos sucesos definidos antes, corresponde a la probabilidad de que la segunda ficha sea blanca dado que la primera lo fue y en consecuencia después de la primera extracción en la bolsa hay catorce fichas de las cuales sólo tres son blancas, por lo cual P(B/A) = P(la 2ª ficha sea blanca dado que la 1ª fue blanca) = 3/14. También, P(B/A') = P(la 2ª ficha sea blanca dado que la 1ª no lo fue) = 4/14 o P(B'/A) = 11/14.

P(B/A) significa que se está calculando la probabilidad de B **referida al espacio muestral reducido A**, en vez de referirla al espacio muestral original S.

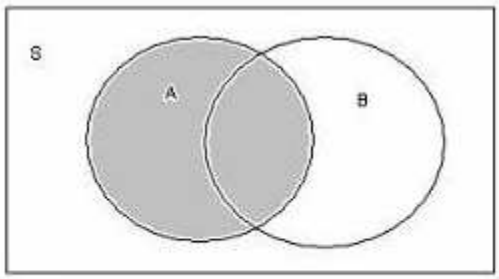


Figura 5.1. Resultado de un experimento condicionado a la ocurrencia de A

Cuando se calcula P(B) se está preguntando que tan probable es que el resultado esté en B sabiendo que está en S, mientras que cuando evaluamos P(B/A) la pregunta es que tan probable es que el resultado esté en B sabiendo que está en A. El área sombreada en la figura 5.1 representa la ocurrencia del suceso A y B/A significa que haya ocurrido B habiendo ocurrido A, representada en la figura 5.2 por el área más oscura, que corresponde a la intersección de A y B, pero referida al suceso A.

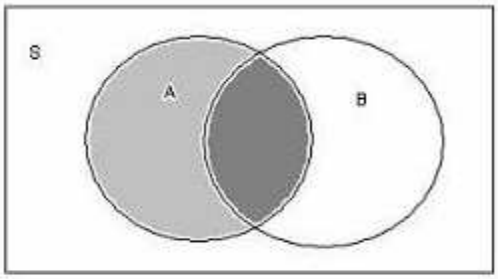


Figura 5.2. La zona más oscura representa B/A.

De los conceptos anteriores surgen las siguientes definiciones.

<u>Definiciones</u>.

Dado dos conjuntos A y B cualesquiera asociados a un espacio muestral S, entonces

1º $P(B/A) = P(A \cap B)/P(A)$, $P(A) > 0$
2º $P(A/B) = P(A \cap B)/P(B)$, $P(B) > 0$

<u>Observaciones</u>.

1) Cuando las probabilidades están condicionada a un suceso cualquiera, denominado A o B o C, entonces tal suceso pasa a tener formalmente las características de un espacio muestral, reducido en relación al espacio original S, de modo que todas las propiedades de la probabilidad que se cumplen en S son también válidas en el espacio muestral reducido. De hecho cuando se plantea la probabilidad de B, P(B), es totalmente concordante a denotarla como P(B/S).
2) Consecuente con la observación anterior es posible demostrar las siguientes propiedades:
P(B'/A) = 1 - P(B/A) , equivalente teorema 1
P(B'/A') = 1 - P(B/A') , equivalente teorema 1
P((B ∪ C)/A) = P(B/A) + P(C/A) - P((B ∩ C)/A) , equivalente teorema 3
P(B ∩ C')/A) = P(B/A) - P((B ∩ C)/A) , equivalente teorema 5

Ejemplos 5.1

a) Si $P(A) = 2/5$, $P(B) = 2/3$, $P(A \cap B) = 1/6$, entonces

- $P(A/B) = P(A \cap B)/P(B) = \frac{1/6}{2/3} = 1/4$
- $P(B/A) = P(A \cap B)/P(A) = \frac{1/6}{2/5} = 5/12$
- $P(B'/A) = 1 - P(B/A) = 1 - 5/12 = 7/12$
- $P(B/A') = P(A' \cap B)/P(A') = \frac{P(B)-P(A\cap B)}{P(A')} = \frac{2/3-1/6}{1-2/5} = \frac{3/6}{3/5} = 5/6$
- $P(B'/A') = 1 - P(B/A') = 1 - 5/6 = 1/6$

b) Se lanza un dado. Si el resultado es par ¿cuál es la probabilidad de que sea el número 6?

- $P(\text{seis}/\text{par}) = 1/3$, porque si ocurre par hay sólo tres resultados posibles de los que uno de ellos es el 6. También haciendo uso de la definición $P(\text{seis}/\text{par}) = \frac{P(\text{par y seis})}{P(\text{par})} = \frac{P(\text{seis})}{P(\text{par})} = \frac{1/6}{3/6} = 1/3.$

c) La siguiente tabla corresponde al ejemplo 4.2 d)

tipo riego \ tamaño	M	P	Total
T	10	5	15
S	2	8	10
Total	12	13	25

.de la cual se pueden calcular las siguientes probabilidades al seleccionar un huerto al azar.

- $P(\text{huerto con riego tecnificado}) = 15/25 = 3/5$
- $P(\text{huerto pequeño con riego tecnificado}) = 5/25 = 1/5$
- $P(\text{huerto pequeño/riego tecnificado}) = 5/15 = 1/3$, pues los huertos con riego tecnificado son 15 de los cuales 5 son de tamaño pequeño.
- $P(\text{riego por surco/huerto mediano}) = 2/12 = 1/6$, pues los huertos medianos son 12 de los cuales 2 riegan por surco.

d) Del ejemplo 4.2. c) se tienen los sucesos $A = \{(x,y)/\, x + y = 6\}$, $B = \{(x,y)/\, x = y\}$ y $C = \{(x,y)/\, x < y\}$, cuyas probabilidades son $P(A) = 5/36$, $P(B) = 1/6$ y $P(C) = 5/12$. Se puede establecer las siguientes probabilidades condicionales.

- $P(A/B) = 1/6$, porque hay 6 pares que cumplen con B y sólo uno de ellos suma seis.
- $P(B/A) = 1/5$, porque hay 5 pares que cumplen con A y sólo uno de ellos es un par.
- $P(A/C) = 2/15$, porque hay 15 pares que cumplen con C de los cuales 2 cumplen con A.
- $P(C/B) =$ 0 , porque hay 6 pares que cumplen con B y ninguno cumple con C.

Las 4 probabilidades anteriores se calcularon usando el camino más sencillo, pero las mismas probabilidades se calculan usando la definición.

e) Con el fin de aportar mayor claridad al concepto de probabilidad condicional considérese el ejemplo introductorio de probabilidad condicional, consistente en extraer sin sustitución dos fichas de una bolsa y los sucesos A = { la 1ª ficha sea blanca } y B = { la 2ª ficha sea blanca }, determinándose, usando el método simplificado, que P(B/A) = 3/14. El procedimiento a continuación es el que se debe realizar para calcular esta probabilidad haciendo uso de la definición. Si se extraen de la bolsa dos fichas sin sustitución, entonces #S = $\binom{15}{2}$ = 105 y #(A ∩ B) = $\binom{4}{2}$ = 6 , luego P(A ∩ B) = 2 / 35. Para calcular P(A) es necesario tener en cuenta que el orden es importante porque así está definido el suceso A, de manera que ahora se trata de variaciones, de modo que #S = 15∗14 = 210 , pues la primera ficha seleccionada puede ser cualquiera de las 15 y la segunda cualquiera de las restantes y #A = 4∗14 = 56 , pues la

primera *debe ser blanca* y la segunda cualquiera de las 14 restantes, de donde P(A) = 56 / 210 = 4/15. En consecuencia P(B/A) = P(A ∩ B) / P(A) = $\frac{2/35}{4/15}$ = 3 /14 como se había establecido. Otra forma de analizar la situación anterior consiste en considerar que, en la situación que se está analizando, el **orden** en que son extraídas las fichas es importante, por lo tanto el espacio muestral son variaciones de 15 fichas tomadas de a 2 en vez de combinaciones, es decir, #S = $P_2^{15} = 15 * 14 = 210$, pero la condicionalidad reduce este espacio muestral al suceso A con #A = 4∗14 = 56, entre los cuales hay 4∗3 = 12 que corresponden a dos fichas blancas, luego P(B / A) = 12 / 56 = 3/14.

Observaciones.

1) Los resultados (b_1 , b_2) y (b_2 , b_1) son dos según las variaciones cuando el **orden importa** y sólo uno cuando el **orden no importa** que corresponde a las combinaciones, cuya notación será { b_1 , b_2} con paréntesis de conjunto, donde "b" se refiere a una ficha blanca.
2) Hay dos maneras de calcular la probabilidad condicional P(A / B), directamente considerando la probabilidad de A respecto al espacio muestral reducido B, o usando la definición donde P(A ∩ B) y P(B) se calculan respecto al espacio muestral original S.

Principio multiplicativo de probabilidades.

Como consecuencia de la probabilidad condicional se obtiene el principio multiplicativo general de probabilidad.

Despejando P(A ∩ B) ya sea de la definición 1 como de la definición 2, se deduce que:

Principio multiplicativo general
$P(A \cap B) = P(A/B) * P(B) = P(B/A) * P(A)$

Por conveniencia se adoptará la notación (s_i, s_j) para indicar un orden en los resultados, primero el resultado s_i y segundo el resultado s_j. La notación (s_i y s_j) denotará que el orden no importa, primero s_i y después s_j o viceversa.

Ejemplos 5.2

a) De la bolsa conteniendo 6 fichas rojas (r), 5 azules (a) y 4 blancas (b), se extraen dos fichas sin sustitución, entonces la probabilidad de obtener

- una roja y una azul en ese orden se plantea y calcula P(r ,a) = P(1ª roja)∗P(2ª azul/1ª roja) $\Rightarrow P$(r, a) = $\frac{6}{15} * \frac{5}{14} = 1/7$.

- una roja y una azul en cualquier orden: $P(\text{r y a}) = \frac{\binom{6}{1} * \binom{5}{1}}{\binom{15}{2}} = 2/7$. Se aprecia que en este caso la probabilidad es el doble de la anterior, evidentemente porque la anterior es más restrictiva, exige un orden. La relación entre ambas formas es que cuando se exige un orden, entonces (r y a) es equivalente a (r, a) y (a, r). Luego P(r y a) = P(r, a) + P(a, r) = 1/7 + 1/7 = 2/7.

- dos fichas blancas: P(b, b) = P(b y b) , pues existe un solo ordenamiento de dos fichas blancas, luego, P(b, b) = P(1ª blanca)∗P(2ª blanca/1ª blanca) = $\frac{4}{15} * \frac{3}{14} = 2/35$. El mismo resultado se obtiene con combinatoria para P(b y b) = $\frac{\binom{4}{2}}{\binom{15}{2}} = \frac{6}{105} = 2/35$.

b) De la bolsa anterior se extraen 3 fichas sin sustitución, entonces la probabilidad de obtener

- una roja, una azul y una roja en ese orden:

P(r, a, r) $= P$(1ª roja)$*P$(2ª azul/1ª roja)$*P$(3ª roja/1ª roja y 2ª azul) $= \frac{6}{15}*\frac{5}{14}*\frac{5}{13} = 5/91.$

- dos rojas y una azul en cualquier orden: $P($ 2 rojas y 1 azul) $= \frac{\binom{6}{2}*\binom{5}{1}}{\binom{15}{3}} = \frac{75}{455} = 15/91$

Se puede constatar que esta última probabilidad es 3 veces la anterior debido a que hay tres ordenamientos posibles para extraer dos fichas rojas y una azul, donde cada ordenamiento tiene una probabilidad de 5/91.

- una blanca, una azul y una roja en ese orden:

P(b, a, r) $= P$(1ª blanca)$*P$(2ª azul/1ª blanca)$*P$(3ª roja/1ª blanca y 2ª azul) $= \frac{4}{15}*\frac{5}{14}*\frac{6}{13} = 4/91$

- una de cada color en cualquier orden P(roja y azul y blanca) $= \frac{\binom{6}{1}*\binom{5}{1}*\binom{4}{1}}{\binom{15}{3}} = \frac{120}{455} = 24/91,$

que es 6 veces la probabilidad anterior, esto debido a que existen $3! = 6$ ordenamientos posibles para obtener una ficha de cada color.

c) En cierta carrera un alumno, si estudia lo suficiente, tiene una probabilidad de 0,6 de aprobar cálculo por primera vez, una probabilidad de 0,9 de aprobar estadística si aprobó cálculo la primera vez y de 0,5 en caso contrario. ¿ Cuál es la probabilidad de que un alumno que toma por primera vez cálculo apruebe estadística, si estudia lo suficiente?

Sea C el suceso aprobar cálculo por primera vez, E el suceso aprobar estadística la primera vez, entonces $P(C) = 0,6$, $P(E/C) = 0,9$ y $P(E/C') = 0,5$, luego $P(E) = P(E/C)*P(C) + P(E/C')*P(C') = 0,9*0,6 + 0,5*0,4 = 0,74.$

d) En un invernadero hay 6 plantas de una especie entre las cuales hay 2 que están enfermas con un virus. Se examinan las plantas una a una hasta encontrar las dos enfermas. ¿Cuál es la probabilidad de que la segunda enferma se encuentre i) al examinar la segunda planta?, ii) al examinar la cuarta planta? , iii) después de examinar la cuarta planta?

i) Para encontrar la segunda enferma (E) en el segundo examen es necesario que la primera planta examinada sea una de las enfermas, luego

P(2ª E en 2º examen) $= P$(1ª E y 2ª E) $= P$(1ª E)$*P$(2ª E/1ª E) $= \frac{2}{6}*\frac{1}{5} = 1/15$

ii) Para encontrar la segunda enferma en la cuarta inspección debe ocurrir que entre las tres primeras plantas examinadas haya una enferma y dos sanas, en cualquier orden, y la cuarta planta examinada esté enferma, entonces

P(2ª E en 4º examen) $= P$(E/ 1 E y 2 S en las tres primeras)$*P$(1 E y 2 S en las tres primeras)

$$= \frac{1}{3}*\frac{\binom{4}{2}*\binom{2}{1}}{\binom{6}{3}} = \frac{1}{3}*\frac{3}{5} = 1/5$$

La probabilidad anterior es equivalente a la suma de las probabilidades de los tres sucesos independientes (E, S, S, **E**), (S, E, S, **E**), (S, S, E, **E**).

iii) P(examinar más de 4 plantas para 2ª E) $= P$(examinar 5 plantas) $+ P$(examinar 6 plantas)

$$= \frac{1}{2}*\frac{\binom{4}{3}*\binom{2}{1}}{\binom{6}{4}} + \frac{1}{1}*\frac{\binom{4}{4}*\binom{2}{1}}{\binom{6}{5}} = \frac{4}{15} + \frac{1}{3} = 3/5.$$

Observaciones.

1) El número de ordenamientos posibles entre *n* elementos distintos está dado por *n*!. El número de ordenamientos con *n* elementos entre los cuales hay grupos de elementos iguales de tamaño *a, b, c* se determina por $n!/a!*b!*c!$. Por ejemplo la cantidad de números distintos, de cuatro cifras, que se pueden escribir utilizando los dígitos {2, 4, 5, 7} es igual a 4! , es decir, 24. En cambio utilizando los dígitos {2, 2, 5, 5} sólo se pueden obtener $4!/2!*2!$, es decir, 6 números distintos, de cuatro cifras que son: 2255, 2525, 2552, 5252, 5225 y 5522. Utilizando los dígitos {2, 4, 5, 5, 5} se pueden escribir $5!/1!*1!*3!$, es decir, 20 números distintos.¡ Intente escribirlos todos!
2) Verifique que P(r, r, a) = P(r, a, r) = P(a, r, r) y que P(r, a, b) = P(a, r, b) = ...= P(b, a, r).
3) En las situaciones de extracciones de elementos en los cuales **el orden** en que son obtenidos **no importa**, la extracción uno a uno es equivalente a extraerlos todos en forma simultánea.

Independencia de sucesos.

Para introducir el concepto se revisarán algunas situaciones anteriores.

1. En el ejemplo introductorio de probabilidad condicional cuando se extraen fichas una a una **con sustitución** se verifica que para los sucesos A = {la 1ª ficha extraída sea blanca} y B = {la 2ª ficha extraída sea blanca}, la P(B/A) = 4/15 y esta probabilidad es coincidente con la P(B) = 4/15. Es decir, la probabilidad de B no se ve afectada por la ocurrencia de A.

2. En el ejemplo 5.1 c) se definieron los sucesos $A = \{(x,y)/\, x+y=6\}$, $B = \{(x,y)/\, x=y\}$ y $C = \{(x,y)/\, x<y\}$, determinándose que P(A/B) = 1/6 $\neq$ P(A) y P(A/C) = 2/15 $\neq$ P(A), o sea, en ambas situaciones la probabilidad de A fue afectada por la ocurrencia del suceso B o por la ocurrencia de C. Sin embargo, al considerar el suceso $D = \{(x,y)/\, y$ *número par*$\}$ con P(D) = $\frac{1}{2}$, se establece que P(D/B) = 3/6 = P(D), pues de los 6 pares ordenados que satisfacen B, sólo (2,2), (4,4) y (6,6) cumplen con la condición que la segunda componente sea par, resultando que la probabilidad de D no es afectada por la ocurrencia del suceso B. En cambio P(D/C) = 9/15 $\neq$ P(D), verificándose que la probabilidad de D es afectada por la ocurrencia de C.

Las situaciones anteriores que resultaron notables dan origen a la siguiente definición.

Definición.

Se dice que dos sucesos A y B asociados a un espacio muestral S son **sucesos independientes** si y sólo si P(A / B) = P(A) y P(B / A) = P(B).

La condición de independencia entre dos sucesos establece que la ocurrencia de uno de ellos no altera la probabilidad de ocurrencia del otro. La condición de independencia da origen a una importante consecuencia.

Principio multiplicativo de probabilidades para sucesos independientes.

Del principio multiplicativo general se tiene que P(A $\cap$ B) = P(A / B)$*$P(B), pero si A y B son sucesos independientes, entonces por definición P(A / B) = P(A), que al sustituirse en la igualdad anterior resulta

Principio multiplicativo para sucesos independientes
A y B sucesos independientes $\Leftrightarrow$ *P(A* $\cap$ *B)* $= P(A)*P(B)$

El principio anterior se puede aplicar en dos direcciones. La más frecuente ocurre cuando mediante un simple racionamiento basado en las condiciones en las que se realiza el experimento permite deducir que dos sucesos son independientes, entonces se aplica $P(A \cap B) = P(A)*P(B)$. La otra ocurre cuando es díficil establecer a priori que dos sucesos son independientes, entonces si se puede establecer que $P(A \cap B) = P(A)*P(B)$, se deduce que A y B son sucesos independientes.

Ejemplos 5.3

a) Del enunciado del problema 3.2 b) no es posible establecer a priori si las enfermedades X e Y son o no independientes, pero considerando la información se puede establecer que $P(X \cap Y) = 2/35 = \frac{1}{5}*\frac{2}{7} = P(X)*P(Y)$, consecuentemente el que una planta tenga la enfermedad X es independiente a que contraiga la enfermedad Y y viceversa. Dicho de otra manera el que una planta tenga una enfermedad no afecta el que contraiga la otra.

b) Del enunciado del ejemplo 5.1 a) no es posible deducir si existe independencia entre los sucesos A y B, pero con la información entregada se establece que:
$P(A \cap B) = 1/6 \neq \frac{2}{5}*\frac{2}{3} = P(A)*P(B)$, luego los sucesos no son independientes.

c) El mecanismo que acciona una línea de embalaje en una exportadora depende de dos subsistemas independientes, A y B, con probabilidades de falla de $1/10$ y $1/15$, respectivamente, durante un día cualquiera. La línea deja de funcionar si fallan simultáneamente ambos subsistemas. Entonces, la probabilidad de que en un día cualquiera:
i) la línea se detenga. Para que esto ocurra deben fallar ambos subsistemas, que corresponde a $P(A \cap B) = P(A)*P(B) = \frac{1}{10}*\frac{1}{15} = 1/150$.
ii) falle sólo el subsistema A, que se calcula:
$P(A \cap B') = P(A) - P(A \cap B) = \frac{1}{10} - \frac{1}{150} = 14/150$
iii) la línea funcione, lo que ocurrirá si al menos un subsistema funcione, esto es, $P(A' \cup B') = 1 - P(A \cap B) = 1 - P(A)*P(B) = 1 - \frac{1}{150} = 149/150$. En este caso se aplicó la propiedad "probabilidad de que al menos uno *no falle*, es igual a uno menos la probabilidad de que *ambos fallen*".

d) Si se lanzan dos dados legales los resultados de ambos dados son independientes, luego
i) P(seis y seis) $= P$(seis)$*P$(seis) $= \frac{1}{6}*\frac{1}{6} = 1/36$
ii) P(exactamente un seis) $= P$(seis , no seis) $+ P$(no seis, seis) $= \frac{1}{6}*\frac{5}{6} + \frac{5}{6}*\frac{1}{6} = 5/18$
iii) P(al menos un seis) $= 1 - P$(ningún seis) $= 1 - \frac{5}{6}*\frac{5}{6} = 11/36$
iv) P(un par) $= 6*P$(un par específico)$=6*\frac{1}{6}*\frac{1}{6} = 1/6$, pues existen 6 pares posibles (1,1),...(6,6).

e) Se lanza un dado cuatro veces y se observa el número de ocurrencias del seis. El espacio muestral para este experimento es $S = \{0, 1, 2, 3, 4\}$, que corresponde a las veces que puede ocurrir el seis en los cuatro lanzamientos. Nótese que este espacio muestral no es equiprobable, así $P(\{0\})$ significa que ninguna vez ocurra el seis, o sea, P(no seis, no seis, no seis, no seis) $= \frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6} = 125/1296$, porque los lanzamientos son

independientes y la probabilidad de *no seis* cada vez es $5/6$. $P(\{1\})$ indica que uno de los lanzamientos muestre seis y los otros tres muestre cualquier valor distinto de seis, luego P(seis y no seis y no seis y no seis), que puede ocurrir de 4 maneras distintas, es decir, $\binom{4}{1}$ maneras, por lo tanto, P(seis y no seis y no seis y no seis) $= 4*\frac{1}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6} = 125/324$. La $P(\{2\}) = \binom{4}{2}*\frac{1}{6}*\frac{1}{6}*\frac{5}{6}*\frac{5}{6} = 25/216$, pues existen $\binom{4}{2}$ = 6 formas de ordenar dos veces el seis en cuatro lanzamientos. De esa manera se sigue calculando la probabilidad para los otros elementos, 3 y 4, del espacio muestral. Realice los cálculos y verifique que la suma de todas las probabilidades es igual a 1.

f) En una cámara de frío hay 1 bins de manzanas Granny , 1 de manzanas Richard y otro de manzanas Fuji, todas de igual apariencia. Se sabe que la probabilidad que una manzana tenga polilla es de 0,05 si es de la variedad Granny, 0,10 si es de la variedad Richard y 0,03 si es de la variedad Fuji. Entonces al elegir una manzana al azar de cada bin

i) P(las tres sanas) $= P$(sana/Gr)$*P$(sana/Ri)$*P$(sana/Fu) $= 0,95*0,90*0,97 = 0,829$

ii) P(dos sanas y una dañada) $= P$(S y S y D)

$= P$(S/Gr)$*P$(S/Ri)$*P$(D/Fu) $+ P$(S/Gr)$*P$(D/Ri)$*P$(S/Fu) $+ P$(D/Gr)$*P$(S/Ri)$*P$(S/Fu)

$= 0,95*0,90*0,03 + 0,95*0,10*0,97 + 0,05*0,90*0,97 = 0,161$

g) Una bolsita A contiene dos semillas de flores rojas y tres de flores blancas y otra B contiene tres semillas de flores rojas y tres de flores blancas. Se extraen, sin sustitución, dos semillas de cada bolsita. Dada la independencia del contenido de ambas bolsitas se puede calcular:

i) P(todas sean de flores de igual color) $= P$(2 rs de A y 2 rs de B) $+ P$(2 bls de A y 2 bls de B)

$= P$(2rs/A)$*P$(2rs/B) $+ P$(2bls/A)$*P$(2bls/B) $= \frac{\binom{2}{2}}{\binom{5}{2}}*\frac{\binom{3}{2}}{\binom{6}{2}} + \frac{\binom{3}{2}}{\binom{5}{2}}*\frac{\binom{3}{2}}{\binom{6}{2}} = 4/50.$

ii) P(sean 2 de cada color) $= P$(2rs/A)$*P$(2bls/B) $+ P$(r y b/A)$*P$(r y b/B) $+ P$(2bls/A)$*P$(2rs/B)

$= \frac{\binom{2}{2}}{\binom{5}{2}}*\frac{\binom{3}{2}}{\binom{6}{2}} + \frac{\binom{2}{1}*\binom{3}{1}}{\binom{5}{2}}*\frac{\binom{3}{1}*\binom{3}{1}}{\binom{6}{2}} + \frac{\binom{3}{2}}{\binom{5}{2}}*\frac{\binom{3}{2}}{\binom{6}{2}} = 11/25.$

2.6 Teorema de la probabilidad total y teorema de Bayes.

Muchas veces la probabilidad de un suceso es difícil obtenerla directamente, pero puede lograrse a partir de la probabilidad de ocurrencia de una serie de sucesos, lo que conduce a lo que se denomina *probabilidad total*. Previamente es necesario recordar el concepto de partición.

Definición.

Se llama **partición** de un espacio muestral S a una serie de k sucesos B_i que cumplan las siguientes condiciones:

1º $B_i \neq \phi$, para todo i = 1, 2, 3, ...,k

2º $B_i \cap B_j = \phi$, si i $\neq$ j

3º $B_1 \cup B_2 \cup B_3 \cup ..... \cup B_k = S$.

La definición establece que los sucesos son no vacíos y excluyentes entre ellos, es decir, no tienen elementos en común y además, son exhaustivos, pues entre todos completan el espacio muestral. Un rompecabezas es una partición, donde cada pieza es un subconjunto del cuadro completo, o sea, un suceso desde el punto de vista probabilístico.

Teorema de la probabilidad total.

Sea $A \subset S$ y { B_i / i = 1, 2, 3, ..., k} una partición de S, la cual induce la partición { $A \cap B_i$ / i = 1, 2, 3, ..., k} en el suceso A, tal que:
1º $(A \cap B_i) \cap (A \cap B_j) = \phi$, si i ≠ j.
2º $(A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup ..... \cup (A \cap B_k) = A$, entonces si son conocidas las $P(A/B_i)$ y $P(B_i)$ para cada i = 1, 2, 3,..., k , se puede establecer que $P(A) = \sum_{i=1}^{k} P(A/B_i) * P(B_i)$, $P(B_i) > 0$.

Demostración.

$P(A) = P((A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup ..... \cup (A \cap B_k)) = \sum_{i=1}^{k} P(A \cap B_i)$, pues se trata de una unión de sucesos mutuamente excluyentes por la condición 1º de partición de A. Pero para elemento $A \cap B_i$ de la partición de A se cumple que $P(A \cap B_i) = P(A/B_i) * P(B_i)$, de acuerdo al principio multiplicativo general de probabilidades. Por lo tanto, sustituyendo en la sumatoria anterior se cumple que $P(A) = \sum_{i=1}^{k} P(A/B_i) * P(B_i)$.

Observaciones.

1) Siguiendo con la analogía del rompecabezas, si el suceso A a que hace referencia el teorema lo asimilamos a la figura central de éste, se tendrá que *algunas de las piezas* contienen parte de la figura central, no importa que la mayoría de las piezas no contribuyan a su formación, lo que equivale a decir que algunas $A \cap B_i$ son vacías, lo fundamental es que al armar el rompecabezas completo la figura central quedará completa.
2) Otra situación se da al considerar un huerto de manzanos donde el 60% de la producción es de la variedad Granny Smith, el 30% de la variedad Fuji y el 10% de la variedad Royal, entonces las tres variedades de manzanas establecen una partición del suceso A = {manzanas calibre 100} correspondientes a {manzanas Granny calibre 100} , {manzanas Fuji calibre 100} y {manzanas Royal calibre 100}.

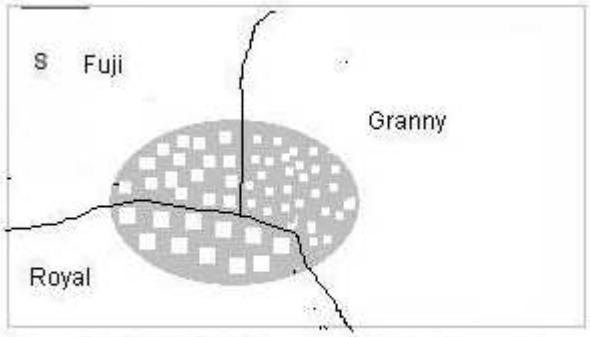


Figura 6.1. Partición del suceso A inducida por la partición de S del ejemplo 6.1 a).

Ejemplos 6.1

a) El 60% de la producción de un huerto de manzanos es de la variedad Granny Smith, el 30% de la variedad Fuji y el 10% de la variedad Royal, y se sabe que son calibre 100 el 15% de las manzanas Granny, el 35% de las Fuji y el 40% de las Royal. Entonces el porcentaje de manzanas calibre 100 de la producción total del huerto se calcula usando el teorema de la probabilidad total, donde A = {manzanas calibre 100}, como

$P(\text{A}) = P(\text{A}/\text{G})*P(\text{G}) + P(\text{A}/\text{F})*P(\text{F}) + P(\text{A}/\text{R})*P(\text{R}) = 0,15*0,6 + 0,35*0,3 + 0,40*0,1$
$= 0,235$, es decir, el 23,5% del total de manzanas es calibre 100.

Téngase en cuenta que la partición la establecen las variedades y por lo tanto la suma de sus probabilidades debe ser 1, sin embargo las probabilidades condicionales P(A/G), P(A/F) y P(A/R) no tienen por qué sumar 1, pues están referidas respecto a cada variedad.

b) Una mezcla de semillas de clavel produce flores blancas, rojas y rosadas en proporción de 50%, 30% y 20% respectivamente. El 5% de las semillas de flores blancas, el 10% de las rojas y el 15% de las rosadas son infértiles (F'). Se desea determinar el porcentaje total de semillas infértiles. Las condiciones del enunciado se disponen adecuadamente a continuación:

$P(\text{flor blanca}) = P(\text{b}) = 0,5$ ..... $P(\text{F' / b}) = 0,05$
$P(\text{flor roja}) = P(\text{r}) = 0,3$ ..... $P(\text{F ' / r}) = 0,10$
$P(\text{flor rosada}) = P(\text{s}) = 0,2$ ..... $P(\text{F' / s}) = 0,15$

Por el teorema de la probabilidad total $P(\text{F'}) = P(\text{F'/ b})*P(\text{b}) + P(\text{F'/ r})*P(\text{r}) + P(\text{F'/ b})*P(\text{b})$

$= 0,05*0,5 + 0,10*0,3 + 0,15*0,2 = 0,085.$

El resultado obtenido permite establecer que el 8,5% del total de semillas son infértiles.

Teorema de Bayes.

En el caso del ejemplo 6.1 b) se puede estar interesado en determinar la probabilidad de que una semilla que resultó ser infértil corresponda a una de flor roja.

En símbolos P(r /F') = $\frac{\text{P(roja e infértil)}}{\text{P(infértil)}}$ = $\frac{\text{P(F' /r)}*\text{P(r)}}{\text{P(F')}}$ = $\frac{0,10*0,3}{0,085}$ = 6/17, por la definición de probabilidad condicional y el principio multiplicativo general de probabilidades. Esta forma de resolver el problema se debe al Rev. Thomas Bayes.

Formalmente, conocidas las probabilidades $P(A/B_i)$ y $P(B_i)$ para todo i, entonces la probabilidad $P(B_j/A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A/B_j)*P(B_j)}{P(A)} = \frac{P(A/B_j)*P(B_j)}{\sum_{i=1}^{k} P(A/B_i)*P(B_i)}$, para j = 1, 2, 3,..., k.

Las explicaciones son las del párrafo anterior, reconociendo, además, que P(A) es la probabilidad total.

Ejemplos 6.2

a) La efectividad de un producto para controlar pudriciones en peras es de 0,80 si el hongo es Botrytis y 0,60 si el hongo es Penicillium. Se estima que el 30% de los frutos está infectado por Botrytis, el 10% está infectado por Penicillium y el resto está sano. Entonces el porcentaje de frutos que se espera que presenten pudriciones después de aplicar el producto se obtiene con la información a continuación, aplicando la probabilidad total.

$P(\text{efectiv /Bot}) = 0,80$ ..... $P(\text{Bot}) = 0,3$
$P(\text{efect /Pen}) = 0,60$ ..... $P(\text{Pen}) = 0,1$
$P(\text{efect /sano}) = 1,0$ ..... $P(\text{sano}) = 0,6$, por lo tanto
$P(\text{sin pudr}) = P(\text{efect /Bot})*P(\text{Bot}) + P(\text{efect /Pen})*P(\text{Pen}) + P(\text{efect /sano})*P(\text{sano})$
$= 0,8*0,3 + 0,6*0,1 + 1,0*0,6 = 0,90$, es decir, el 90% de los frutos estará sano y en consecuencia el 10% presentará pudriciones.

Podría interesar establecer la probabilidad de que un fruto haya estado infectado por Penicillium si está sano después de aplicar el producto. En este caso hay que aplicar el teorema de Bayes:

$P(\text{Pen / sano}) = \frac{P(\text{sano/Pen}) * P(\text{Pen})}{P(\text{sano})} = \frac{0,6*0,1}{0,9} = 1/15.$

b) En la situación del problema 6.1 b) se seleccionan 250 semillas de la mezcla y se siembran. Es necesario saber la proporción de flores de cada color que se obtendrán. Esta situación se resuelve aplicando sucesivamente el teorema de Bayes, pues se necesita P(b /fértil), P(r /fértil) y P(s /fértil). La información a utilizar es

$P(\text{flor blanca}) = P(\text{b}) = 0,5$ ..... $P(\text{F' / b}) = 0,05$ .....$P(\text{F /b}) = 0,95$

$P(\text{flor roja}) = P(\text{r}) = 0,3$ ..... $P(\text{F '/ r}) = 0,10$ .....$P(\text{F /r}) = 0,90$

$P(\text{flor rosada}) = P(\text{s}) = 0,2$ ..... $P(\text{F' / s}) = 0,15$ .....$P(\text{F /s}) = 0,85$

$P(\text{F}) = 1 - P(\text{F'}) = 1 - 0,085 = 0,915$

$P(\text{b /F}) = \frac{P(\text{F /b}) * P(\text{b})}{P(\text{F})} = \frac{0,95*0,5}{0,915} = 95/183 = 51,9\%$

$P(\text{r /F}) = \frac{P(\text{F /r}) * P(\text{r})}{P(\text{F})} = \frac{0,90*0,3}{0,915} = 54/183 = 29,5\%$

$P(\text{s /F}) = \frac{P(\text{F /s}) * P(\text{s})}{P(\text{F})} = \frac{0,85*0,2}{0,915} = 34/183 = 18,6\%$, es decir, 51,9% serán flores blancas, 29,5% serán rojas y 18,6% serán flores rosadas. Estas probabilidades reciben el nombre de *probabilidades a posteriori*.

c) Una empresa M considera que su rival la empresa W tiene una probabilidad de 0,6 de presentarse a una licitación. Si W se presenta la probabilidad de que M gane (G) la licitación es 0,2, mientras que si W no se presenta (W') la probabilidad de ganarla es 0,9. A la empresa M le interesa conocer sus posibilidades de hacerse con la licitación. Esto corresponde a la probabilidad total

$P(\text{G}) = P(\text{G / W}) * P(\text{W}) + P(\text{G/ W' }) * P(\text{W' }) = 0,2*0,6 + 0,9*0,4 = 0,48$, es decir, M tiene una probabilidad de 48% de ganar la licitación.

Un financista tiene la curiosidad de saber la probabilidad de que W no se haya presentado a la licitación si M ganó la licitación, entonces

$P(\text{W' /G}) = \frac{P(\text{G / W' }) * P(\text{W' })}{P(\text{G})} = \frac{0,9*0,4}{0,48} = 3/4$. Esta probabilidad es de 75%.

3. DISTRIBUCIONES DE PROBABILIDAD

3.1 Introducción

Recuérdese que un modelo matemático es la descripción matemática de una situación real en cuya elaboración se hacen algunos supuestos y en el que se consideran algunas simplificaciones de la realidad. La bondad de un modelo depende de cuán bien se aproxima a la realidad que pretende describir y además de cuán simple sea. En síntesis un modelo es una forma matemática de describir el *comportamiento* de un fenómeno.

Los fenómenos determinísticos, como lo son por ejemplo, los físicos de la **cinemática**, la **energía** , la **óptica**, la **termodinámica** o en la química inorgánica como sucede con compuestos y sustancias de gran importancia biológica tales como los **fertilizantes** o los **pesticidas**, son descritos mediante modelos determinísticos. Estos modelos se traducen en fórmulas que establecen las interrelaciones entre los factores que intervienen en el fenómeno, mediante la cual se puede determinar con **certeza** el comportamiento de éste si se conocen las condiciones en que actúa un número determinado de los factores. Por ejemplo, se puede predecir con certeza la distancia recorrida por un móvil si se conocen las condiciones en que se realiza el movimiento.

Por el contrario, en los fenómenos no determinísticos, como lo son todos los **juegos de azar** y también innumerables fenómenos naturales, como los **climáticos,** la **producción** de frutales o de cultivos, no se pueden predecir con **certeza** el resultado. En consecuencia la única manera de describirlos es a través de su comportamiento probabilístico mediante modelos estocásticos. Para comprender estos modelos se requiere conocer una serie de términos, notaciones y conceptos que les son propios y que se desarrollarán en esta unidad.

3.2 Distribuciones de variable aleatoria.

El concepto básico en el que se sustenta toda la teoría de las distribuciones de probabilidad cuyo objetivo es formular los modelos estocásticos en términos puramente matemáticos, es el de *variable aleatoria*.

Definición.

Se llama **variable aleatoria** (v.a) a una función X cuyo dominio es el espacio muestral S y con recorrido en los reales, tal que a cada elemento del espacio muestral le asigna una imagen en los números reales.

En términos matemáticos: $X: S \rightarrow \Re$, tal que $\forall s \in S \Rightarrow X(s) \in \Re$.

Observaciones.

1) Se conviene en designar las variables aleatorias por letras mayúsculas X, Y, Z,.....

2) El recorrido de una variable aleatoria, R_X, está formado por todas las imágenes de X en $\Re$. Conceptualmente es otro espacio muestral del experimento. Este nuevo espacio muestral generalmente **no es equiprobable**, aunque S si lo sea.

Los siguientes ejemplos servirán para clarificar el concepto.

Ejemplos 2.1.

a) ε_1**: lanzamiento de una moneda**, con espacio muestral S = $\{s, c\}$, y sea la v.a X_1 tal que, $X_1(s) = 1$, $X_1(c) = 2$ con $R_{X_1} = \{1, 2\}$, es decir, la función X transforma al resultado *sello* en el real 1 y *cara* en el real 2.

Observe que la definición de variable aleatoria no impone ninguna restricción respecto al número real que se asigne, ni tampoco en que los valores asignados tengan alguna interpretación, aunque lo habitual es que si la tenga, como se ilustra en el siguiente caso:
Sea X_2: nº de sellos obtenidos al lanzar una moneda. De acuerdo a esta definición de X_2 $X_2(s) = 1$, $X_2(c) = 0$ con $R_{X_2} = \{0, 1\}$ que se explica en el sentido que si al lanzar la moneda ocurre *sello* el número de sellos obtenidos es *uno*, mientras que si ocurre *cara* el número de sellos obtenidos es *cero.*

Ambas variables aleatorias, X_1 y X_2, son conceptualmente correctas.

b) ε_2 **: lanzamiento de dos monedas** , con $S = \{(c, c), (c, s), (s, c), (s, s)\}$. Si X: nº de sellos obtenidos con ε_2, entonces $X(c, c) = 0$, $X(s, c) = X(c, s) = 1$, $X(s, s) = 2$ y $R_X = \{0, 1, 2\}$.

c) ε_3 **: lanzamiento de un dado** , con espacio muestral S = $\{1, 2, 3, 4, 5, 6\}$. Si X_1: puntos obtenidos con ε_3, entonces $X_1(s_i) = s_i$, $\forall\, s_i \in S$ y por lo tanto $R_{X_1} = \{1, 2, 3, 4, 5, 6\}$ resulta igual a S, pues X_1 es la función identidad. Si para este mismo experimento se define X_2: nº de seis obtenidos con ε_3, entonces $X_2(1) = X_2(2) = X_2(3) = X_2(4) = X_2(5) = 0$, mientras que $X_2(6) = 1$, luego $R_{X_2} = \{0, 1\}$. Otra posible variable aleatoria en este experimento es $X_3 = \begin{cases} 1 \text{ si el valor es par} \\ 2 \text{ si el valor es impar} \end{cases}$, con $R_{X_3} = \{1, 2\}$.

d) ε_4 **: se lanzan dos dados y se observan los valores obtenidos**. Si se define X_1 : suma de puntos obtenidos, entonces $R_{X_1} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, mientras que si la variable aleatoria es X_2 : nº de ases obtenidos , entonces $R_{X_2} = \{0, 1, 2\}$.

Para continuar con el desarrollo del modelo se debe tener una función que le asigne probabilidades a los elementos de R_X. Para ello, hay que distinguir entre variables aleatorias discretas y continuas.

Distribuciones de variables aleatorias discretas (v.a.d).

Una variable aleatoria es discreta si R_X es un conjunto finito ó infinito numerable. Todos los ejemplos 2.1 corresponden a este tipo de variable.

Definición.

Sea X variable aleatoria discreta, entonces una función *p*, denominada función de probabilidad puntual (f.p.p) ó de cuantía, que le asigne probabilidades a los elementos x_i de R_X , debe satisfacer las siguientes condiciones:

1º) $p(x_i) \geq 0$, $\forall\, x_i \in R_X$

2º) $\sum_{x_i \in R_X} p(x_i) = 1$

Ejemplos 2.2

a) Sea X variable aleatoria discreta con

$p(x_i) = \begin{cases} 1/2 & \text{si } x_i = 2 \\ 1/3 & \text{si } x_i = 3 \\ 1/6 & \text{si } x_i = 6 \end{cases}$, entonces *p* es una correcta función de probabilidad puntual en R_X = { 2 , 3 , 6 } porque sus imágenes son no negativas y su suma es igual a 1.

b) La distribución $p(x_i) = \frac{1}{6}, \forall\, x_i \in \{1, 2, 3, 4, 5, 6\}$ para la variable aleatoria X_1 del experimento ε_3 del ejemplo 2.1, constituye una correcta función de probabilidad puntual. En este caso se establece que el espacio R_{X_1} es equiprobable lo que ocurre si el dado es simétrico. Si el dado estuviese cargado entonces la función *p* indicaría diferentes valores para cada x_i. En el mismo experimento la variable aleatoria X_2 tiene por función de cuantía $p(x_i) = \begin{cases} 5/6 & \text{si } x_i = 0 \\ 1/6 & \text{si } x_i = 1 \end{cases}$.

c) En ε_4 la variable aleatoria X_2 tiene función de cuantía $p(x_i) = \begin{cases} 25/36 & \text{si } x_i = 0 \\ 10/36 & \text{si } x_i = 1 \\ 1/36 & \text{si } x_i = 2 \end{cases}$

Distribuciones de variables aleatorias continuas (v.a.c).

Una variable aleatoria es continua si el conjunto R_X es un conjunto infinito no numerable. En este tipo de variable aleatoria el conjunto R_X corresponde a un intervalo o a una unión de intervalos de números reales. Así si ε consiste en medir la cantidad de agua lluvia caida en Quinta Normal durante un año dado, habría que establecer, por ejemplo, $R_X = \{h/0 \leq h \leq 1000\}$, donde h es la altura en mm., o más simplemente, como se adoptará en lo sucesivo, R_X será el conjunto de los reales, $\Re$. Tenga en cuenta que el espacio muestral no tiene por qué estar ajustado a lo que realmente suceda, pues lo importante es que no deje fuera valores posibles, y $\Re$ cumple con ser el conjunto R_X más amplio posible.

Definición.

Sea X variable aleatoria continua, entonces una función *f*, denominada función de densidad de probabilidad (f.d.p), que asigne probabilidades en $\Re$, debe satisfacer las siguientes condiciones:

1º) $f(x) \geq 0$, $\forall\, x \in \Re$

2º) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

3º) $P(a \leq X \leq b) = \int_a^b f(x) dx$

Observaciones.

1) La definición anterior establece que una función que asigne probabilidades a una variable aleatoria continua debe ser **no negativa**.

2) Las probabilidades se asignan en términos de área bajo la curva cuya función es *f*, por esta razón la segunda condición establece que el área total es uno, porque corresponde a la probabilidad del espacio muestral.

3) La tercera condición dice que la probabilidad del *suceso* definido por el intervalo [*a , b*] la determina el área limitada por la recta $x = a$, la curva $f(x)$, la recta $x = b$ y el eje OX.

4) De la definición se establece que las probabilidades puntuales, es decir en un punto, tienen el valor *cero*, pues el área bajo una curva en un punto es nula, luego, $P(X = c) = 0$. Esta situación es intuitivamente correcta, porque cualquier intervalo contiene infinitos puntos y si cada uno tuviera probabilidad superior a cero, entonces la probabilidad del intervalo superaría al valor 1. Se debe tener presente que la probabilidad es del intervalo y no de los puntos que están en él. Como consecuencia de la definición se concluye que para **variables aleatorias continuas**

$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)dx$

Algunos ejemplos ayudarán a comprender mejor estos conceptos que son válidos por su sencillez y no necesariamente por su interpretación a alguna situación real.

Ejemplos 2.3.

a) Sea X variable aleatoria continua con $f(x) = \begin{cases} 1 - \frac{1}{2}x & \text{si } 0 \leq x \leq 2 \\ 0 & \text{para otros valores} \end{cases}$

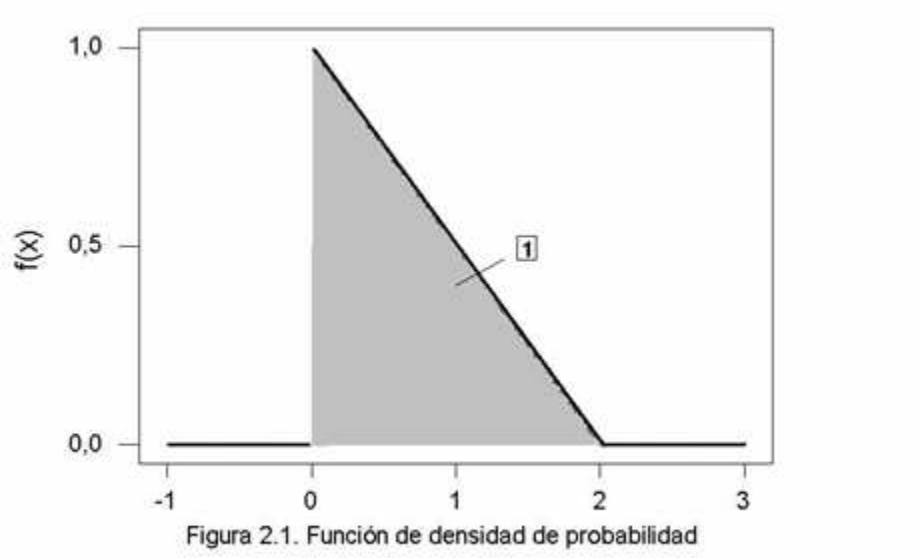


Figura 2.1. Función de densidad de probabilidad

La figura 2.1 muestra el comportamiento de esta variable aleatoria, que coincide con OX en los negativos y en el intervalo] 2 , +∞[, y que en [0 , 2] es un segmento de recta ubicada por arriba del eje OX, por lo que la variable toma valores en este último intervalo, formando un triángulo rectángulo con el eje coordenado cuya área es igual a uno, cumpliéndose que el área total bajo $f(x)$ es la unidad. También se aprecia que asigna probabilidades mayores a intervalos cercanos al cero y probabilidades decrecientes a intervalos cercanos al dos. Algunos cálculos de probabilidades asociadas a esta variable se desarrollan a continuación.

- $P(\frac{1}{2} \leq X \leq \frac{3}{2}) = \int_{\frac{1}{2}}^{\frac{3}{2}} (1 - \frac{1}{2}x)\, dx = 1/2$

- $P(X < 1) = \int_{-\infty}^{1} f(x)\, dx = \int_{-\infty}^{0} 0 * dx + \int_0^1 (1 - \frac{1}{2}x)\, dx = 0 + 3/4 = 3/4$, esto se explica porque la función vale cero en los negativos y por lo tanto el área es nula.

- $P(X \geq 2/3) = \int_{\frac{2}{3}}^{\infty} f(x)\, dx = \int_{\frac{2}{3}}^{2} (1 - \frac{1}{2}x)\, dx + \int_2^{\infty} 0 * dx = 4/9$, pues la función toma el valor cero en el intervalo]2 , ∞[.

- $P(1 \leq X \leq \frac{5}{2}) = \int_1^2 (1 - \frac{1}{2}x)\, dx = 1/4$

b) Sea la variable aleatoria continua X con $f(x) = \begin{cases} x^2/3 & \text{si} \quad -2 \leq x \leq 1 \\ 0 & \text{para otros valores} \end{cases}$

La función de densidad de probabilidad es un arco de parábola positiva con vértice en O en el intervalo [-2 , 1] y coincide con OX fuera del intervalo (figura 2.2). El área total bajo la curva corresponde a $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{-2} 0 * dx + \int_{-2}^{1} \frac{x^2}{3}\,dx + \int_{1}^{\infty} 0 * dx = 0 + 1 + 0 = 1.$

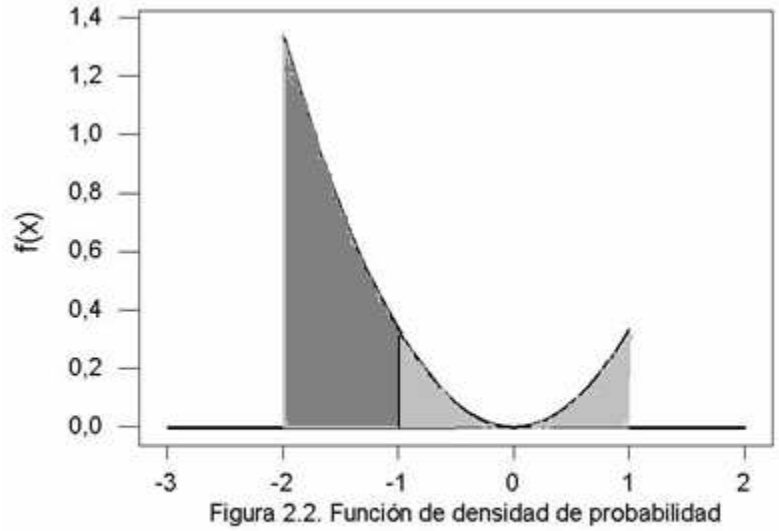


Figura 2.2. Función de densidad de probabilidad

Algunas probabilidades asociadas a esta distribución se calculan a continuación.

- $P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{x^2}{3}\,dx = 1/36$
- $P(X < 0) = \int_{-2}^{0} \frac{x^2}{3}\,dx = 8/9$
- $P(X > -1) = \int_{-1}^{\infty} f(x)\,dx = \int_{-1}^{1} \frac{x^2}{3}\,dx + \int_{1}^{\infty} 0 * dx = \frac{2}{9} + 0 = 2/9$, representada por el área más clara en la figura 2.2.

Función de distribución acumulativa (f.d.a).

Otra forma de expresar el comportamiento de una variable aleatoria consiste en hacerlo mediante su distribución acumulativa, la que aporta una serie de ventajas, en especial para las distribuciones notables, por razones que se van a explicar posteriormente.

Definición.

Se llama función de distribución acumulativa, de una variable aleatoria X, discreta o continua, a una función F tal que $F(x) = P(X \leq x)$.

Tal como lo da a entender su nombre esta es una función que va acumulando probabilidades. En el caso discreto lo hace sumando probabilidades punto a punto, similar a la frecuencia relativa acumulada H_i en descriptiva. En el caso continuo corresponde al área total bajo la curva desde -∞ hasta el punto x en el eje real. Formalmente:

1. Si X variable aleatoria discreta $F(x) = \sum_{x_i \leq x} p(x_i)$
2. Si X variable aleatoria continua $F(x) = \int_{-\infty}^{x} f(x)\,dx$

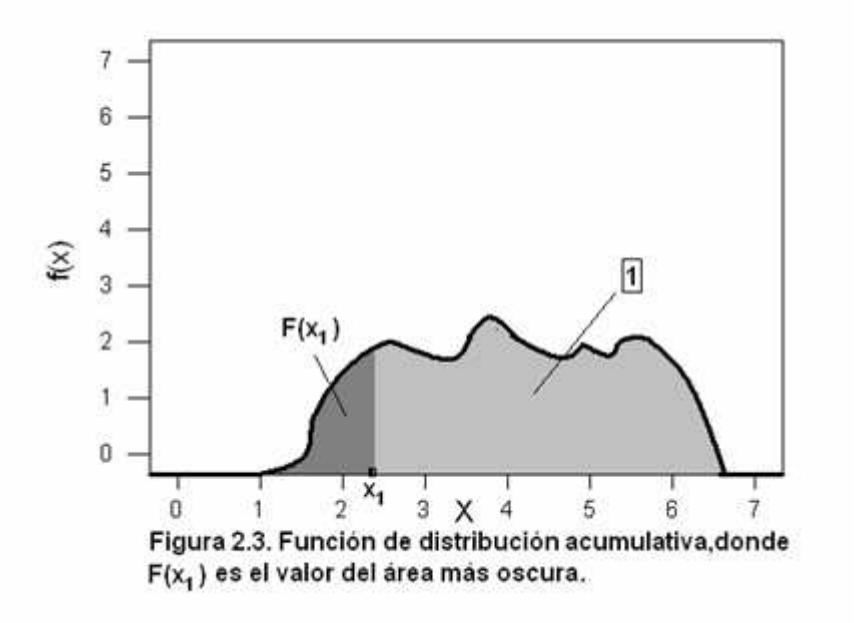


Figura 2.3. Función de distribución acumulativa,donde $F(x_1)$ es el valor del área más oscura.

La figura 2.3 ilustra el concepto de función de distribución acumulativa en el caso de una variable aleatoria continua *X*. La interpretación es que en la medida que el punto x_1 avanza hacia la derecha el área bajo la curva se va incrementando y por tanto el valor de $F(x_1)$, es decir, aumenta la probabilidad de que ocurra un valor de *X* **menor o igual** que x_1, hasta alcanzar el valor 1 cuando x_1 llegue al final del recorrido.

Propiedades.

1° F es una función **no decreciente**, esto es, si $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$.

2° $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to \infty} F(x) = 1$, es decir, $F(x)$ varía entre 0 y 1.

3° Si X variable aleatoria discreta, entonces F es una función escalonada, con saltos de altura $p(x_i)$ en cada punto $x_i \in R_X$ y con probabilidad puntual $P(X = x_i) = F(x_i) - F(x_{i-1})$.

4° Si X variable aleatoria continua, entonces F es una función continua en $\Re$ con $f(x) = \frac{d}{dx}(F(x))$, con probabilidades $P(a \leq X \leq b) = F(b) - F(a)$ y $P(X > b) = 1 - F(b)$.

Ejemplos 2.4.

a) Si X variable aleatoria con $p(x_i) = \begin{cases} 1/2 & \text{si } x_i = 2 \\ 1/3 & \text{si } x_i = 3 \\ 1/6 & \text{si } x_i = 6 \end{cases}$, entonces

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 2 \\ 1/2 & \text{si} \quad 2 \le x < 3 \\ 5/6 & \text{si} \quad 3 \le x < 6 \\ 1 & \text{si} \quad x \ge 6 \end{cases}$$

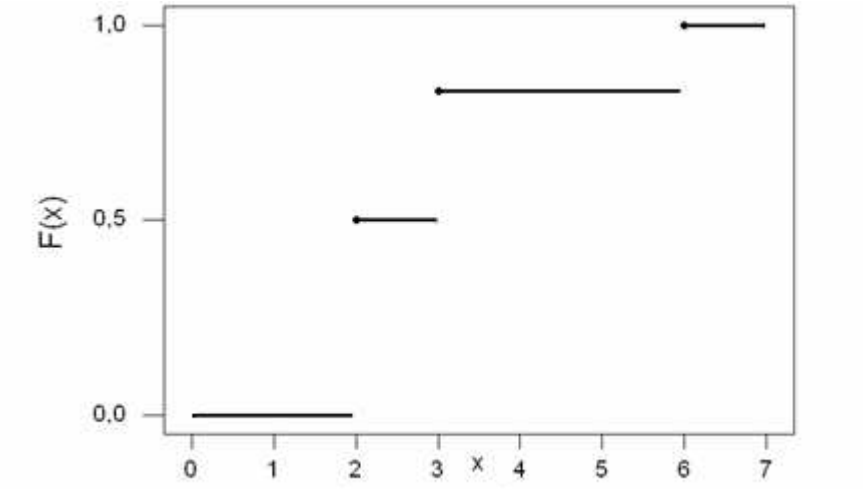


Figura 2.4. Función de distribución acumulativa de X variable aleatoria discreta.

b) Si X variable aleatoria con $f(x) = \begin{cases} x^2/3 & \text{si} \quad -2 \le x \le 1 \\ 0 & \text{para otros valores} \end{cases}$, entonces

$$F(x) = \begin{cases} 0 & \text{si} \quad x < -2 \\ \frac{x^3}{9}+\frac{8}{9} & \text{si} \quad -2 \le x \le 1 \\ 1 & \text{si} \quad x > 1 \end{cases}$$, pues $\int_{-\infty}^{x} f(x)\,dx = \int_{-2}^{x} \frac{x^2}{3}\,dx = \frac{x^3}{9}+\frac{8}{9}$

La función cuyo gráfico es el de la figura 2.5 no tiene área asociado, sino que valores sobre la curva, así en la gráfica *F*(-1) = 7/9 como se aprecia en la gráfica, valor que corresponde al área más oscura de la figura 2.2.

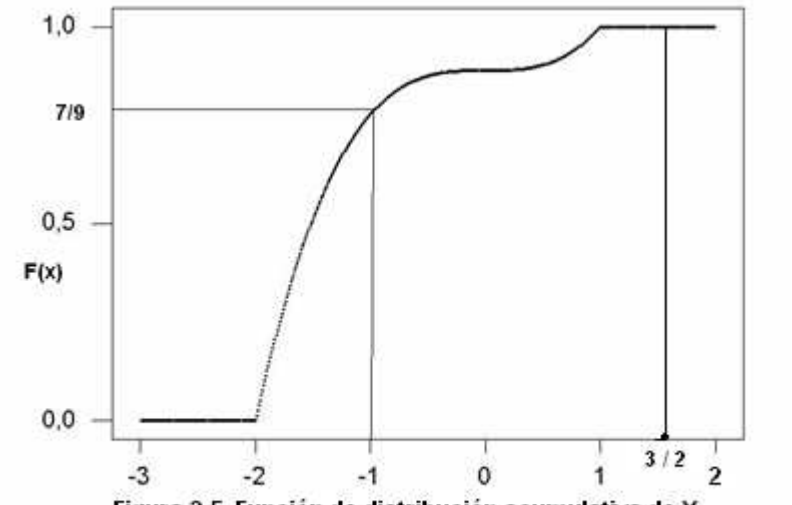


Figura 2.5. Función de distribución acumulativa de X variable aleatoria continua.

Como ejemplos del cálculo de probabilidades utilizando la función de distribución acumulativa F, se utilizarán los mismos casos del ejemplo 2.3. b), de modo que sirvan de comparación.

- $P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = F(\frac{1}{2}) - F(\text{-}\frac{1}{2}) = (\frac{1}{72}+\frac{8}{9}) - (\text{-}\frac{1}{72}+\frac{8}{9}) = 1/36$
- $P(X < 0) = F(0) = (0 + \frac{8}{9}) = 8/9$
- $P(X > -1) = 1 - P(X \leq -1) = 1 - F(-1) = 1 - (\frac{\text{-}1}{9} + \frac{8}{9}) = 2/9$
- $P(-1 \leq X \leq \frac{3}{2}) = F(\frac{3}{2}) - F(\text{-}1) = 1 - (\frac{\text{-}1}{9} + \frac{8}{9}) = 1 - 7/9 = 2/9$, pues $3/2$ está en el intervalo $x > 1$, así que $F(\frac{3}{2}) = 1$, como se ve en la figura 2.5.

3.3 Valores característicos de variables aleatorias.

Son valores que permiten resumir mediante un número ciertas características de una variable aleatoria. Muchas veces este valor característico coincide con el parámetro de la distribución. Los dos más importantes se refieren al *valor esperado* o *esperanza matemática* y el otro a la *varianza.*

Valor esperado de una variable aleatoria.

Definiciones.

1. Se llama **valor esperado** de una variable aleatoria discreta al número $E[X] = \sum_{x_i \in R_X} x_i * p(x_i)$.
2. Se llama **valor esperado** de una variable aleatoria continua al número $E[X] = \int_{-\infty}^{\infty} x * f(x) dx$.
3. Para cualquier función H de la variable aleatoria X, $E[H(X)] = \sum_{x_i \in R_X} H(x_i) * p(x_i)$ si X es variable aleatoria discreta, o $E[H(X)] = \int_{-\infty}^{\infty} H(x) * f(x)\, dx$ si X es variable aleatoria continua.

Ejemplos 3.1

a) Sea X: puntos obtenidos al lanzar un dado ("legal") con $p(x_i) = \frac{1}{6} \ \forall x_i \in \{1,2,3,4,5,6\}$, entonces $E[X] = 1*\frac{1}{6} + 2*\frac{1}{6} + 3*\frac{1}{6} + 4*\frac{1}{6} + 5*\frac{1}{6} + 6*\frac{1}{6} = \frac{1}{6}*(1+2+3+4+5+6) = 3,5$

b) Sea X variable aleatoria discreta con $p(x_i) = \begin{cases} \frac{4}{35} \text{ si } x_i = 0 \\ \frac{18}{35} \text{ si } x_i = 1 \\ \frac{12}{35} \text{ si } x_i = 2 \\ \frac{1}{35} \text{ si } x_i = 3 \end{cases}$, luego

$E[X] = 0*\frac{4}{35} + 1*\frac{18}{35} + 2*\frac{12}{35} + 3*\frac{1}{35} = \frac{9}{7}$

c) Sea X : suma de puntos al lanzar dos veces un dado legal, entonces la distribución de X es $p(x_i) = \begin{cases} 1/36 & \text{si } x_i = 2\,,\ 12 \\ 2/36 & \text{si } x_i = 3\,,\ 11 \\ 3/36 & \text{si } x_i = 4\,,\ 10 \\ 4/36 & \text{si } x_i = 5\,,\ 9 \\ 5/36 & \text{si } x_i = 6\,,\ 8 \\ 6/36 & \text{si } x_i = 7 \end{cases}$

La distribución especifica que la probabilidad de obtener una suma de 2 puntos es $1/36$ igual a la probabilidad de obtener 12 puntos o que obtener 5 o 9 puntos tienen ambas la misma probabilidad de $4/36$. Entonces el número esperado de puntos obtenidos es $E[X] = 2*\frac{1}{36} + 3*\frac{2}{36} + ... + 10*\frac{3}{36} + 11*\frac{2}{36} + 12*\frac{1}{36} = \frac{252}{36} = 7$

Observación.

De los tres ejemplos anteriores, especialmente en el a), es posible deducir que el valor esperado de una distribución es equivalente al "promedio" de los valores que esta variable aleatoria puede tomar, pero no como promedio simple de sus valores, sino como un **promedio ponderado** por su probabilidad $p(x_i)$. Esto es equivalente a pensar que en una tabla de frecuencia de variable discreta, $\mu = \frac{\sum f_i * X_i}{N} = \sum \frac{f_i}{N} * X_i = \sum h_i * X_i$, pues $h_i = f_i/N$ es la frecuencia relativa y ésta equivale a una *probabilidad empírica*. En este sentido, en los tres ejemplos, se puede interpretar que si se observa la variable aleatoria un número "infinito" de veces, entonces el promedio de los valores obtenidos es su valor esperado.

Ejemplos 3.2

a) Sea X variable aleatoria continua con $f(x) = \begin{cases} 2x & \text{si } 0 \le x \le 1 \\ 0 & \text{para otros valores} \end{cases}$

En la figura 3.1 se aprecia que la gráfica de esta distribución está representada por el segmento que une los puntos $(0,0)$ y $(1,2)$ y por el eje OX en el resto de los reales. De esta manera asigna probabilidades mayores a valores en intervalos cercanos a 1 y probabilidades pequeñas a intervalos cercanos al cero.

La interpretación de $E[X] = \int_{-\infty}^{\infty} x*f(x)\,dx = \int_0^1 x*2x\,dx = 2/3$, es que si se observa un número muy grande de veces el valor de la variable su "promedio" es $2/3$, lo cual es consistente porque sus valores son más cercanos al 1 que al cero, dentro del intervalo $[0,1]$.

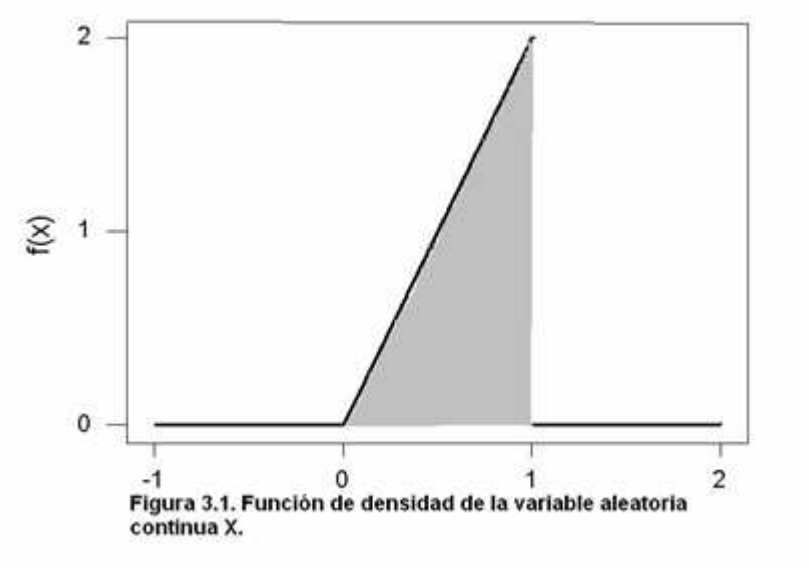


Figura 3.1. Función de densidad de la variable aleatoria continua X.

b) Sea la función de distribución de X

$$f(x) = \begin{cases} x^2/3 & \text{si } -2 \le x \le 1 \\ 0 & \text{para otros valores} \end{cases}, \text{ entonces}$$

$E[X] = \int_{-2}^{1} x * \frac{x^2}{3} dx = \int_{-2}^{1} \frac{x^3}{3} dx = -5/4$, o sea, su valor "promedio" es $-5/4$, valor consistente, porque de acuerdo al ejemplo 2.3 b) esta variable aleatoria toma valores negativos con una probabilidad de $8/9$.

Propiedades del valor esperado.

Las propiedades que se exponen a continuación son equivalentes a las establecidas para la media poblacional en la unidad de descriptiva.

1º $\boldsymbol{E[k] = k}$. El valor esperado de una constante es igual a la constante.
La propiedad es trivial, pues corresponde a la misma propiedad del promedio.

2º $\boldsymbol{E[cX] = c * E[X]}$. La propiedad establece que la constante que multiplica a la variable aleatoria multiplica al valor esperado.

3º $\boldsymbol{E[cX \pm k] = c * E[X] \pm k}$. Esta corresponde a la propiedad de linealidad del valor esperado e incluye a las dos primeras como casos especiales.

Demostración.

Por facilidad en la demostración se considerará a X como variable aleatoria continua, pero como la integral y la sumatoria tienen las mismas propiedades a utilizar en la demostración, también es válida para las variables aleatorias discretas.

$E[cX \pm k] = \int_{-\infty}^{\infty} (cx \pm k) * f(x)\, dx$, por la definición 3 de valor esperado

$$= \int_{-\infty}^{\infty} (\, cx * f(x) \pm k * f(x)\,)\, dx$$

$E\,[cX \pm k] = \int_{-\infty}^{\infty} cx*f(x)\,dx \pm \int_{-\infty}^{\infty} k*f(x)\,dx$

$= c*\int_{-\infty}^{\infty} x*f(x)\,dx \pm k*\int_{-\infty}^{\infty} f(x)\,dx$

$= c*E$[x] $\pm k$, pues la primera integral es el valor esperado de X y la segunda es igual a uno por definición.

4º Sean X e Y variables aleatorias cualesquiera, entonces $\boldsymbol{E[X \pm Y] = E[X] \pm E[Y]}$

5º Sean X e Y variables aleatorias cualesquiera, entonces $\boldsymbol{E[X*Y] \neq E[X]*E[Y]}$, salvo que X e Y sean **variables aleatorias independientes**.

Estas dos últimas propiedades se demuestran en la sección 4.3, ejemplo 4.2 y en la consecuencia 3 de variables aleatorias independientes.

Ejemplos 3.3

Como ejemplos se mostrarán algunas aplicaciones del valor esperado.

a) Una compañía aseguradora desea ofrecer un seguro agrícola anual para la producción de cerezas por un monto de 2500 UF. La compañía estima que puede tener que pagar el monto total con probabilidad 0,02 , el 50% del total con probabilidad 0,06 y un 25% del monto con probabilidad 0,1. ¿ Cuánto debe ser la prima anual que la compañía debe cobrar si desea tener una utilidad promedio de 50 UF anual por cada uno de estos seguros?

Sea X: la pérdida anual por cada siniestro de la compañía, cuya distribución de probalidad es

$p(x_i) = \begin{cases} 0,02 & \text{si } x_i = 2500 \\ 0,06 & \text{si } x_i = 1250 \\ 0,10 & \text{si } x_i = 625 \\ 0,82 & \text{si } x_i = 0 \end{cases}$

Por lo tanto $E\,[X]$ = $2500*0,02$+$1250*0,06$+$625*0,1$ = $187,5$ UF , es el monto promedio anual que debería pagar la compañía por cada seguro. Si desea tener una ganancia de 50 UF, entonces debería cobrar 237,5 UF, que corresponde a la pérdida más la utilidad esperada.

b) La función $p(x_i)$ representa la distribución de probabilidad de calidad de un productor de repollos $p(x_i) = \begin{cases} 1/6 & \text{si } x_i = 1, \text{ es decir primera} \\ 1/2 & \text{si } x_i = 2, \text{ es decir segunda} \\ 1/4 & \text{si } x_i = 3, \text{ es decir tercera} \\ 1/12 & \text{si } x_i = 4, \text{ si es desecho} \end{cases}$.

Si la ganancia por unidad está dada según la función g(x) = $18x^2$-144x+281, calcular la ganancia promedio del productor por cada unidad.

Forma 1.

1º Por la propiedad del valor esperado $E\,[g(X)] = E\,[18X^2 - 144X + 281]$

$= 18*E\,[X^2] - 144*E[X] + 281$

2º $E\,[X^2] = 1^2*\frac{1}{6} + 2^2*\frac{1}{2} + 3^2*\frac{1}{4} + 4^2*\frac{1}{12} = 23/4$ y $E\,[X] = 1*\frac{1}{6} + 2*\frac{1}{2} + 3*\frac{1}{4} + 4*\frac{1}{12} = 9/4$

3º Sustituyendo estos valores donde corresponde se obtiene una ganancia promedio por unidad de $ 60,5.

Forma 2.

1º Se obtiene la distribución de probabilidad de la ganancia. Los valores de g_i se obtienen sustituyendo los valores 1, 2, 3 y 4 de x_i, respectivamente, en la función ganancia, obteniéndose:

$$p\,(g_i) = \begin{cases} 1/6 & \text{si} \quad g_i = 155 \\ 1/2 & \text{si} \quad g_i = 65 \\ 1/4 & \text{si} \quad g_i = 11 \\ 1/12 & \text{si} \quad g_i = -7 \end{cases}$$

2º Se calcula $E\,[G] = 155*\frac{1}{6} + 65*\frac{1}{2} + 11*\frac{1}{4} + (-7)*\frac{1}{12} =$ $ 60,5 , coincidente con el resultado anterior.

c) La variable aleatoria X representa el peso (en kg) de pollos broiler de un productor, cuya distribución está dada por:

$$f\,(x) = \begin{cases} \frac{3}{10}(3x - x^2) & \text{si } 1 \le x \le 3 \\ 0 & \text{en otro caso} \end{cases} .$$

Si el productor tiene una ganancia de 0,01 UF por cada pollo que pese entre 1 y 1,5 kg , de 0,02 UF por cada pollo que pese entre 1,5 kg y 2,5 kg y de 0,015 UF cuando pesa más de 2,5 kg. ¿ Cuál será su ganancia total al vender su producción de 5000 pollos ?

De la función de distribución del peso de los pollos se establece que $P(1 \le X \le 1,5) = 13/40$, que $P(1,5 \le X \le 2,5) = 23/40$ y que $P(2,5 \le X \le 3) = 4/40$, luego la distribución de probabilidad de la ganancia queda establecida por

$$p\,(g_i) = \begin{cases} 13/40 & \text{si } g_i = 0,010 \\ 23/40 & \text{si } g_i = 0,020 \\ 4/40 & \text{si } g_i = 0,015 \end{cases}$$

, en consecuencia la ganancia promedio por pollo es $E\,[G] = 0,010*\frac{13}{40} + 0,020*\frac{23}{40} + 0,015*\frac{4}{40} =$ 0,01625 UF, po lo tanto, la ganancia total se obtiene multiplicando la ganancia promedio por unidad por el total de pollos vendidos, resultando una ganancia de 81,25 UF.

Observe que la distribución de la variable ganancia es **discreta.**

Varianza de una variable aleatoria.

Definición.

Se llama varianza de una variable aleatoria a $\boldsymbol{E[X - E[X]]^2}$.

Observación.

La definición establece que la varianza es *un promedio de desvíos al cuadrado*. Por la misma razón que en estadística descriptiva, ésta es una medida de la variabilidad del comportamiento de la variable aleatoria.

Proposición.

$$\boldsymbol{V[X] = E[X^2] - (E[X])^2}$$

Demostración.

Sea $E[X] = \mu$, entonces $V[X] = E[X-\mu]^2$, por cuadrado de binomio

$$= E[X^2 - 2\mu X + \mu^2], \text{ usando propiedades}$$
$$= E[X^2] - 2\mu * E[X] + \mu^2$$
$$= E[X^2] - 2\mu\mu + \mu^2 = E[X^2] - \mu^2$$
$$= E[X^2] - (E[X])^2$$

Ejemplos 3.3

a) Del ejemplo 3.1 a) se tiene que $p(x_i) = \frac{1}{6}\ \forall x_i \in \{1,2,3,4,5,6\}$ y que $E[X] = 3,5$, entonces $V[X] = E[X^2] - (E[X])^2 \Rightarrow E[X^2] = 1^2 * \frac{1}{6} + 2^2 * \frac{1}{6} + ... + 5^2 * \frac{1}{6} + 6^2 * \frac{1}{6} = 91/6$, luego $V[X] = \frac{91}{6} - (\frac{7}{2})^2 = 35/12$.

b) Del ejemplo 3.1 b), $p(x_i) = \begin{cases} \frac{4}{35} \text{ si } x_i = 0 \\ \frac{18}{35} \text{ si } x_i = 1 \\ \frac{12}{35} \text{ si } x_i = 2 \\ \frac{1}{35} \text{ si } x_i = 3 \end{cases}$ y $E[X] = \frac{9}{7}$. Como

$E[X^2] = 0^2 * \frac{4}{35} + 1^2 * \frac{18}{35} + 2^2 * \frac{12}{35} + 3^2 * \frac{1}{35} = 15/7 \Rightarrow V[X] = \frac{15}{7} - (\frac{9}{7})^2 = 24/49.$

c) Del ejemplo 3.2 a), $f(x) = \begin{cases} 2x & \text{si } 0 \le x \le 1 \\ 0 & \text{en otro caso} \end{cases}$ y $E[X] = 2/3$, luego

$E[X^2] = \int_0^1 x^2 * 2x\, dx = \frac{1}{2} \quad \Rightarrow V[X] = \frac{1}{2} - (\frac{2}{3})^2 = 1/18.$

d) Del ejemplo 3.2 b), $f(x) = \begin{cases} x^2/3 & \text{si } -2 \le x \le 1 \\ 0 & \text{para otros valores} \end{cases}$ y $E[X] = -\frac{5}{4}$ de donde

$E[X^2] = \int_{-2}^{1} x^2 * \frac{x^2}{3}\, dx = 11/5 \Rightarrow V[X] = \frac{11}{5} - (-\frac{5}{4})^2 = 51/80.$

La variabilidad relativa de X se obtiene con el $CV = \frac{\sqrt{V[X]}}{E[X]} = \frac{\sqrt{\frac{51}{80}}}{[-\frac{5}{4}]} = 0,639$

Propiedades de la varianza.

Tal como sucede con la esperanza las propiedades a continuación se corresponden con las vistas en estadística descriptiva.

1° $\boldsymbol{V[k] = 0}$. Se establece que la varianza de una constante es igual a cero, situación trivial por que una constante no varía.

2° $\boldsymbol{V[cX] = c^2 * V[X]}$. La propiedad establece que la constante que multiplica a la variable aleatoria multiplica al cuadrado a su varianza.

3° $\boldsymbol{V[cX \pm k] = c^2 * V[X]}$

Demostración.

$V[cX \pm k] = E(cX \pm k)^2 - (E(cX \pm k))^2$

$= E(c^2X^2 \pm 2ckX + k^2) - (cE(X) \pm k)^2$

$= c^2E(X^2) \pm 2ck\,E(X) + k^2 - (c^2(E(X))^2 \pm 2ck\,E(X) + k^2)$

$= c^2\,E(X^2) - c^2(E(X))^2$, pues k^2 y los dobles productos se anulan

$= c^2\,(E(X^2) - (E(X))^2) = c^2 * V(X)$

Esta demostración sirve para validar las dos propiedades anteriores que resultan como casos particulares de ésta.

4° Sean X e Y variables aleatorias ***independientes***, entonces $V[X \pm Y] = V[X] + V[Y]$.

Esta es una propiedad importante en estadística, porque establece que al tener dos *variables aleatorias independientes* la varianza de su suma o diferencia es **siempre** igual a la **suma** de sus varianzas, cuya demostración es el ejemplo 4.3. c) de la sección 3.4.

3.4 Nociones sobre distribuciones de variables aleatorias bidimensionales.

En muchas situaciones interesa considerar simultáneamente dos o más características en un mismo individuo, como por ejemplo, su altura y su peso ; su edad, años de educación y su ingreso mensual. Para tal efecto es necesario desarrollar algunos conceptos.

Definiciones.

1. El par $(X\,,\,Y)$ recibe el nombre de *variable aleatoria bidimensional* o *vector aleatorio* si y solo si X e Y son variables aleatorias *unidimensionales.*

Notación: $\vec{X} = (X\,,\,Y)$; $\vec{X} : S \to \Re^2$; $R_{\vec{X}}$ es su recorrido.

2. El vector aleatorio $\vec{X}$ es discreto si su recorrido es un conjunto finito o infinito numerable.

3. En $\vec{X}$ vector aleatorio discreto, una función p(x_i, y_j) que le asigne probabilidades a los elementos $(x_i\,,\,y_j)$ de $R_{\vec{X}}$, denominada función de probabilidad puntual conjunta ó de cuantía conjunta, debe satisfacer las siguientes condiciones:

1º $p(x_i\,,\,y_j) \geq 0$, $\forall\,(x_i\,,\,y_j) \in R_{\vec{X}}$

2º $\sum\limits_{(x_i,y_j)\in R_{\vec{X}}} \sum p(x_i\,,\,y_j) = 1$

4. Si $B \subset R_{\vec{X}}$, entonces $P(B) = \sum\limits_{(x_i,y_j)\in B} \sum p(x_i\,,\,y_j)$

5. El vector aleatorio $\vec{X}$ es continuo si su recorrido es una región de $\Re^2$.

6. Sea $\vec{X}$ vector aleatorio continuo, entonces una función $f\,(x\,,\,y)$, denominada función de densidad conjunta , que le asigne probabilidades a toda región B de $\Re^2$, debe satisfacer las siguientes condiciones:

1º $f\,(x\,,\,y) \geq 0$, $\forall\,(x\,,\,y) \in \Re^2$

2º $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\,(x\,,\,y)\,dA = 1$, donde $dA = dx*dy = dy*dx$

7. Si $B \subset \Re^2$, entonces $P(B) = \int_B\int f\,(x\,,\,y)\,dA$

8. Se llama *covarianza* de X e Y a $Cov(X\,,\,Y) = E[X{*}Y] - E[X]{*}E[Y]$.
La covarianza es una medida del grado de asociación entre dos variables. Si $Cov(X,Y) > 0$ la asociación entre las variables es directa, en cambio si $Cov(X,Y) < 0$ la asociación es inversa.

El inconveniente de la covarianza es que su unidad de medida depende de las de las variables y que puede tomar cualquier valor real lo que dificulta una interpretación más fina. Similar a lo que ocurre en estadística descriptiva, donde el coeficiente de variación facilita la interpretación de la variabilidad, en este caso se establece el coeficiente de correlación entre dos variables aleatorias.

9. Se llama coeficiente de correlación entre *X* e *Y* al número adimensional rho (ρ) que se calcula como la covarianza entre las variables, dividida por la raíz del producto de sus varianzas, así:
$\rho = Cov(X,Y)/\sqrt{V[X]{*}V[Y]}$, -1 $\leq \rho \leq$ 1, es decir, es un valor acotado.

10. Si Z es una función H de dos variables aleatorias, $Z = H(X\,,\,Y)$, entonces, según la distribución sea discreta o continua

$$E[Z] = \sum_{(x_i,y_j)\in R_{\vec{X}}} \sum H\,(x_i\,,\,y_j){*}\,p(x_i\,,\,y_j) \quad \text{ó} \quad E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H\,(x\,,\,y){*}f\,(x\,,\,y)\,dA$$

Ejemplos 4.1

a) La siguiente tabla de doble entrada define una función de probabilidad conjunta $p(x_i\,,\,y_j)$ para un vector aleatorio discreto

$x_i \backslash y_j$	0	2	4	6	Total
1	0,05	0,03	0,07	0,05	0,20
3	0,10	0,12	0,05	0,03	0,30
5	0,15	0,25	0,08	0,02	0,50
Total	0,30	0,40	0,20	0,10	1,00

Según la tabla $P(X = 3\,,\,Y = 0) = p(x_2\,,\,y_1) = 0,10$, $P(X = 1\,,\,Y = 4) = p(x_1\,,\,y_3) = 0,07$
$P(X > 3\,,\,Y < 4) = P(X = 5$ e $Y = 0$ ó $2) = P(X = 5, Y = 0)$ + $P(X = 5, Y = 2) = 0,40$
Observe que la suma de **todas** las casillas es uno, lo que corresponde a la probabilidad del espacio muestral.

Sea $Z = X{*}Y$, entonces de la distribución conjunta, $E[Z] = \sum_{(x_i,y_j)\in R_{\vec{X}}} \sum (\,x_i * y_j){*}\,p(x_i\,,\,y_j)$

$\Rightarrow E[Z]$= $1{*}0{*}0,05$+$1{*}2{*}0,03$+$1{*}4{*}0,07$+$1{*}6{*}0,05$+....+$5{*}0{*}0,15$+$5{*}2{*}0,25$+$5{*}4{*}0,08$+$5{*}6{*}0,02$
= 7, 2 .
Esto significa que la esperanza o promedio del *producto de* X *por* Y es 7,2 .

b) Sea $\vec{X} = (X\,,\,Y)$ con distribución conjunta $f\,(x\,,\,y) = \begin{cases} x^2 + \frac{xy}{3} & \text{si } 0 \leq x \leq 1\ ,\ 0 \leq y \leq 2 \\ 0 & \text{en otra situación} \end{cases}$
Entonces $P(X < \frac{1}{3}\,,\,Y > 1) = \int_0^{1/3}\int_1^2 (x^2 + \frac{xy}{3})\,dy\,dx = 13/324$

Sea $Z = 2X + Y$, entonces $E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (2x + y)*f\,(x\,,\,y)\,dA$
$= \int_0^1\int_0^2 (2x + y)*(x^2 + \frac{xy}{3})\,dy\,dx$
$= \int_0^1\int_0^2 (2x^3 + \frac{5}{3}x^2y + \frac{1}{3}xy^2)\,dy\,dx = 23/9.$ Significa que el valor esperado de *dos veces la variable aleatoria X más la variable aleatoria Y* es 23 / 9.

Distribuciones marginales**.**

En la tabla que define la distribución conjunta $p(x_i\,,\,y_j)$ los totales por fila, se llama *distribución marginal* de X. Los totales de columnas, se llama *distribución marginal* de Y. Las distribuciones marginales corresponden a distribuciones unidimensionales de las variables aleatorias X e Y por separado, las que se deducen de la distribución conjunta. Así del ejemplo 4.1 a), la última columna y la última fila respectivamente son las distribuciones marginales

$$p(x_i) = \begin{cases} 0,20 \text{ si} & x_i = 1 \\ 0,30 \text{ si} & x_i = 3 \\ 0,50 \text{ si} & x_i = 5 \end{cases} \qquad p(y_j) = \begin{cases} 0,30 \text{ si} & y_j = 0 \\ 0,40 \text{ si} & y_j = 2 \\ 0,20 \text{ si} & y_j = 4 \\ 0,10 \text{ si} & y_j = 6 \end{cases}$$

Las distribuciones marginales se pueden utilizar como cualquier distribución unidimensional, lo que se ilustra en los siguientes ejemplos.

$P(X = 3) = 0,30$; $P(Y > 0) = P(Y = 2) + P(Y = 4) + P(Y = 6) = 0,70$
$E[X] = 1*0,20 + 3*0,30 + 5*0,50 = 3,6$ y análogamente $E[Y] = 2,2.$
$V[X] = (1^2*0,20 + 3^2*0,30 + 5^2*0,50) - (3,6)^2 = 2,44$ y $V[Y] = 3,56$
Considerando que $E[X*Y] = 7,2$ (del ejercicio 4.1 a)) se obtiene que $Cov(X,Y) = 7,2 - 3,6*2,2 = -0,72$, con un coeficiente de correlación $\rho = -0,72/\sqrt{2,44*3,56} = -0,244.$

La función $g\,(x)$, obtenida integrando $f\,(x\,,\,y)$ respecto a y en todo su recorrido, se llama *función de ditstribución marginal de* X, luego $g(x) = \int_{-\infty}^{\infty} f(x\,,\,y)\,dy.$

La función $h\,(y)$, obtenida integrando integrando $f\,(x\,,\,y)$ respecto a x en todo su recorrido, se llama *función de ditstribución marginal de* Y, luego $h\,(y) = \int_{-\infty}^{\infty} f(x\,,\,y)\,dx.$

Para la función de densidad conjunta del ejemplo 4.1 b), las funciones marginales respectivas son:

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^2 (x^2 + \tfrac{xy}{3})\,dy = 2x^2 + \tfrac{2}{3}x \;\Rightarrow\; g\,(x) = \begin{cases} 2x^2 + \frac{2}{3}x \text{ si } 0 \le x \le 1 \\ 0 \quad \text{en otro caso} \end{cases}$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 (x^2 + \tfrac{xy}{3})\,dx = \tfrac{1}{3} + \tfrac{y}{6} \quad \Rightarrow\; h\,(y) = \begin{cases} \frac{1}{3} + \frac{y}{6} \text{ si } 0 \le y \le 2 \\ 0 \quad \text{en otro caso} \end{cases}$$

Las distribuciones $g(x)$ y $h(y)$, igual que en el caso discreto, se utilizan en situaciones como las siguientes.

$P(X > \frac{1}{2}) = \int_{\frac{1}{2}}^1 (2x^2 + \frac{2}{3}x)\,dx = 5/6$; $P(Y < 2/5) = \int_0^{2/5} (\frac{1}{3} + \frac{y}{6})\,dy = 11/7$
$E\,[Y] = \int_0^2 y\,(\frac{1}{3} + \frac{y}{6})\,dy = 10/9.$

Ejemplo 4.2

Se demostrará la propiedad que $E[X \pm Y] = E[X] \pm E[Y]$

Sea $Z = X \pm Y \quad \Rightarrow E[X \pm Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x \pm y) * f(x, y)\, dy\, dx$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x * f(x, y)\, dy\, dx \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y * f(x, y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} x\, dx * \left(\int_{-\infty}^{\infty} f(x, y)\, dy\right) \pm \int_{-\infty}^{\infty} y\, dy * \left(\int_{-\infty}^{\infty} f(x, y)\, dx\right)$$

$$= \int_{-\infty}^{\infty} x\, dx * g(x) \pm \int_{-\infty}^{\infty} y\, dy * h(y)$$

$$= E[X] \pm E[Y]$$

Variables aleatorias independientes.

Un caso importante en estadística es aquel en que dos variables aleatorias son independientes, lo cual se establece en la siguiente

Definición.

Se dice que dos variables aleatorias son independientes si y solo si su distribución conjunta es igual al producto de sus distribuciones marginales y sus rangos no dependen una de la otra.

Consecuencias.

1. Si (X, Y) es un vector aleatorio discreto, X e Y independientes $\Leftrightarrow p(x_i, y_j) = p(x_i) * p(y_j)$
2. Si (X, Y) es un vector aleatorio continuo, X e Y independientes $\Leftrightarrow f(x, y) = g(x) * h(y)$
3. X e Y independientes $\Rightarrow E[X * Y] = E[X] * E[Y] \Rightarrow Cov(X, Y) = 0$.

Demostración.

$E[X * Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy * f(x, y)\, dA$, por definición

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy * g(x) * h(y)\, dy\, dx$, por ser X e Y independientes

$= \int_{-\infty}^{\infty} x * g(x)\, dx * \left(\int_{-\infty}^{\infty} y * h(y)\, dy\right)$

$= \left(\int_{-\infty}^{\infty} y * h(y)\, dy\right) * \left(\int_{-\infty}^{\infty} x * g(x)\, dx\right)$

$= E[Y] * E[X]$

Por lo tanto $Cov(X, Y) = E[X * Y] - E[Y] * E[X] = 0$

Ejemplos 4.3

a) La siguiente tabla describe la distribución conjunta del vector aleatorio (X, Y)

$x_i \backslash y_j$	4	5	$p(x_i)$
1	0,09	0,21	0,30
2	0,15	0,35	0,50
3	0,06	0,14	0,20
$p(y_j)$	0,30	0,70	1,00

Se puede verificar que la tabla describe la distribucíon conjunta de variables aleatorias independientes, porque en cada casilla $p(x_i, y_j) = p(x_i) * p(y_j)$. Además

$E[X] = 1 * 0,30 + 2 * 0,50 + 3 * 0,20 = 1,9 \quad ; \quad E[Y] = 4 * 0,30 + 5 * 0,70 = 4,7$

$E[X * Y] = 1 * 4 * 0,09 + 1 * 5 * 0,21 + 2 * 4 * 0,15 + 2 * 5 * 0,35 + 3 * 4 * 0,06 + 3 * 5 * 0,14 = 8,93$

$\Rightarrow Cov(X, Y) = E[X * Y] - E[X] * E[Y] = 8,93 - 1,9 * 4,7 = 0$

b) Sea $f\ (x,y) = \begin{cases} \frac{1}{18}xy^2 & \text{si } 0 \le x \le 2\,,\ 0 \le y \le 3 \\ 0 & p.o.v \end{cases}$, la distribución conjunta de dos variables X e Y.

Sus funciones de distribución marginales son

$g(x) = \int_0^3 \frac{1}{18}xy^2\,dy = \frac{1}{2}x$ si $0 \le x \le 2$, luego $g(x) = \begin{cases} \frac{1}{2}x & \text{si } 0 \le x \le 2 \\ 0 & \text{para otros valores} \end{cases}$

$h(y) = \int_0^2 \frac{1}{18}xy^2\,dx = \frac{1}{9}y^2$ si $0 \le y \le 3$, luego $h(y) = \begin{cases} \frac{1}{9}y^2 & \text{si } 0 \le y \le 3 \\ 0 & \text{para otros valores} \end{cases}$

Los valores esperados de X e Y son respectivamente

$E\,[X] = \int_0^2 x*(\frac{1}{2}x)\,dx = 4/3 \quad ; \qquad E\,[Y] = \int_0^3 y*(\frac{1}{9}y^2)\,dy = 9/4$

Si $Z = X*Y$, entonces $E\,[X*Y] = \int_0^2\int_0^3 xy*(\frac{1}{18}xy^2)\,dy\,dx$
$= \int_0^2 \frac{9}{8}x^2\,dx = 3$

Se puede establecer que X e Y son variables aleatorias independientes, pues

$f(x,y) = \frac{1}{18}xy^2 = (\frac{1}{2}x)*(\frac{1}{9}y^2) = g(x)*h(y)$ y $Cov(X,Y) = 3 - \frac{4}{3}*\frac{9}{4} = 0.$

c) Se demostrará que si X e Y son variables aleatorias independientes, entonces $V\,[X \pm Y]$ es igual a $V\,[X\,] + V\,[Y]$.

Demostración.

$$\begin{aligned} V\,(X \pm Y) &= E\,(X \pm Y)^2 - (E(X \pm Y))^2 \\ &= E\,(\,X^2 + Y^2 \pm 2X*Y) - (E(X) \pm E(Y))^2 \\ &= \{E\,(X^2) + E\,(Y^2) \pm 2\,E\,(X*Y)\} - \{(E\,(X))^2 + (E\,(Y))^2 \pm 2\,E(X)*E\,(Y)\} \\ &= \{E\,(X^2) - (E\,(X))^2\} + \{E\,(Y^2) - (E\,(Y))^2\} \pm 2\,\{E\,(X*Y) - E(X)*E\,(Y)\} \\ &= \quad V\,(X) \quad + \quad V\,(Y) \quad \pm \ 2 \quad Cov(X\,,\,Y) \\ &= V\,(X) + V\,(Y)\ , \text{ pues como } X \text{ e } Y \text{ son independientes} \Rightarrow Cov(X\,,\,Y) = 0 \end{aligned}$$

Observación.

De la demostración anterior se deduce la propiedad más general de la varianza de una suma o diferencia de variables aleatorias que establece: $V\,[X \pm Y] = V\,[X\,] + V\,[Y] \pm 2*Cov\,(X\,,\,Y)$.

Ejemplos 4.4

a) Del ejemplo 4.1 b), se obtiene que $E\ [X] = 13/18$ y $E\,[Y] = 10/9$, entonces utilizando propiedades, $E\ [2X + Y] = 2*E\,[X]\ + E\,[Y] = 2*\frac{13}{18} + \frac{10}{9} = 23/9$, lo que coincide con el resultado obtenido antes por definición.
También, por propiedades $E\ [2 - X + 3Y] = 2 - E\,[X] + 3*E\,[Y] = 2 - \frac{13}{18} + 3*\frac{10}{9} = 83/18$
En este caso se puede verificar que $V\,[X + Y] \neq V\,[X]\ + V\,[Y]$, porque X e Y no son variables aleatorias independientes.

b) Con la distribución conjunta de variables aleatorias discretas del ejemplo 4.3 a) se obtiene que:

$E[X] = 1,9$

$E[Y] = 4,7$

$V[X] = (1^2 * 0,3 + 2^2 * 0,5 + 3^2 * 0,2) - (1,9)^2 = 0,49$

$V[Y] = (4^2 * 0,3 + 5^2 * 0,7) - (4,7)^2 = 0,21.$

A partir de los cuales, utilizando propiedades, se calcula:.

$E[2 - 3X + Y] = 2 - 3 * E[X] + E[Y] = 1,0$

$V[X + Y] = V[X] + V[Y] = 0,49 + 0,21 = 0,70$, porque X e Y son variables aleatorias independientes.

$V[X - Y] = V[X] + V[Y] = 0,70$, por la misma razón anterior. También:

$V[-X + 3Y + 2] = V[((-X) + 3Y) + 2] = V[(-X) + 3Y] = V[(-X)] + V[3Y]$

$= (-1)^2 * V[X] + (3)^2 * V[Y] = 0,49 + 9 * 0,21 = 2,38$

.

4. DISTRIBUCIONES DE PROBABILIDAD NOTABLES

4.1 Introducción.

El diagrama de la figura 1.1 establece las dos formas de describir el comportamiento de una población, empirica o teoricamente, la primera de las cuales requiere realizar una observación exhaustiva de la población, es decir, un censo.

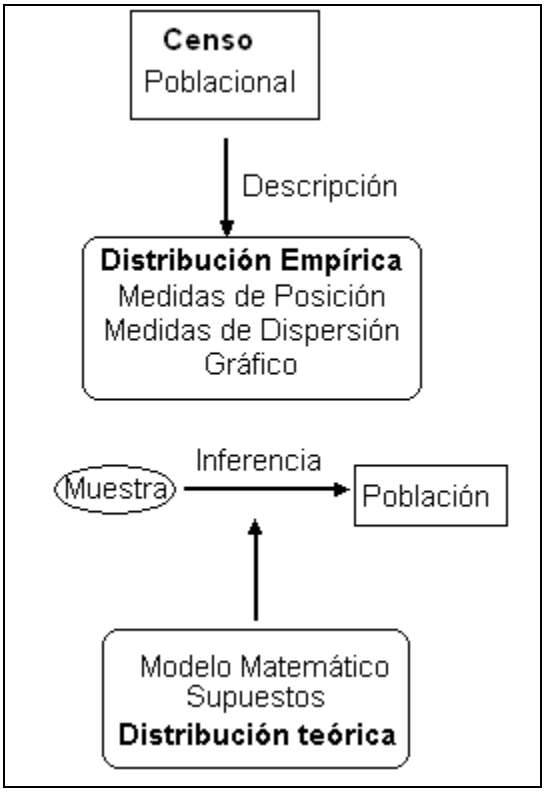


Figura 1.1. Distribuciones empíricas y teóricas.

Por lo general es difícil realizar censos para grandes poblaciones por razones principales de costo y tiempo , pero igualmente existe la necesidad de caracterizarlas basándose, si es posible, en un número razonable de observaciones, es decir, en una muestra de la población. Para tal efecto hay que recurrir a supuestos sobre el comportamiento de la población, es decir, distribuciones teóricas, de las cuales se puede asumir su *forma*, pero no sus parámetros, los cuales se deducirán a partir de la muestra, es decir, se hará una estimación. La forma de la distribución teórica se puede deducir a partir de comportamientos anteriores del fenómeno o a partir de un análisis descriptivo de la muestra si ésta contiene un número relativamente grande de observaciones como para construir un histograma de frecuencias.

Existe un gran número de distribuciones teóricas, tanto de variables continuas como de variables discretas, cada una de las cuales se expresa en términos de una función matemática, como se estudió en distribuciones de variables aleatorias. Entre las distribuciones de variable aleatoria continua más notables se debe mencionar la distribución Normal, la más importante de todas las distribuciones, la distribución Uniforme y la distribución Exponencial. De las distribuciones discretas son importantes la distribución Binomial, la más notable entre las

discretas, la distribución de Poisson y la distribución Binomial negativa, todas con aplicaciones en el ámbito agronómico.

4.2 Distribución Normal.

Es la distribución que aparece con mayor frecuencia en el comportamiento de fenómenos reales, en especial en el área de las ciencias naturales. Johann Carl Friedrich Gauss genio matemático , físico y astrónomo, de nacionalidad alemana, fue el que mayormente contribuyó a su formulación y aplicación en diferentes áreas del saber como por ejemplo en su aplicación a la teoría de los errores, de importancia en ingeniería.

Es una distribución de *variable aleatoria continua* cuya función matemática, función de densidad de probabilidad, es $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$; $-\infty \leq x \leq \infty$, cuya gráfica corresponde a una curva en forma de campana denominada *Campana de Gauss*, que como se puede apreciar depende de los parámetros μ y σ^2, que corresponden a su *valor esperado* y *varianza* respectivamente.

Notación: $X = N(\mu, \sigma^2)$

Características de la distribución normal.

1° la curva tiene forma acampanada, asintótica al eje X hacia $-\infty$ y $+\infty$. El área total encerrada por ésta y X es igual a 1, como corresponde a toda función de distribución de probabilidad.

2º la curva tiene un máximo en μ y es simétrica respecto a la recta x = μ. Luego, en esta distribución son coincidentes la media aritmética, la mediana y la moda, es decir, μ = Me = Mo.

3º la curva tiene dos puntos de inflexión que se ubican en x = $\mu - \sigma$ y x = $\mu + \sigma$

4º el área bajo la curva comprendida entre los puntos de inflexión es igual a 0,6826 (68,26%) y el área entre $\mu - 2\sigma$ y $\mu + 2\sigma$ es igual a 0,9544 (95,44%), cualesquiera sean los valores de sus parámetros μ y σ^2. Se debe recordar que el área bajo la curva, en variables aleatorias continuas, corresponde a la probabilidad de sucesos que son intervalos de números reales. En consecuencia, lo anterior se puede interpretar en el sentido que el 68,26% de los individuos que componen la población teórica tienen un valor de la variable en estudio entre $\mu - \sigma$ y $\mu + \sigma$ y en el 95,44% el valor de la variable quedará comprendida entre $\mu - 2\sigma$ y $\mu + 2\sigma$.

Ejemplo 2.1

En una lechería la producción diaria de leche por vaca, X, se distribuye $N(18, 9)$, cuya gráfica es la de la figura 2.1. De acuerdo al enunciado se puede deducir que el 68,26% de las vacas tienen una producción diaria de leche entre 15 y 21 litros de leche, que corresponde a valores entre $\mu \pm \sigma$, mientras que el 95,44% de las vacas producirían entre 12 y 24 litros diarios, que corresponde a $\mu \pm 2\sigma$.

Si la lechería cuenta con 3000 vacas la pregunta de cuántas de ellas producen entre 15 y 21 litros, se resuelve considerando que el 68,26% de ellas está en esa condición y por lo tanto el 68,26% de 3000 corresponde a 2048 vacas

Para contestar la pregunta de cuántas vacas producirán más de 24 litros, se debe considerar que el 95,44% de ellas produce entre 12 y 24 litros diarios y que en los *extremos*, es decir bajo 12 litros y sobre 24 litros, está el (100 - 95,44)% = 4,56% de las observaciones y

como la distribución es simétrica, la mitad, o sea, el 2,28% produce más de 24 litros, lo que implica que 68 son las vacas que estarían en esa condición.

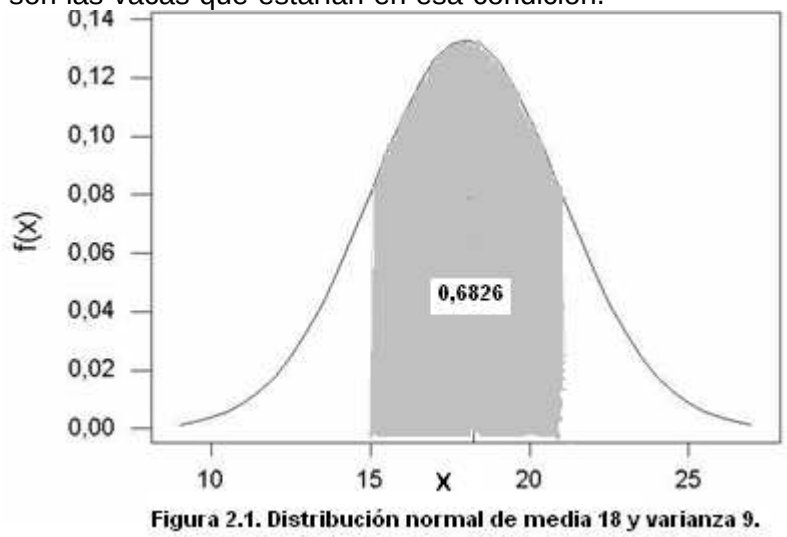


Figura 2.1. Distribución normal de media 18 y varianza 9.

Distribución Normal Típica o estándar.

Se llama distribución normal típica a $Z = N(0\,,\,1)$, con función de distribución de probabilidad $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Su función de distribución acumulativa es $\phi(z) = \int_{-\infty}^{z} f(z)\ dz$, que representa el área bajo la curva normal estándar desde -∞ hasta el valor real z. Por ejemplo el área acumulada hasta el punto *a* es $\phi(a) = \int_{-\infty}^{a} f(z)\ dz$, representada en la figura 2.2.

La relación entre el área bajo la curva normal típica con la probabilidad de Z se expresa así:

1) $P(Z \leq a) = \phi(a)$, por definición y que corresponde al área desde -∞ hasta *a.*

2) $P(a \leq Z \leq b) = \phi(b) - \phi(a)$, corresponde al área entre *a y b,* según la siguiente deducción

$\int_{-\infty}^{b} f(z)\ dz = \int_{-\infty}^{a} f(z)\ dz + \int_{a}^{b} f(z)\ dz \Rightarrow \int_{a}^{b} f(z)\ dz = \int_{-\infty}^{b} f(z)\ dz - \int_{-\infty}^{a} f(z)\ dz$

$\Rightarrow P(a \leq Z \leq b) = \phi(b) - \phi(a)$

3) $P(Z > b) = 1 - P(Z \leq b) = 1 - \phi(b)$, corresponde al área desde *b* hasta +∞.

La función de distribución acumulativa ϕ está tabulada para diferentes valores de z La razón de la tabulación radica en la situación práctica de obviar los cálculos rutinarios de la integración debido a que la $\int_{-\infty}^{z} f(z)\ dz$ no se puede resolver utilizando el Teorema Fundamental del cálculo debido a que la función $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ no tiene primitiva.

En el anexo 1 (tabla A1) se incluye una tabla de la función $\phi(z)$. Los valores para z con un decimal se presentan en la primera columna y las siguientes columnas corresponden al segundo decimal de z. Así, por ejemplo $\phi(\text{-}2,10) = 0,0179$ se lee en la intersección de línea -2,10 con la columna 0,00 y $\phi(\text{-}2,14) = 0,0162$ se lee en la misma línea en la columna 0,04. La probabilidad $\phi(1,38) = 0,9162$ se lee en la línea con z igual 1,30 y en la columna del 0,08.

Ejemplo 2.2

Se mostrarán algunos ejemplos de cálculo de probabilidades asociada a una distribución normal típica.

- $P(Z \leq 1,2) = \phi(1,2) = 0,8849$

- $P(Z < -0,65) = \phi(-0,65) = 0,2578$

-$P(0,5 \leq Z \leq 0,82) = \phi(0,82) - \phi(0,5) = 0,7939 - 0,6915 = 0,1024$

- $P(-0,5 \leq Z \leq 0,82) = \phi(0,82) - \phi(-0,5) = 0,7939 - 0,3085 = 0,4854$

- $P(-1,2 \leq Z \leq -0,7) = \phi(-0,7) - \phi(-1,2) = 0,2420 - 0,1151 = 0,1269$

- $P(Z > 1,45) = 1 - P(Z \leq 1,45) = 1 - \phi(1,45) = 1 - 0,9265 = 0,0735$

- $P(Z \geq -0,65) = 1 - P(Z < -0,65) = 1 - 0,2578 = 0,7422$

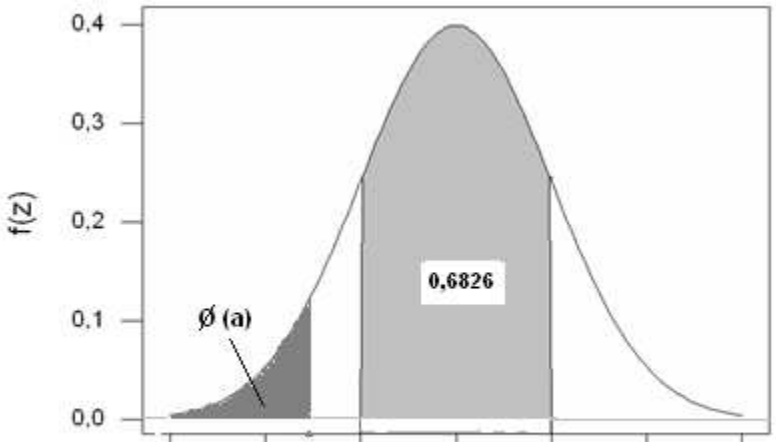


Figura 2.2. Distribución normal tipica

A continuación se enunciará un teorema de enorme importancia estadística, porque establece la relación entre una distribución normal cualquiera y la distribución normal típica.

Teorema.

Sea X variable aleatoria con distribución $N(\mu\,,\,\sigma^2)$, entonces la variable tipificada $Z = \frac{X-\mu}{\sigma}$ tiene distribución $N(0\,,\,1)$.

Esto es de especial relevancia porque limita los cálculos de probabilidad de distribuciones normales al uso de una tabla única, como la A1 del anexo.

Consecuencia.

Se fundamentará matemáticamente cómo probabilidades asociadas a una variable $X = N(\mu,\,\sigma^2)$ se pueden obtener a partir de probabilidades de una normal típica.

Por definición $P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} EXP(-(\frac{x-\mu}{\sigma})^2/2)\ dx$, al realizar en la integral la sustitución $z = \frac{x-\mu}{\sigma}$ se deduce que $dx = \sigma\, dz$ y que los límites de integración de la integral transformada son respectivamente $z_1 = \frac{a-\mu}{\sigma}$ y $z_2 = \frac{b-\mu}{\sigma}$. Entonces

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} EXP(-(\frac{x-\mu}{\sigma})^2/2)\, dx \\ &= \int_{z_1}^{z_2} \frac{1}{\sigma\sqrt{2\pi}} EXP(-z^2/2)\, \sigma\, dz \\ &= \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} EXP(-z^2/2)\, dz \\ &= P(z_1 \leq Z \leq z_2) \\ &= P(\tfrac{a-\mu}{\sigma} \leq Z \leq \tfrac{b-\mu}{\sigma}). \\ &= \phi(\tfrac{b-\mu}{\sigma}) - \phi(\tfrac{a-\mu}{\sigma}). \end{aligned}$$

Ejemplos 2.3

a) Para ilustrar el uso del teorema en el cálculo de probabilidades, considérese la variable $X = N(22, 25)$ cuya transformación $Z = \frac{X-22}{5} = N(0\,,\,1)$, entonces para obtener probabilidades de eventos de X se procede como a continuación

$$P(X < 12) = P(\tfrac{X\text{- }22}{5} < \tfrac{12\text{ - }22}{5}) = P(Z < \tfrac{12\text{ - }22}{5}) = \phi(\tfrac{12\text{ - }22}{5}) = \phi(\text{- }2) = 0,0228$$

$$\begin{aligned} P(20 \leq X \leq 25) &= P(\tfrac{20\text{ - }22}{5} \leq \tfrac{X-22}{5} \leq \tfrac{25-22}{5}) = P(\tfrac{20\text{ - }22}{5} \leq Z \leq \tfrac{25-22}{5}) \\ &= \phi(0,6) - \phi(\text{- }0,8) = 0,7257 - 0,2119 = 0,5138 \end{aligned}$$

$$P(X \geq 29) = 1 - P(X < 29) = 1 - P(Z < \tfrac{29\text{ - }22}{5}) = 1 - \phi(1,4) = 1 - 0,9192 = 0,0808$$

b) En una lechería la producción de leche por vaca tiene distribución $X = N(18, 9)$, representada en la figura 2.1. ¿Cuál es la probabilidad que una vaca elegida al azar:

1) produzca menos de 12 litros

Es necesario transformar $X = N(18, 9)$ a $Z = N(0, 1)$, lo que implica que $Z = \frac{X\text{ - }18}{3}$, luego $P(X < 12) = P(\frac{X\text{ - }18}{3} < \frac{12-18}{3})$ $P(Z < \text{-}2) = 0,0228$, por lo tanto la probabilidad que una vaca elegida al azar produzca menos de 12 litros es de 0,0228. También se puede decir que el 2,28% de las vacas de la lechería producen menos de 12 litros diarios.

2) tenga una producción entre 21 y 24 litros?

Esto es, $P(21 \leq X \leq 24) = P(1,0 \leq Z \leq 2,0) = \phi(2,0) - \phi(1,0) = 0,9772 - 0,8413 = 0,1359.$
Por lo tanto la probabilidad que una vaca cualquiera produzca entre 21 y 24 litros diarios es de 0,1359.

3) produzca entre 15 y 22 litros?

$P(15 \leq X \leq 22) = P(-1 \leq Z \leq 1,33) = \phi(1,33) - \phi(-1,0) = 0,9082 - 0,1587 = 0,7495.$
En consecuencia la probabilidad que una vaca elegida al azar tenga una producción entre 15 y 22 litros es de 0,7495.

4) tenga una producción mayor a 25 litros?

$P(X > 25) = 1 - P(Z \leq 2,33) = 1 - \phi(2,33) = 1 - 0,9901 = 0,0099.$ Es decir, el 0,99% de las vacas de la lechería produce más de 25 litros.

Valores percentiles de la distribución normal típica.

Los valores percentiles de distribuciones de probabilidad son de gran importancia en estadística. En el caso de la distribución normal el valor percentil consiste en obtener el valor de *a* tal que $P(Z \leq a) = \alpha,\ 0 < \alpha < 1$. Conceptualmente esta situación es la inversa de la desarrollada en la sección anterior. Es decir, si la distribución tabulada es $P(Z \leq a) = \phi(a)$, entonces $P(Z \leq a) = \alpha$ implica $\phi(a) = \alpha$, luego $a = \phi^{-1}(\alpha)$ es la inversa de la función de distribución acumulativa normal.

Notación. Se utilizará la notación percentil $z_\alpha = \phi^{-1}(\alpha)$

Los valores percentiles z_α se obtienen de la misma tabla, función acumulativa de la normal estándar, usándola en forma inversa.

Ejemplos 2.4

En cada caso obtener el valor de a que cumpla con la probabilidad dada a partir de una tabla de la distribución acumulativa normal estándar:

a) $P(Z < a) = 0,1093 \Rightarrow \phi(a) = 0,1093 \Rightarrow a = \phi^{-1}(0,1093) = z_{0,1093} = -1,23$

b) $P(Z < a) = 0,8159 \Rightarrow \phi(a) = 0,8159 \Rightarrow a = \phi^{-1}(0,8159) = z_{0,8159} = 0,90$

c) $P(Z > a) = 0,20 \Rightarrow 1 - P(Z \leq a) = 0,20 \Rightarrow P(Z \leq a) = 0,80 \Rightarrow a = \phi^{-1}(0,80) = 0,84$

d) $P(Z > a) = 0,1093 \Rightarrow 1 - \phi(a) = 0,1093 \Rightarrow \phi(a) = 0,8907 \Rightarrow a = \phi^{-1}(0,8907) = 1,23$

e) $P(Z > a) = 0,10 \Rightarrow \phi(a) = 0,90 \Rightarrow a = \phi^{-1}(0,90) = z_{0,90} = 1,28$

f) $P(Z < a) = 0,10 \Rightarrow \phi(a) = 0,10 \Rightarrow z_{0,10} = -1,28$

Observaciones.

1) Los ejemplos a) y d) , e) y f) corresponden a situaciones simétricas en la distribución normal típica, por lo cual sus valores percentiles tienen el mismo valor pero con signos opuestos. Ello siempre ocurrirá con los valores percentiles complementarios, esto es $z_\alpha = - z_{1-\alpha}$.

2) Una situación de gran importancia en estadística son los intervalos de probabilidad central $(1 - \alpha)$ de la distribución normal típica Z, cuyos extremos son valores percentiles simétricos, que en términos probabilísticos es $P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$ o en forma equivalente $P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$, pues por la observación anterior $z_{\alpha/2} = -z_{1-\alpha/2}$ por corresponder a valores percentiles simétricos.

Así $P(z_{0,10} \leq Z \leq z_{0,90}) = 0,80 \Rightarrow P(-1,28 \leq Z \leq 1,28) = 0,80$ (ver ejemplo e) y f) anterior) y $P(z_{0,025} \leq Z \leq z_{0,0975}) = 0,95 \Rightarrow P(-1,96 \leq Z \leq 1,96) = 0,95$, valores que se encuentran en el cuadro 2.1.

Valores percentiles notables de la distribución normal típica.

Ciertos valores percentiles de la distribución normal típica tienen uso frecuente en inferencia y por esta razón se denominarán valores notables, los cuales se resumen en la siguiente tabla.

α	z_{α}	$z_{1-\alpha}$
$0,10$	$-1,28$	$1,28$
$0,05$	$-1,645$	$1,645$
$0,025$	$-1,96$	$1,96$
$0,01$	$-2,33$	$2,33$

Cuadro 2.1. Valores percentiles notables de la distribución Z

También se pueden calcular valores percentiles de distribuciones normales cualesquiera y para ello se debe realizar el proceso de tipificación tal como en el cálculo de probabilidades.

Ejemplo 2.5

Si en un cierto huerto ocurre que el peso X de manzanas Granny, tiene distribución normal con media 140 gr y desviación típica de 20 gr, entonces se pueden determinar situaciones como las siguientes.

a) El peso máximo del 10% de las manzanas de menor peso, o sea, el percentil 10.
La distribución de las manzanas es $X = N(140, 400)$ para la cual se está pidiendo *a* tal que $P(X < a) = 0,10 \Rightarrow P(Z < \frac{a-140}{20}) = 0,10 \Rightarrow \phi(\frac{a-140}{20}) = 0,10 \Rightarrow \quad \frac{a-140}{20} = \phi^{-1}(0,10) = -1,28.$
Despejando, se obtiene que $a = 114,40$ gr, en consecuencia el 10% de las manzanas más pequeñas pesan **menos de 114,4 gr**.

b) El peso mínimo del 5% de las manzanas más grandes, es decir, el percentil 95.
$P(X > a) = 0,05 \Rightarrow 1 - \phi(\frac{a-140}{20}) = 0,05 \Rightarrow \phi(\frac{a-140}{20}) = 0,95 \Rightarrow \frac{a-140}{20} = \phi^{-1}(0,95) = 1,645.$
Despejando se obtiene que $a = 172,9$ gr. Luego el 5% de las manzanas más grandes pesan **sobre los 172,9 gr**.

c) Entre que peso se encuentra el 90% central de las manzanas.
El 90% central se encuentra entre el percentil 5 y el 95 de la distribución del peso de las manzanas, designados respectivamente por *a* y *b,* valores simétricos en relación a la media 140. Por lo tanto $P(a \leq X \leq b) = 0,90 \Rightarrow P(\frac{a-140}{20} \leq Z \leq \frac{b-140}{20}) = 0,90,$ luego
$\phi(\frac{a-140}{20}) = 0,05 \quad \Rightarrow \frac{a-140}{20} = \phi^{-1}(0,05) = -1,645 \quad \Rightarrow a = 107,1.$
$\phi(\frac{b-140}{20}) = 0,95 \Rightarrow \frac{b-140}{20} = \phi^{-1}(0,95) = 1,645 \Rightarrow b = 172,9,$ por consiguiente el 90% central de las manzanas pesa entre 107,1 gr y 172,9 gr.

4.3 Distribución Uniforme.

La distribución uniforme es la símil de la distribución equiprobable de variable aleatoria discreta, y establece que en cualquier posición, dentro del rango de valores de la variable, la probabilidad de un suceso está en relación con la longitud del intervalo que lo define. Por ejemplo si *a* , *b* , *c* y *d*, en orden de magnitud, pertenecen al rango de valores y si *b* - *a* = *d* - *c* , entonces en una distribución uniforme se cumple que $P(a \leq X \leq b) = P(c \leq X \leq d)$. En consecuencia la función de distribución de probabilidad de la distribución uniforme debe ser una función constante en el intervalo de valores de X. Una variable aleatoria continua X tiene

distribución uniforme de parámetros *a* y *b* si su función de distribución de probabilidad es de la forma $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{para otros valores} \end{cases}$.

Notación: $X =$ *Unif (a , b)*

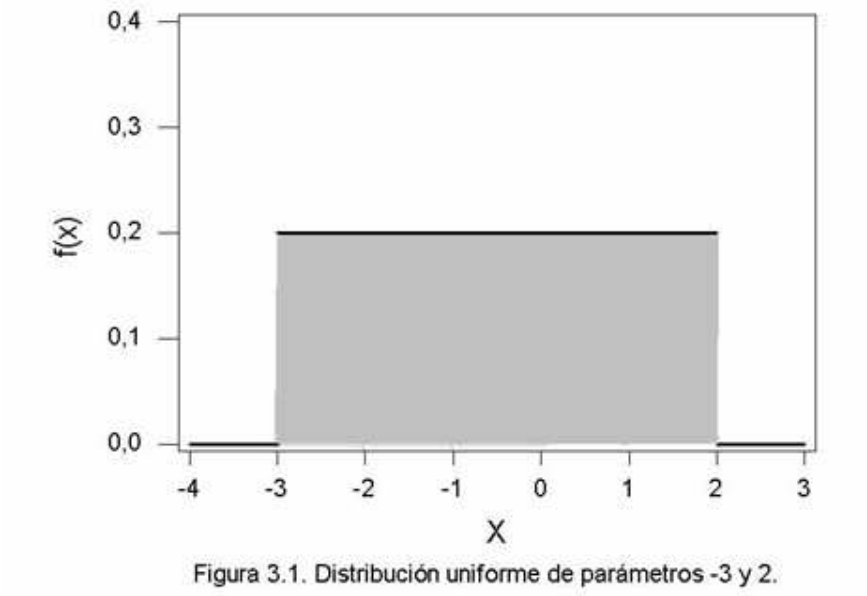


Figura 3.1. Distribución uniforme de parámetros -3 y 2.

Valores característicos.

Es fácil deducir aplicando las definiciones de valor esperado y varianza que $E[X] = \frac{a+b}{2}$ y $V[X] = \frac{(b-a)^2}{12}$.

Ejemplo 3.1

En un terminal de buses la frecuencia de salida a un cierto destino es de treinta minutos a partir de las 7:00 AM. Un usuario frecuente llega al terminal en un instante que está distribuido uniformemente entre las 7:30 y las 8:00 hras. Si llega justo a la hora de salida ya no puede abordarlo y debe esperar el siguiente, de modo que su espera máxima es de 30 minutos.

La hora de llegada del usuario ocurre, también, en un intervalo de 30 minutos, luego es una distribución uniforme en un intervalo de longitud treinta, por lo tanto las probabilidades de tiempos de espera es la razón con respecto a 30 del tiempo desde su hora de llegada hasta las 8 hras.

Si, por ejemplo, el usuario tuviera que esperar al menos 10 minutos, su hora de llegada debe ser entre las 7:30 y las 7:50, es decir, en un intervalo de longitud 20, lo que implica una probabilidad de ocurrencia de 20/30 o 2/3.

Si interesa la probabilidad de que tenga que esperar menos de 16 minutos, su llegada debe ser entre las 7:44 y las 8:00, correspondiente a un intervalo de longitud 16, cuya probabilidad es 8/15.

Para una espera de al menos 5 minutos, cuando la frecuencia de salida es cada 15 minutos, su llegada debe ser entre las 7:30 y 7:40 o entre las 7:45 y 7:55, esto es, dos intervalos de longitud 10, pero la llegada del usuario sigue siendo uniforme en un intervalo de 30 minutos, luego la probabilidad de tal evento es 20/30. Por otra parte la probabilidad de una

espera de menos de 8 minutos es de 16/30, pues su llegada debe ser entre 7:37 y 7:45 o 7:52 y 8:00, que corresponde a dos intervalos de longitud 8.

4.4 Distribución Exponencial

La variable aleatoria continua X tiene distribución exponencial de parámetro α si su función de densidad tiene la forma $f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x \geq 0\,,\ \alpha > 0 \\ 0 & \text{si } x < 0 \end{cases}$.

Notación: $X = Exp(\alpha)$

Valores característicos.

Aplicando las definiciones de valor esperado y varianza a la distribución exponencial y utilizando un poco de cálculo integral se puede establecer que $E[X] = 1/\alpha$ y $V[X] = 1/\alpha^2$.

Ejemplo 4.1

Un pesticida, que se degrada inicialmente en forma muy rápida, tiene un promedio de residualidad de 8 días. Por residualidad se entenderá que el producto es aún efectivo en ese instante. Por experiencias anteriores se sabe que la variable aleatoria *T*, días de residualidad, se ajusta a una distribución exponencial.

a) determinar la función de distribución de probabilidad del tiempo de residualidad *T*

$$E[T] = 8 = 1/\alpha \Rightarrow \alpha = 1/8 = 0,125 \Rightarrow f(t) = \begin{cases} 0,125 * e^{-0,125*t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

b) ¿ cuál es la probabilidad que el insecticida tenga una residualidad mayor a 16 días ?

Para este propósito conviene obtener la función de distribución acumulativa que implica $\int_0^t 0,125 * e^{-0,125*x} dx = 1 - e^{-0,125*t}$, si $t \geq 0$, o sea, $F(t) = \begin{cases} 0 & \text{, si } t < 0 \\ 1 - e^{-0,125*t} & \text{, si } t \geq 0 \end{cases}$, debiéndose calcular $P(T \geq 16) = 1 - F(16) = e^{-16/8} = e^{-2} = 0,135$. Luego el insecticida tiene efectividad después de los 16 días con una probabilidad de 0,135.

c) ¿ cuál es el valor mediano de la residualidad del insecticida ?

La mediana corresponde al valor de t tal que $F(t) = 0,5 \Rightarrow 1 - e^{-0,125*t} = 0,5$ $\Rightarrow e^{-0,125*t} = 0,5$. Aplicando logaritmo y resolviendo se obtiene que $t \approx 5,5$ días. Esto significa que hay una probabilidad del 50% que el producto dure menos de 5,5 días.

d) ¿ después de cuántos días existe una probabilidad menor a 0,05 de que haya residualidad del producto ?

Se debe calcular $P(T > t) < 0,05 \Rightarrow 1 - F(t) < 0,05 \Rightarrow F(t) > 0,95 \Rightarrow e^{-t/8} < 0,05$, usando logaritmo y despejando se obtiene $t > 24$ días, es decir, después de los 24 días.

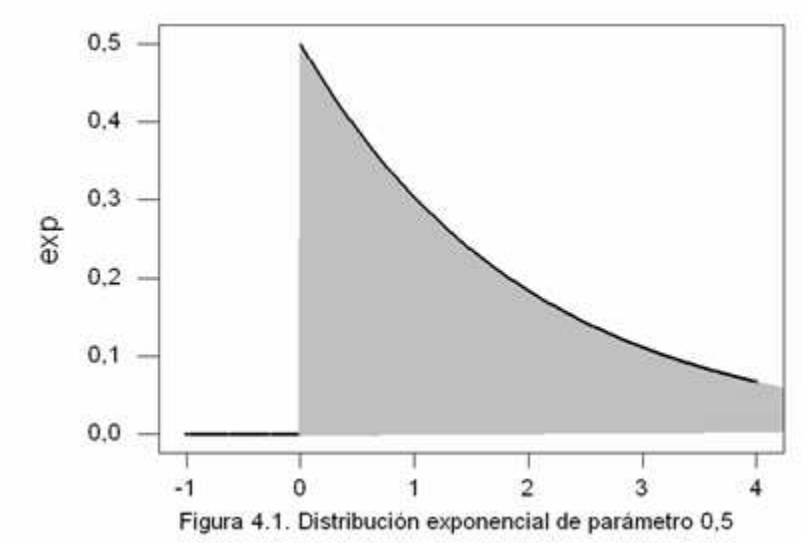


Figura 4.1. Distribución exponencial de parámetro 0,5

4.5 Distribución de Bernoulli.

Existen experimentos dicotómicos en los cuales el resultado se puede establecer en términos de *éxito* o *fracaso*, es decir, ocurre A o A', como por ejemplo un individuo puede estar sano o enfermo, vivo o muerto, defectuoso o no defectuoso. En estos casos la variable asociada es una variable aleatoria Y denominada variable Bernoulli tal que $Y = 1$ si ocurre A, es decir, se obtiene un éxito cuya probabilidad de ocurrencia es p, e $Y = 0$ si ocurre A', o sea, se obtiene un fracaso cuya probabilidad de ocurrencia es $1 - p$. Formalmente la distribución de Y es

$$p(y_i) = \begin{cases} 1-p & \text{si } y_i = 0 \\ p & \text{si } y_i = 1 \end{cases}$$

Valores característicos.

$$E[Y] = p \; ; \; V[Y] = p*(1-p)$$

Demostración.

Aplicando las definiciones de esperanza y varianza se tiene que

1) $E[Y] = 0*(1-p) + 1*p = p$.
2) $V[Y] = E[Y^2] - (E[Y])^2 = (0^2*(1-p) + 1^2*p) - (p)^2 = p - p^2 = p*(1-p)$.

4.6 Distribución Binomial.

La distribución binomial se origina cuando se seleccionan al azar individuos para establecer si poseen o no una determinada característica A. La elección debe ser *independiente* y la probabilidad de que un individuo presente la característica A es la misma de un individuo a otro. Estas condiciones se dan por ejemplo cuando de un conjunto muy grande de semillas de una determinada variedad, entre las cuales el porcentaje de germinación es *p*, se seleccionan *n* semillas en forma independiente, luego el número de semillas germinadas sigue una distribución binomial. O cuando en un vivero la probabilidad de que una planta esté enferma es *p* y se seleccionan *n* plantas al azar para ser examinadas, entonces el número de plantas

enfermas entre las *n* seleccionadas tiene distribución binomial. Es condición en estos casos que la *selección sea una tras otra y con sustitución*, al ser finitas las poblaciones definidas, pero si el número de individuos, semillas o plantas, es muy grande la sustitución es irrelevante.

Considérese el caso de un vivero en el cual $P(\text{enferma}) = p$, $P(\text{sana}) = 1 - p = q$ y del cual se seleccionan 10 plantas al azar. Si X es el número de plantas enfermas que hay entre las 10 seleccionadas, entonces para calcular las probabilidades se procede como se estableció en probabilidades para sucesos independientes. Así las probabilidades de obtener 4, 7 o x_i plantas enfermas se calculan según el siguiente procedimiento:
$P(X = 4) = c * P(E, E, E, E, S, S, S, S, S, S)$, donde $(E, E, E, E, S, S, S, S, S, S)$ es una de las formas en que puede ocurrir 4 enfermas y 6 sanas, cuya probabilidad es $p^4 * q^6$, debido a la independencia de la selección y c es el número de formas distintas en que puede ocurrir esa combinación y por lo tanto $c = \binom{10}{4} = 210$. Se deduce, entonces que $P(X = 4) = 210 * p^4 * q^6$.

De manera análoga se establece que $P(X = 7) = \binom{10}{7} * p^7 * q^3 = 120 * p^7 * q^3$ y que $P(X = x_i) = \binom{10}{x_i} * p^{x_i} * q^{10-x_i}$, porque al haber x_i plantas enfermas *p* debe aparecer x_i veces en el producto, *q* debe aparecer el resto de las veces y $\binom{10}{x_i}$ es el número de ordenamientos posibles en que puede ocurrir x_i enfermas y $(10 - x_i)$ sanas.

Formalmente, la variable aleatoria discreta X, correspondiente al *número de veces que ocurre un suceso A, cuya probabilidad de ocurrencia es p, en n observaciones independientes de un experimento ϵ, tiene distribución binomial de parámetros n y p*, con función de distribución de probabilidad $p(x_i) = \binom{n}{x_i} * p^{x_i} * q^{n-x_i}$, $x_i = 0, 1, 2, ..., n$, con función de distribución acumulativa $F(x; n,p) = \sum_{x_i=0}^{x} \binom{n}{x_i} * p^{x_i} * q^{n-x_i}$.

Notación: $X = Bin\ (n\ ,\ p)$

Observaciones.

1) La distribución recibe el nombre de binomial porque cada uno de los valores $p(x_i)$ resulta ser un término del desarrollo del binomio $(q + p)^n = \sum_{xi=0}^{n} \binom{n}{x_i} * p^{x_i} * q^{n-x_i}$, cuya suma evidentemente es igual a 1, pues $p + q = 1$.

2) Es fácil establecer la relación entre la distribución binomial y la distribución de Bernoulli, de parámetro *p*, ya que a cada una de las *n* observaciones independientes de ϵ corresponde un valor 1 o 0, según haya ocurrido o no el suceso *A*, por la cual la variable binomial X corresponde a la suma de *n* variables independientes Bernoulli, es decir, $X = \sum_{i=1}^{n} Y_i = Bin\ (n\ ,\ p)$.

Por ejemplo el valor X = 4, donde X es el número de plantas enfermas al examinar 10 plantas al azar, puede ocurrir de varias maneras, como por ejemplo, $(E, E, E, E, S, S, S, S, S, S)$ = $(1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$ ó $(E, S, S, E, S, S, S, E, E, S)$ = $(1, 0, 0, 1, 0, 0, 0, 1, 1, 0)$, que corresponden a una sucesión de 10 variables Bernoulli, cuya *suma* es el *valor* de X igual a 4.

3) Existen tablas de la distribución acumulativa binomial, para ciertas combinaciones de *n* y *p.*

Valores característicos.

$$E[X] = n*p \;;\; V[X] = n*p*(1-p)$$

Demostración.

Dado que $X = \sum_{i=1}^{n} Y_i$, y que para cada Y_i, $E[Y_i] = p$, $V[Y_i] = p*(1-p)$, luego por la propiedad del valor esperado y de la varianza para variables independientes

1) $E[X] = E[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} (E[Y_i]) = \sum_{i=1}^{n} (p) = n*p$

2) $V[X] = V[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} (V[Y_i]) = \sum_{i=1}^{n} p*(1-p) = n*p*(1-p)$.

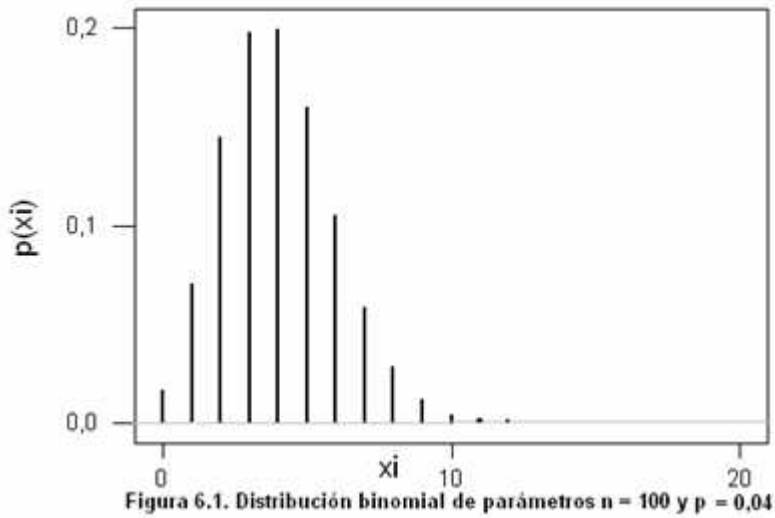


Figura 6.1. Distribución binomial de parámetros n = 100 y p = 0,04.

Ejemplo 6.1

En un vivero la probabilidad que una planta de vid tenga virus es de 0,04. Un viticultor necesita comprar 100 parras al vivero.

Si X *es el número de plantas con virus que hay entre las compradas por el viticultor*, entonces $X = Bin(100,\ 0,04)$, con $p(x_i) = \binom{100}{x_i} * (0,04)^{x_i} * (0,96)^{100-x_i}$, $x_i = 0, 1, 2, ..., 100$, cuyo gráfico es el de la figura 6.1.

a) ¿Cuántas plantas con virus se espera que adquiera el viticultor?

Esto se refiere al $E[X] = 100*0,04 = 4$, es decir se espera, pero no tiene que ocurrir necesariamente así, que éste en promedio adquiera 96 plantas sanas y 4 enfermas con virus .

b) ¿ Cuál es la probabilidad de que el viticultor adquiera
1) ninguna planta con virus ?
Corresponde a $P(X=0) = \binom{100}{0} * (0,04)^0 * (0,96)^{100} = 0,0169$, luego existe una probabilidad de 1,7% de que este suceso ocurra.
2) al menos una planta con virus ?
Esto es $P(X \geq 1) = 1 - P(X=0) = 1 - 0,0169 = 0,9831$, o sea, esto ocurrirá en el 98,31% de los casos.
3) entre 5 y 10 plantas con virus, ambos valores incluidos ?
Luego, $P(5 \leq X \leq 10) = \sum_{xi=5}^{10} \binom{100}{x_i} * (0,04)^{x_i} * (0,96)^{100-x_i}$, pero el cálculo de esta probabilidad, aún con calculadora científica, es tedioso, por lo que es conveniente utilizar tablas de la distribución acumulativa binomial como la del anexo 2 (tabla A2). .

Entonces $P(5 \leq X \leq 10) = F(10;100,0,04) - F(4;\ 100,0,04) = 0,9978 - 0,6289 = 0,3689$. Este suceso ocurrirá el 36,89% de las veces.

4) exactamente 4 plantas con virus ?
De la tabla A2, $P(X=4) = F(4;\ 100,0,04) - F(3;\ 100,0,04) = 0,6289 - 0,4295 = 0,1994$. Esto corrobora que la probabilidad de ocurrencia del valor esperado de una variable aleatoria discreta no es necesariamente un valor alto y en este caso ocurre aproximadamente el 20% de las veces.

Aproximaciones de la distribución binomial.

Existen dos aproximaciones para la distribución binomial. Una de estas aproximaciones es a la distribución de Poisson y ocurre cuando *n* es "grande" y *p* o *q* pequeño. Esta distribución se tratará a continuación. La otra aproximación es a la distribución normal la que resulta bastante satisfactoria cuando $n*p*q > 4$. Según esta condición la aproximación no resulta buena para el problema del viticultor, pues $100*0,04*0,96 < 4$. En la figura 6.2 se ilustra el caso de la aproximación a la normal de una binomial con *p* = 0,3 y con *n* de 30, 120 y 270 a distribuciones N(9, 6,3), N(36, 25,2) y N(81, 56,7) respectivamente.

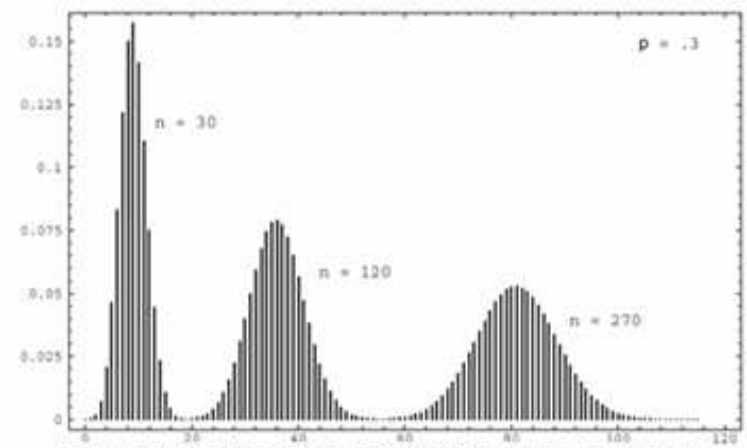


Figura 6.2. Aproximación de l distribución binomial a la normal. (Fuente internet)

A continuación un ejemplo ilustrativo.

Ejemplo 6.2

En un vivero una planta de kiwi tiene una probabilidad de 0,2 de estar enferma. Se examinan una a una 64 plantas seleccionadas al azar.

Si X es el número de plantas enfermas detectadas en las 64 examinadas, entonces

1) Como $X = Bin\ (64, 0{,}2)$, se utilizará la distribución exacta de la tabla A2 para calcular las siguientes dos probabilidades:

- $P(5 \leq X \leq 10) = F(10;\ 64,\ 0,2) - F(4;\ 64,\ 0,2) = \ 0,2410 - 0,0021 = 0,2389$

- $P(8 \leq X \leq 18) = F(18;\ 64,\ 0,2) - F(7;\ 64,\ 0,2) = 0,9579 - 0,0420 = 0,9159$

2) Dado que $E\,[X] = 64/5$, $V\,[X] = 256/25$ y que $n*p*q = 64*0{,}2*0{,}8 = 10{,}24 > 4$, entonces $X \approx N(\frac{64}{5}\ ,\ \frac{256}{25})$, entonces se utilizará esta aproximación para calcular las probabilidades anteriores:

- $P(5 \leq X \leq 10) \approx P(4.5 \leq X \leq 10.5) = P(\text{-}2,59 \leq Z \leq \text{-}0,72) = \phi(\text{-}0,72) - \phi(\text{-}2,59)$, luego $P(5 \leq X \leq 10) \approx 0,2310$.

- $P(8 \leq X \leq 18) \approx P(7,5 \leq X \leq 18,5) = \phi(1,78) - \phi(\text{-}1,66) = 0,9140$

Las probabilidades de la distribución normal fueron obtenidas de la tabla A1. Al usar la aproximación se debe realizar la *corrección por continuidad* propuesta por Yates, que consiste en restar 0,5 en el límite inferior del intervalo y sumar 0,5 en el límite superior, pues se debe asumir que en un intervalo con límites números enteros al pasar a otro con límites en un contínuo, el intervalo parte media unidad antes y termina media unidad después. Esta aproximación se aplicó en los cálculos anteriores. Comparando los resultados exactos con los de la aproximación, se aprecia que las diferencias son del orden de milésimas, con la ventaja que es más práctico aplicar la distribución normal como se verá más adelante.

4.7 Distribución de Poisson.

La distribución de Poisson se puede presentar como una distribución límite de la distribución binomial cuando *n* tiende a infinito y *p* ó *q* tiende a cero. La deducción de esta distribución se hará aplicando esta situación límite.

Sean *n* y *p* tal que $n*p$ se mantenga constante e igual a un valor λ, entonces para obviar el doble limite se sustituirá $p = \lambda / n$, pues de esta manera *p* tenderá a cero cuando *n* tienda a infinito. Luego

$$p(x_i) = \lim_{n\to\infty} \lim_{p\to 0} Bin(n,p) = \lim_{n\to\infty} Bin(n, \tfrac{\lambda}{n}) = \lim_{n\to\infty} \binom{n}{x_i} * (\tfrac{\lambda}{n})^{x_i} * (1 - \tfrac{\lambda}{n})^{n-x_i}$$

$$= \lim_{n\to\infty} \binom{n}{x_i} * \frac{\lambda^{x_i}}{n^{x_i}} * (1 - \tfrac{\lambda}{n})^{n} \ * (1 - \tfrac{\lambda}{n})^{-x_i}.$$

Pero $\binom{n}{x_i} = \frac{n*(n\text{-}1)*(n\text{-}2)\text{.....}*(n-x_i+1)}{x_i\,!}$ tiene x_i factores tanto en numerador como en denominador.

Por lo tanto, $p(x_i) = \lim_{n\to\infty} \frac{n*(n\text{-}1)*(n\text{-}2)\text{.....}*(n-x_i+1)}{x_i\,!} * \frac{\lambda^{x_i}}{n^{x_i}} * (1 - \tfrac{\lambda}{n})^{n} * (1 - \tfrac{\lambda}{n})^{-x_i}$

$$= \lim_{n\to\infty} \frac{n*(n\text{-}1)*(n\text{-}2)\text{.....}*(n-x_i+1)}{n^{x_i}} * \frac{\lambda^{x_i}}{x_i\,!} * (1 - \tfrac{\lambda}{n})^{n} \ * (1 - \tfrac{\lambda}{n})^{-x_i}$$

$$= \frac{\lambda^{x_i}}{x_i\,!} * \lim_{n\to\infty} \frac{n*(n\text{-}1)*(n\text{-}2)\text{.....}*(n-x_i+1)}{n^{x_i}} * (1 - \tfrac{\lambda}{n})^{n} \ * (1 - \tfrac{\lambda}{n})^{-x_i}$$

$$= \frac{\lambda^{x_i}}{x_i!} * \lim_{n\to\infty} \frac{n*(n-1)*(n-2).....*(n-x_i+1)}{n^{x_i}} * \lim_{n\to\infty} (1-\frac{\lambda}{n})^n * \lim_{n\to\infty} (1-\frac{\lambda}{n})^{-x_i}.$$

Se determinará cada uno de los tres límites por separado.

u = $\lim_{n\to\infty} \frac{n*(n-1)*(n-2).....*(n-x_i+1)}{n^{x_i}} = \lim_{n\to\infty} (\frac{n}{n} * \frac{n-1}{n} * \frac{n-2}{n} * ..... * \frac{n-x_i+1}{n}) = 1$, pues el límite de cada uno de los x_i factores es 1.

v = $\lim_{n\to\infty} (1-\frac{\lambda}{n})^n$ = $e^{-\lambda}$, que corresponde a un límite matemático notable

w = $\lim_{n\to\infty} (1-\frac{\lambda}{n})^{-x_i}$ = ($\lim_{n\to\infty} (1-\frac{\lambda}{n})$)$^{-x_i}$ = $1^{-x_i} = 1$.

En consecuencia $p(x_i) = \frac{\lambda^{x_i}}{x_i!} * u*v*w = \frac{\lambda^{x_i}}{x_i!} * e^{-\lambda}$.

Así, una variable aleatoria discreta *X* con función de probabilidad $p(x_i) = \frac{\lambda^{x_i}}{x_i!} * e^{-\lambda}$, $x_i = 0, 1, 2, 3.....$, se denomina distribución de Poisson de parámetro $\lambda > 0$, con función de distribución acumulativa $F(x; \lambda) = \sum_{x_i=0}^{x} \frac{\lambda^{x_i}}{x_i!} * e^{-\lambda}$.

Notación: $X = \mathcal{P}$ *(*λ*).*

Existen tablas de la distribución acumulativa de Poisson para diferentes valores de λ.

Ejemplo 7.1

Un ejemplo típico asociado a la aproximación binomial tiene relación con el cálculo actuarial, que corresponde al que usan compañías de seguros cuando tienen que calcular las primas a cobrar por seguros con un alto número de asegurados y con una baja probabilidad de ocurrencia del siniestro. La siguiente situación ilustra uno de estos casos.

Una compañía tiene 50.000 afiliados al Seguro Obligatorio por Accidentes Personales. La compañía sabe que la probabilidad anual de muerte por accidente automovilístico es de 0,0001.

La variable aleatoria *X número de muertes accidentales anuales entre sus asegurados*, es una variable binomial de parámetros *n* = 50000 y *p* = 0,0001, cuya función de distribución es $p(x_i) = \binom{50000}{x_i} * (0,0001)^{x_i} * (0,9999)^{50000-x_i}$, pero estas probabilidades son molestas de calcular. Como este es un caso con *n grande* y *p pequeño*, es válida la aproximación de Poisson. En este ejemplo $\lambda = n*p = 50000*0{,}0001 = 5$, luego *Bin* $(50000, 0,0001)$ es aproximadamente una Poisson de parámetro 5, denotada $\mathcal{P}(5)$. Usando la distribución acumulativa, del anexo 3 (tabla A3), la compañía puede calcular la probabilidad de los siguientes sucesos.

a) ¿Cuál es la probabilidad de no pagar ningún siniestro durante el año ?

$P(X = 0) \approx F(0;5) = 0,0067 = 0,7\%$

b) ¿ Cuál es la probabilidad de pagar a lo más 3 siniestros durante ese periodo ?

$P(X \leq 3) \approx F(3;5) = 0,2650$

c) ¿ Cuál es la probabilidad de tener que pagar exactamente 5 siniestros en el año ?

$P(X = 5) \approx F(5;5) - F(4;5) = 0,6160 - 0,4405 = 0,1755$

d) ¿ Cuál es la probabilidad de pagar más de 5 siniestros en el periodo ?

$$P(X > 5) = 1 - P(X \leq 5) \approx 1 - F(5{:}5) = 1 - 0,6160 = 0,3840$$

Valores característicos.

$$E[X] = \lambda \; ; \; V[X] = \lambda$$

Esta distribución se caracteriza por el hecho que la varianza es igual que el valor esperado. No se dará un demostración formal de los valores característicos anteriores, pero sí una fundamentación.

El valor esperado de la binomial es $n*p$ que por definición es igual a λ, por lo cual debe corresponder al valor esperado de la Poisson. Por otra parte cuando p tiende a cero, entonces q tiende a uno y como la varianza de la distribución binomial es $n*p*q = \lambda*q$ y como el valor límite de q es uno resulta igual a λ, que corresponde a la varianza de la distribución de Poisson.

Observación.

En el ejemplo 7.1 el valor de λ es 5, que corresponde al número promedio de muertes anuales por accidentes automovilísticos que le suceden a la compañía de seguros.

La distribución de Poisson, además de su utilización como aproximación a la distribución binomial, sirve como modelo probabilístico de un número grande de situaciones, varias de ellas en el área biológica y agronómica. La distribución de bacterias en un cultivo, la distribución de glóbulos rojos en una muestra de sangre, la distribución de ciertas plagas de insectos en un huerto, se modelan de acuerdo a la distribución de Poisson. Otro tipo de situaciones surgen cuando los eventos ocurren a lo largo del tiempo, por ejemplo: número de camiones llegados a un centro de acopio o barcos a un puerto durante un día, número de llamadas recibidas en una central telefónica en un lapso de una hora específica o número de personas haciendo fila en un banco entre las 13:30 y 14:00. En general la distribución de Poisson se deriva del denominado ***proceso de Poisson*** que se asocia al *número de ocurrencias de un suceso A en una* ***región continua,*** que puede ser un intervalo , una superficie o un volumen, cuando la ocurrencia de *A* en un *punto* de la región es **independiente** a la ocurrencia en otro *punto*.

El proceso de Poisson presupone principalmente que:

1° el número de eventos que ocurren en regiones dis juntas son **independientes**

2° la probabilidad que un evento ocurra dos o más veces en una región pequeña es virtualmente cero.

3° el parámetro de la distribución del número de eventos que ocurre en una región dada es proporcional al tamaño de la región.

En el caso del ejemplo 7.1 el número promedio de muertes anuales por accidentes automovilísticos es 5, entonces el promedio de muertes mensuales es 5/12 o el promedio de muertes en dos años por la causal anterior es 10, por la condición tercera anterior.

Ejemplos 7.2

a) En una cierta localidad se estima que el número promedio de madrigueras de conejos que existen por hectárea es 2 y sea *X el número de madrigueras por ha*, entonces $X = \mathcal{P}\ (2)$. De la tabla A3 se obtienen los valores para calcular las probabilidades de que en un cultivo de:

1) una hectárea no haya madriguera, se determina como $P(X = 0) = F(0; 2) = 0,1353$

2) una hectárea haya exactamente 2 madrigueras, lo que corresponde a
$P(X = 2) = F(2; 2) - F(1; 2) = 0,6767 - 0,4060 = 0,2707$

3) una hectárea se encuentren menos de 3 madrigueras, es decir,
$P(X < 3) = P(\leq 2) = F(2; 2) = 0,6767$

4) una hectárea haya más de 5 madrigueras, se plantea
$P(X > 5) = 1 - P(X \leq 5) = 1 - F(5; 2) = 1 - 0,9834 = 0,0166$

5) dos hectáreas no haya madrigueras. En esta situación $Y = \mathcal{P}\ (4)$ y en consecuencia se debe utilizar una tabla para lambda igual a 4. Sin embargo, se verá como con los supuestos del proceso de Poisson se puede resolver utilizando la distribución de lambda igual a dos. Las dos hectáreas corresponden a **dos** regiones de una hectárea, en cada hectárea las ocurrencias son
independientes, de acuerdo a la condición primera anterior. Así, $P(Y = 0) = P(X = 0) * P(X = 0) = (0,1353)^2$, cuyo resultado $0,0183$ es coincidente con el valor $F(0;4)$.
6) dos hectáreas haya exactamente dos madrigueras ?
Dos madrigueras **en dos ha.** en relación al suceso por **cada** ha. puede ocurrir de varias maneras, **dos** en la primera ha. y **cero** en la segunda, o viceversa o **una** madriguera en cada ha. Luego
$P(Y = 2) = P(X = 2) * P(X = 0) + P(X = 0) * P(X = 2) + P(X = 1) * P(X = 1)$
$= 0,2707 * 0,1353 + 0,1353 * 0,2707 + (0,2707)^2 = 0,1465,$
coincidente con $F(2; 4) - F(1; 4)$.

b) Si por una parada P pasan en promedio 3 buses cada 15 minutos en forma aleatoria, este se trata de un comportamiento con distribución de Poisson de parámetro 3.
¿ Cuál es la probabilidad que un usuario que llega a P puntualmente a las 8 AM tenga que esperar por un bus

1) a lo menos 15 minutos ?
Esto significa que desde las 8 a las 8:15 no pasen buses, luego $P(X = 0/\lambda = 3) = F(0; 3) = 0,0498.$

2) a lo más 15 minutos ?
La distribución es la misma, pero se trata que en el intervalo pase por lo menos un bus, luego $P(X \geq 1/\lambda = 3) = 1 - F(0; 3) = 0,9502$

3) a lo más 5 minutos ?
En esta situación se trata de una distribución de Poisson de parámetro 1, pues el intervalo es la tercera parte del anterior, en consecuencia

$P(X \geq 1/\lambda = 1) = 1 - F(0;1) = 1 - 0,3679 = 0,6321.$

4) a lo menos 30 minutos ?

El parámetro de la distribución es 6, porque el intervalo es el doble de 15 minutos y por lo tanto $P(X = 0/\lambda = 6) = F(0;6) = 0,0025 = F(0;3) * F(0;3)$. Nótese que la probabilidad de que ello ocurra *existe*, pero es muy baja.

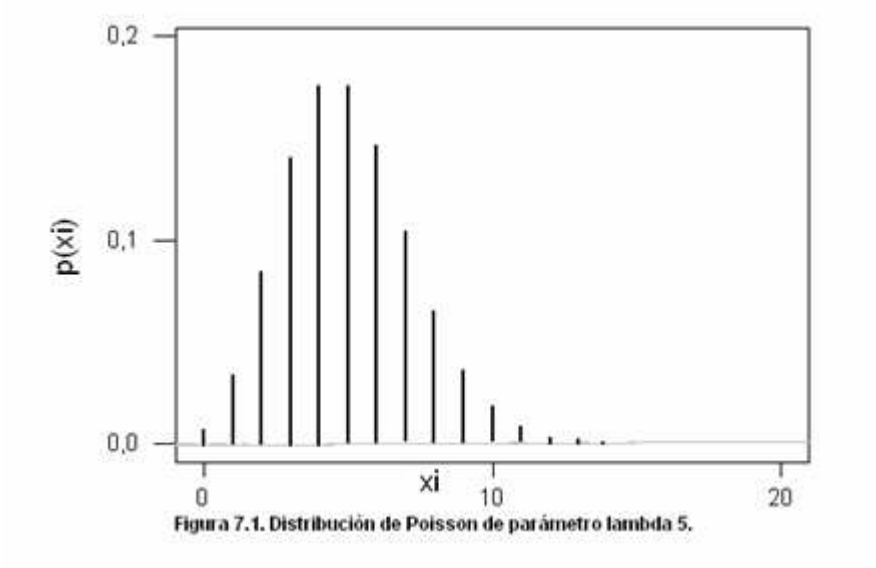


Figura 7.1. Distribución de Poisson de parámetro lambda 5.

4.8 Distribución de Pascal.

Esta distribución también se denomina *binomial negativa*, porque su distribución está asociada al desarrollo de $(1-q)^{-r}$. Se genera cuando interesa el número de observaciones necesarias para que un suceso *A* ocurra *r* veces en *n* observaciones independientes de un experimento ε. Es una generalización de la distribución geométrica, en cuyo caso $r = 1$.

Definición.

La variable aleatoria discreta X *número de veces que debe repetirse, en forma independiente, un experimento ε hasta que un suceso A, asociado al experimento y cuya probabilidad es* ***p*** *, ocurra* ***r*** *veces, tiene distribución de Pascal de parámetros* ***r*** *y* ***p***, con función de probabilidad $p(x_i) = \binom{x_i - 1}{r-1} * p^r * q^{x_i - r}$, $x_i = r, r+1, r+2, \ldots\ldots$.

Notación: $X = B^{-}(r, p)$

Valores característicos.

$E[X] = \frac{r}{p}$; $V[X] = \frac{r*q}{p^2}$

Observaciones.

1) Cada una de las $p(x_i)$ corresponde a un término del desarrollo de $p^r * (1-q)^{-r}$, de donde proviene su nombre de binomial negativa.

2) Esta distribución tiene la característica que $V[X] > E[X]$, a diferencia de la binomial en la cual $V[X] < E[X]$ y de la Poisson en que ocurre que $V[X] = E[X]$.

3) No hay tablas para esta distribución, pero para el cálculo de probabilidades se hace uso de su relación con la distribución binomial, siendo $X = B^{-}(r, p)$ e $Y = Bin(n, p)$, entonces:
$P(X \leq n) = P(Y \geq r)$ y $P(X > n) = P(Y < r)$.

Ejemplo 8.1

En un procedimiento de inspección sanitaria para la detección de plagas en plantas que se internan al país con fines de propagación deben examinarse *n* plantas. La norma establece que si **no** se detectan plantas con problemas el lote es aceptado; si se detectan hasta 2 plantas con problemas el lote es puesto en cuarentena y con 3 ó más plantas con problemas el lote es rechazado. Para un lote de cierta especie de planta la probabilidad que una planta venga con problemas es de 0,12 ¿ cuál es la probabilidad de :

a) tener que examinar 20 plantas o menos para encontrar 1 planta con problema?

Esto corresponde a una distribución de Pascal $X = \text{B}^{-}(1, 0{,}12)$ y sea la distribución asociada $Y = \text{Bin}(20, 0{,}12)$, luego $P(X \leq 20) = P(Y \geq 1) = 1 - 0,0776 = 0,9224$

b) no detectar plantas con problemas al examinar 20 plantas ?

Esto corresponde a la binomial Y, por lo cual $P(Y = 0) = (0,88)^{20} = 0,0776$

c) tener que examinar 30 plantas o menos para encontrar 1 planta con problema?

En este caso se asocia $Y = \text{Bin}(30, 0{,}12)$, luego $P(X \leq 30) = P(Y \geq 1) = 0,9784$

d) tener que examinar 30 plantas o menos para encontrar 3 plantas con problemas?

Para este caso $X = \text{B}^{-}(3, 0{,}12)$, luego $P(X \leq 30) = P(Y \geq 3) = 1 - 0,2847 = 0,7153$

Del análisis de las probabilidades anteriores se aprecia que es insuficiente examinar 20 plantas en las condiciones planteadas para decretar cuarentena, porque la probabilidad no es suficientemente alta, por lo que 30 plantas es más adecuado, aún cuando su probabilidad no es suficientemente alta para decretar rechazo.

5. DISTRIBUCIONES DE PROBABILIDAD EN EL MUESTREO DE POBLACIONES

5.1 Introducción.

El estudio del muestreo es la introducción a la teoría estadística propiamente tal. La experimentación es parte del Método Científico, mediante el cual un investigador obtiene un conjunto de datos a partir de los cuales desea obtener conclusiones válidas para un conjunto más amplio o población. El paso de lo particular a lo general se denomina inferencia inductiva y la Estadística aporta diversas metodologías para llevar a cabo este proceso, todas ellas basadas en el comportamiento de variables aleatorias en cierto tipo de muestreo.

Así, si es de interés conocer las características de una cierta población es posible, en base a experiencias previas establecer el supuesto de su comportamiento probabilístico,es decir, su distribución de probabilidad. Sin embargo, a menos que se haya censado, **sus parámetros** no serán conocidos, lo que implica que su caracterización es incompleta, porque sólo saber su comportamiento carece de valor práctico como para sacar conclusiones respecto a ella. Surge, entonces, la interrogante de cuál sería la forma para obtener información acerca de ellos. Una manera sería, como ya se mencionó, realizar un censo, pero éste es un proceso lento que incluso puede ser irrealizable y además muy oneroso. Otra forma consiste en seleccionar unos pocos elementos de la población y a partir de ellos conseguir información para los parámetros. Este último procedimiento se denomina *muestreo.* Dos ejemplos de la realidad ayudarán a ilustrar las ventajas del muestreo y la inferencia inductiva.

Ejemplos 1.1

a) una cocinera no necesita tomarse toda la sopa para saber si ésta está bien sazonada, por el contrario prueba una pizca y de lo que ahí concluya lo hace extensivo al total, es decir, realiza una *inferencia inductiva*.

b) por exigencia legal un productor de semillas de flores debe informar en el envase del porcentaje de semillas infértiles. Para obtener la información en ningún momento piensa en sembrar *todas* las semillas producidas, sino que tomará un conjunto *bien mezclado* de ellas, las pondrá a germinar y a partir del resultado sacará conclusiones. Es posible que tenga que repetir varias veces el proceso.

En ambos ejemplos no fue necesario usar toda la población para obtener conclusiones acerca de la característica de interés. Sin embargo, toda inferencia inductiva conlleva riesgo o un cierto grado de incertidumbre, pues una inferencia inductiva exacta es imposible. Una de las metodologías estadísticas, la inferencia, establece técnicas para realizar inferencias inductivas y dar una *medida*, con apoyo del cálculo de probabilidades, del grado de incertidumbre de tales inferencias, siempre que se respeten ciertos principios.

5.2 Población, muestra y tipos de muestreo.

La población corresponde a la *totalidad* de los **valores** de una característica medida en el conjunto de los individuos que son de interés en un cierto estudio y para los cuales se obtendrán las conclusiones respecto a tal característica, es decir, es el espacio muestral.

Una muestra de la población es cualquier subconjunto de ésta. Surge la cuestión, entonces, de cómo seleccionar la muestra. Dos tipos de muestras son las *muestras probabilísticas* y las

muestras no probabilísticas. En las muestras probabilísticas cada individuo tiene una probabilidad dada, habitualmente la misma probabilidad, de ser escogido. Esta forma de muestreo requiere que los individuos sean seleccionados *aleatoriamente*. En las muestras no probabilísticas los individuos son seleccionados de acuerdo al criterio del o los investigadores, basado en sus experiencias y de su supuesto conocimiento de la población en estudio. Esta forma de muestreo da, por lo general, muestras *sesgadas*. Con frecuencia se le pregunta al Estadístico como hacer para seleccionar una *muestra representativa*. De partida es imposible saber si la muestra seleccionada lo es, porque **no** se conoce lo que se quiere representar, esto es, la población. Sin embargo el único procedimiento que garantiza, con algún grado de certeza conocido, seleccionar una muestra *representativa* es la *aleatorización.*

Cuando el muestreo se aplica a poblaciones pequeñas o relativamente grandes se le denomina *muestreo en poblaciones finitas*. Existen varios tipos entre los cuales los más importantes y de mayor uso son: el muestreo aleatorio simple, el muestreo estratificado, el muestreo por conglomerados y el muestreo sistemático.

Muestreo aleatorio simple.

Es el muestreo más sencillo de todos y consiste en que la elección de los individuos de la población se realiza en forma irrestricta, de modo que cada individuo tiene la misma probabilidad de ser seleccionado en la muestra.

Muestreo estratificado.

Se aplica cuando en la población existen claramente identificados dos o más subpoblaciones o estratos de interés para el estudio a realizar y se quiere asegurar una muestra con una cantidad de individuos de cada estrato en relación al tamaño de éste. Por lo general, en cada estrato se realiza un muestreo aleatorio simple. Ejemplos de estratos son: clases socioeconómicas (ABC1, C2, C3, D, E) ; sexo (hombres, mujeres).

Muestreo por conglomerado.

Hay situaciones en los cuales la población está conformada por conglomerados que son grupos de individuos que tienen la particularidad de estar muy cercanos unos a otros. Cuando establecer una lista de todos los individuos resulta muy difícil o cuando una selección aleatoria de estos implicara tener observaciones que podrían quedar muy distantes una de otras, lo que resultaría muy costoso, es posible seleccionar primero conglomerados, en forma aleatoria, y dentro de estos a los individuos de interés para el estudio.

Si los individuos son heterogéneos dentro del conglomerado se observan varios o todos sus componentes, de lo contrario si son muy homogéneos basta con pocas observaciones.

Por ejemplo para estimar el ingreso promedio por hogar en el Gran Santiago, resulta muy conveniente seleccionar manzanas (conglomerados homogéneos) y dentro de estas diferentes hogares, pues es más fácil tener un mapa con las diferentes manzanas que un listado de todos los hogares. Además, es menos costoso encuestar dentro de la manzana que muchos hogares repartidos por toda la ciudad.

Un manzano es un conglomerado de frutos si lo que se necesita es medir la infestación por polilla de la manzana. En inspecciones sanitarias para detectar presencia de insectos cuarentenarios en fruta de exportación el *palet* es un conglomerado de cajas.

Muestreo sistemático.

Consiste en realizar la elección de los individuos en forma sistemática a intervalos regulares, en el espacio o el tiempo, hasta obtener el número de individuos necesarios para la muestra, donde el primer seleccionado fue elegido al azar. Por la razón descrita éste no es propiamente un muestreo probabilístico por lo que se dice que es un muestreo seudoaleatorio. Se utiliza por razones prácticas de selección. Por ejemplo, si se necesita estimar el porcentaje de fruta de descarte por defectos o daños de insectos en una exportadora, una forma práctica de hacerlo consiste en seleccionar fruta en la línea de embalaje (correa transportadora) a intervalos de tiempo iguales hasta conseguir un número adecuado de frutos.

En este tipo de muestreo se corre el riesgo de obtener muestras sesgadas cuando existen periodicidades dentro de la población.

El muestreo propio de la inferencia estadística no corresponde a ninguno de los anteriores, aunque se parece bastante al primero. Su diferencia radica en que se trata de un muestreo en *poblaciones infinitas*, lo cual puede resultar extraño, pero que se puede explicar porque se trata de muestras de variables aleatorias y en teoría a una variable aleatoria se le pueden realizar infinitas observaciones.

Muestreo aleatorio simple (m.a.s) en poblaciones infinitas.

Supóngase que se desea establecer la distribución de la población de alturas de las chilenas adultas. Situaciones previas han permitido establecer que el comportamiento de las alturas en poblaciones grandes tiene aproximadamente forma acampanada. Se puede en consecuencia hacer el supuesto de normalidad de las alturas de la población de interés ¿ pero qué se sabe de sus parámetros media y varianza ? En la realidad casi nada, a lo más, una idea vaga del promedio de las alturas. Por lo tanto se debe obtener información para determinar valores de los parámetros.

Una manera de proceder sería medirle las alturas a **todas** las chilenas adultas, es decir, censarlas. Con tales datos, si es que no hay errores de medición, se calculan los *verdaderos* valores de μ y σ^2. Es fácil darse cuenta de las dificultades, tiempo y costo son las principales, de llevar a cabo tal proyecto.

Otra manera consiste en obtener la información mediante una muestra aleatoria.

Si el propósito es comparar con la población de alturas de las mujeres estadounidenses adultas, se debe repetir el procedimiento de medirles las alturas a una muestra de esta otra población, con el fin de obtener información de los nuevos parámetros para esta otra distribución normal.

Definición.

Una muestra aleatoria simple de la variable aleatoria X es un conjunto de *n* observaciones **independientes** $X_1, X_2, X_3, ..., X_n$ de X, todas ellas con la misma distribución de probabilidad. El número natural ***n*** recibe el nombre de **tamaño de la muestra**.

Observaciones.

1) Conceptualmente una muestra aleatoria consiste en *n* observaciones repetidas de X **todas** realizadas bajo condiciones idénticas, pero como en la práctica esto es imposible, hay que contentarse con que las condiciones sean similares y las variaciones irrelevantes.

2) Por la condición de independencia de las observaciones, si X variable aleatoria con función de distribución $p(x_i)$ o $f(x)$, según sea discreta o continua, entonces la función de distribución conjunta, g , de la muestra $\{X_1,\ X_2, X_3, ..., X_n\}$ de X es, respectivamente:

$$g(X_1,\ X_2, X_3, ..., X_n) = \prod_{i=1}^{n} p(x_i)$$

$$g(X_1,\ X_2, X_3, ..., X_n) = \prod_{i=1}^{n} f(x)$$

3) Como cada variable aleatoria X_i tiene la misma distribución que X, entonces:

$$E[X_i] = E[X] = \mu \quad \text{y} \quad V[X_i] = V[X] = \sigma^2\ ,\ i = 1, 2, 3, ...n$$

5.3 Estadígrafos.

El estadígrafo es un elemento muy importante en estadística, porque se refiere a un resultado obtenido a partir de las observaciones muestrales.

Definición.

Sea $\{X_1,\ X_2, X_3, ..., X_n\}$ una muestra aleatoria de X y $\{x_1, x_2, x_3, ..., x_n\}$ los valores observados en la muestra obtenida, entonces se llama **estadígrafo** a una función real H de la muestra y **valor del estadígrafo** a la función H de los valores observados.

Observaciones.

1) Un estadígrafo es una variable aleatoria $Y = H(X_1,\ X_2, X_3, ..., X_n)$, mientras que el valor del estadígrafo es un número real $y = H(x_1, x_2, x_3, ..., x_n)$. Por ejemplo, en una población de **pesos de manzanas** se tomará una m.a.s tamaño 3, es decir, se seleccionarán **tres** manzanas para medirle su peso representados por las variables aleatorias X_1, X_2 y X_3 , donde las tres observaciones tienen la misma distribución de la población, representada por la variable aleatoria X. Considérese como estadígrafo el promedio de la muestra, es decir, $Y = (\sum_{i=1}^{3} X_i)/3$. El valor de Y es desconocido mientras no se seleccionen las manzanas y se pesen, luego es una variable aleatoria. Suponga que los pesos de las manzanas seleccionadas resultaron ser 165 gr , 142 gr y 155 gr. respectivamente, por lo tanto el valor del estadígrafo Y es $y = (\sum_{i=1}^{3} x_i)/3 = (165 + 142 + 155)/3 = 154$ gr.

2) Un estadígrafo puede ser cualquier función de la muestra. Algunos posibles estadígrafos son: la media muestral, la varianza muestral, mínimo muestral, máximo muestral, rango muestral, mediana muestral, proporción muestral, y así muchos otros. Todos conceptualmente equivalentes a lo visto en descriptiva, con el apelativo de muestral para diferenciarlos de los parámetros respectivos.

De los anteriores los más importantes son la media, la varianza y la proporción muestral, por sus propiedades y su vinculación a la distribución normal.

Media o promedio muestral.

A la media poblacional, como se describió en estadística descriptiva o como valor esperado de una variable aleatoria, se asocia la media muestral, simbolizada por $\overline{X}$, según la siguiente

función de la muestra aleatoria simple: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. Esta definición tiene como consecuencias varias propiedades importantes para el promedio muestral.

Teorema 3.1.

Sea X una variable aleatoria con $E[X] = \mu$ y $V[X] = \sigma^2$ y $\overline{X}$ la media de una muestra tamaño *n* de X, entonces $E[\overline{X}] = \mu$ y $V[\overline{X}] = \sigma^2/n$.

Demostración.

Usando propiedades del valor esperado y la consecuencia 3) de la definición de muestra aleatoria simple

$E[\overline{X}] = E[\frac{\sum_{i=1}^{n} X_i}{n}] = \frac{1}{n}\sum_{i=1}^{n}(E[X_i]) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}*(n*\mu) = \mu = E[X]$.

Usando propiedades de la varianza y de la condición de independencia de las X_i de la definición de muestra aleatoria simple

$V[\overline{X}] = V[\frac{\sum_{i=1}^{n} X_i}{n}] = \frac{1}{n^2}\sum_{i=1}^{n}(V[X_i]) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}*(n*\sigma^2) = \sigma^2/n = \frac{V[X]}{n}$.

Notación : $E[\overline{X}] = \mu_{\overline{X}}$; $V[\overline{X}] = \sigma^2_{\overline{X}}$.

El teorema anterior es válido cualquiera sea la forma de la distribución de X y es un resultado de enorme importancia para la estadística inferencial, como se verá en ese capítulo, principalmente porque demuestra que si el tamaño de la muestra **crece**, la magnitud de la varianza de la media muestral **decrece** en proporción inversa al ***n*** .

Ejemplos 3.1

a) sea $X = N(22, 40)$ y $\overline{X}$ la media de una muestra tamaño 10 de X, entonces $E[\overline{X}] = 22$ y $V[\overline{X}] = 40/10 = 4$.

b) si $X = Bin\ (n\ ,\ p)$ se recordará que $E[X] = n*p$ y que $V[X] = n*p*(1\text{-}p)$, luego $E[\overline{X}] = n*p$ y $V[\overline{X}] = \frac{n*p*(1-p)}{n} = p*(1\text{-}p)$.

Varianza muestral.

Es una medida de variabilidad de los datos de la muestra, en forma similar a la varianza poblacional y por ser un estadígrafo corresponde a una función de la muestra aleatoria, según la siguiente definición: $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$, donde (n - 1) recibe el nombre de grados de libertad.

Observaciones.

1) Nótese la semejanza con la definición de varianza en descriptiva, es decir, suma de desvíos (respecto a la media muestral) al cuadrado, dividido por (n - 1). El denominador en el cálculo de un **estimador de varianza** siempre se llaman grados de libertad.

2) La división por los grados de libertad (n - 1), en vez de n, es necesaria, porque es deseable que suceda que $E[S^2] = \sigma^2$, como se demostrará.

Teorema 3.2.

Sea S^2 la varianza muestral definida como antes, entonces:

1. $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{n-1}$

2. $E[S^2] = \sigma^2$

Demostración.

1) $S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = \frac{1}{n-1}\sum(X_i^2 - 2*X_i*\overline{X} + \overline{X}^2) = \frac{1}{n-1}[\sum X_i^2 - 2*\overline{X}*\sum X_i + \sum \overline{X}^2]$

$= \frac{1}{n-1}[\sum X_i^2 - 2*\overline{X}*(n*\overline{X}) + n*\overline{X}^2] = \frac{1}{n-1}[\sum X_i^2 - n*\overline{X}^2] = \frac{1}{n-1}\left[\sum X_i^2 - n*\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2\right]$

$= \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{n-1}$.

2) $E[S^2] = E\left[\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1}E\left[\sum\left((X_i - \mu) - (\overline{X} - \mu)\right)^2\right]$

$= \frac{1}{n-1}E\left[\sum(X_i - \mu)^2 - n*(\overline{X} - \mu)^2\right] = \frac{1}{n-1}\left[\sum(E(X_i - \mu)^2) - n*E(\overline{X} - \mu)^2\right]$

$= \frac{1}{n-1}\left[\sum(\sigma^2) - n*\sigma^2_{\overline{X}}\right] = \frac{1}{n-1}\left[n*\sigma^2 - n*\frac{\sigma^2}{n}\right] = \frac{1}{n-1}*(n-1)*\sigma^2 = \sigma^2$.

5.4 Distribución de las muestras de una población normal.

Con el fin de entender plenamente el concepto de distribución muestral hay que tener presente las siguientes consideraciones. Una distribución hace referencia a una población, en este caso a una población de muestras aleatorias. El caso es que se tiene originalmente una población, que es la que interesa conocer a través de **una muestra**. La consideración que se hace, entonces, es que a partir de un tamaño muestral *n* dado, se genera una población *teórica* correspondiente a **todas las posibles muestras tamaño *n*** que se pueden obtener de la población de interés. La distribución de la población *teórica* de las medias es la que se va desarrollar, aún cuando para los propósitos de investigación será necesario tener sólo **una** muestra de esa población.

En el desarrollo de este tema es necesario hacer uso del teorema que establece propiedades reproductivas de la distribución normal, el que establece que cualquier función en primer grado de una variable aleatoria con distribución normal también será normal, el que se enunciará sin demostración.

Teorema 4.1.

1. Sea $X = N(\mu\,,\,\sigma^2)$ y la función de X, $Y = aX + c$, entonces la variable aleatoria Y tiene **distribución normal** con media $\mu_Y = a*\mu + c$ y varianza $\sigma_Y^2 = a^2*\sigma^2$.

2. Sea $\{X_i = N(\mu_i\ ,\ \sigma_i^2)\ ,\ i = 1, 2, 3, ..., k\}$ un conjunto de k variables aleatorias normales e independientes y sea $Y = \sum_{i=1}^{k} X_i$, entonces la variable aleatoria Y **tiene distribución normal** con media $\mu_Y = \sum_{i=1}^{k} \mu_i$ y varianza $\sigma_Y^2 = \sum_{i=1}^{k} \sigma_i^2$.

Consecuencia.

Dado que $\overline{X}$ corresponde a una suma de variables aleatorias normales, en virtud del teorema anterior su distribución es normal y en virtud del teorema 3.1 su valor esperado es μ y su varianza σ^2/n , por lo tanto si $\boldsymbol{X = N(\mu\,,\,\sigma^2)\ \Rightarrow\ \overline{X} = N(\mu\,,\,\sigma^2/n)}$. Del resultado anterior se deduce que la transformación $\boldsymbol{Z}$ de $\overline{X}$, $\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$, tiene distribución normal típica, al igual que la transformación $\boldsymbol{Z}$ de X, $\frac{X-\mu}{\sqrt{\sigma^2}}$. Con este resultado es posible calcular probabilidades de sucesos vinculados a la media muestral de una población normal a partir de una distribución normal estándar.

Ejemplos 4.1.

a) Sea $X = N(22, 40)$ y $\overline{X}$ la media de muestra tamaño 10 de X, entonces $\overline{X} = N(22, 4)$. El gráfico de ambas distribuciones se ilustra en la figura 4.1, en la cual el área en blanco es común a ambas distribuciones y por lo tanto las áreas sombreadas en una y otra son iguales.

b) A partir de la distribución $\overline{X} = N(22, 4)$, se calculan las siguientes probabilidades:

- $P(\overline{X} \leq 19) = P(\frac{\overline{X}-\mathbf{22}}{\sqrt{\mathbf{4}}} \leq \frac{19-22}{2}) = P(Z \leq\ -1,5) = \phi(-1,5) = 0,0668$
- $P(21 \leq \overline{X} \leq 24) = P(-0,5 \leq Z \leq 1,0) = \phi(1,0) - \phi(-0,5) = 0,8413 - 0,3085 = 0,5328$
- $P(\overline{X} \geq 23,5) = P(Z \geq 0,75) = 1 - \phi(0,75) = 1 - 0,7734 = 0,2266$

Los resultados anteriores significan que, bajo las condiciones enunciadas, sólo el 6,68% de las muestras dará valores promedios menores a 19, es decir, 3 unidades por debajo de la media poblacional; que el 53,28% de las muestras entregará valores promedios entre 21 y 24 y que el 22,66% de las muestras dará como resultado promedios por sobre 23,5. Estos resultados sirven para **determinar que tan probable** resultará que una media muestral tenga la **aproximación deseada** respecto a la media poblacional que es la que interesa conocer. Se puede apreciar a través del gráfico anterior, o realizando los cálculos respectivos, que las probabilidades de alejarse del valor central son bastante menores para el promedio muestral que las que les correspondería a cualquier valor poblacional X. Por el contrario la probabilidad

de una $\overline{X}$ de estar "alrededor" del valor central es bastante mayor que para un valor poblacional, como consecuencia de la **menor varianza** de la población de $\overline{X}$.

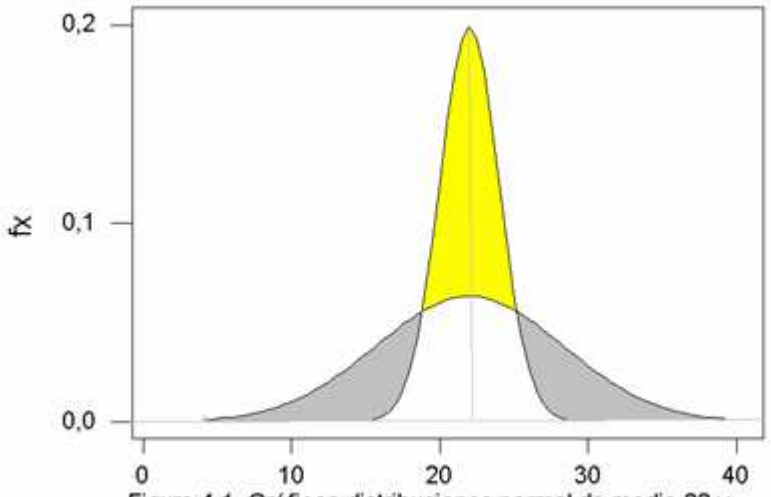


Figura 4.1. Gráficos distribuciones normal de media 22 y varianza 40 y de la media de muestras tamaño 10.

c) De un criadero donde el peso de los cerdos tiene distribución normal con media 82 kg y varianza 25, se toma una muestra de 16 cerdos seleccionados al azar. ¿Cuál es la probabilidad que el peso promedio obtenido de la muestra :

- sea menor a 80 kg ?

Primero es conveniente establecer que $\overline{X} = N(82,\ 25/16) \ \Rightarrow \frac{\overline{X}-82}{5/4} = N(0\ ,\ 1)$. luego $P(\overline{X} < 80) = P(Z < \frac{80-82}{1,25}) = \phi(-1,6) = 0,0548$, o sea, la probabilidad de obtener con la muestra una media menor a 80 es de 0,0548

- tenga una diferencia de $\pm$ 1 kg respecto a μ ?

Esto se plantea $P(-1 \leq \overline{X} - \mu \leq 1) = P(-\frac{1}{1,25} \leq \frac{\overline{X}-\mu}{1,25} \leq \frac{1}{1,25}) = P(-0,8 \leq Z \leq 0,8)$

$= \phi(0,8) - \phi(-0,8) = 0,7881 - 0,2119 = 0,5762,$

es decir, la probabilidad de obtener de la muestra una media que difiera de μ en a lo más 1 kg es de 0,5762.

d) En las condiciones del ejemplo anterior, determinar:

- los valores a y b equidistante de μ tal que $P(a \leq \overline{X} \leq b) = 0,95$, en consecuencia

$P(a \leq \overline{X} \leq b) = 0,95 \Rightarrow P(\frac{a-82}{1,25} \leq \frac{\overline{X}-82}{1,25} \leq \frac{b-82}{1,25}) = 0,95 \Rightarrow P(\frac{a-82}{1,25} \leq Z \leq \frac{b-82}{1,25}) = 0,95$

Para que la probabilidad anterior se dé, $\frac{a-82}{1,25}$ debe corresponder al percentil 0,025 de la Z y $\frac{b-82}{1,25}$ al percentil 0,975, por lo tanto $\frac{a-82}{1,25}$ = - 1,96 y $\frac{b-82}{1,25}$ = 1,96. Despejando *a* y *b* se obtiene que *a* = 79,55 kg y *b* = 84,45 kg.

- el valor de c tal que $P(\overline{X} > c) = 0,05$, luego

$P(\overline{X} > c) = 0,05 \Rightarrow P(Z > \frac{c-82}{1,25}) = 0,05 \Rightarrow 1 - \phi(\frac{c-82}{1,25}) = 0,05 \Rightarrow \phi(\frac{c-82}{1,25}) = 0,95$

$\Rightarrow \frac{c-82}{1,25} = \phi^{-1}(0,95) = 1,645 \Rightarrow c = 84,06$ kg.

e) ¿ Cuál será el tamaño de muestra necesario a tomar de la población de pesos de los cerdos para que la probabilidad de obtener una media mayor a 83 sea de 0,10 ?

La distribución de las medias muestrales es $\overline{X} = N(82,\ 25/n)$, luego $P(\overline{X} > 83) = 0,10$ $\Rightarrow P(Z > \frac{1}{5/\sqrt{n}}) = 0,10 \Rightarrow 1 - \phi(\frac{\sqrt{n}}{5}) = 0,10 \Rightarrow \phi(\frac{\sqrt{n}}{5}) = 0,90 \Rightarrow \frac{\sqrt{n}}{5} = 1,28 \Rightarrow n = 41.$

Por lo tanto para que se cumpla la probabilidad deseada la muestra debe corresponder a 41 cerdos del criadero, seleccionados al azar. Observe que el resultado aritmético es 40,96, pero *n* debe ser un número natural, luego se aproxima a 41. En cálculos de tamaño de muestra el criterio, cuando el resultado es decimal, es **siempre** aproximar hacia arriba.

5.5 Distribuciones que incluyen a la varianza muestral de una población normal.

Con la varianza muestral o con la combinación de la varianza con la media muestral resultan tres distribuciones de enorme importancia, en especial para la inferencia estadística.

Distribución ji cuadrada.

Karl Pearson, destacado Estadístico británico, con el fin de aportar un enfoque estadístico al estudio de la herencia y la evolución biológica es su creador, así como del concepto de correlación lineal . Tiene múltiples aplicaciones y una muy importante en el área de la genética.

Pearson estableció que el estadígrafo $D^2 = \frac{(n-1)*S^2}{\sigma^2}$ tiene la distribución denominada ji cuadrada (chi-square) con (n $-$ 1) grados de libertad, correspondiente a los de S^2.

Notación: $D^2 = \frac{(n-1)*S^2}{\sigma^2} = \chi^2(n-1)$; Notación percentil alfa : $\chi^2_\alpha(n-1)$

Observaciones.

1) La distribución tiene por representación una curva como la de la figura 5.1 , donde grados de libertad (g.l) es el parámetro de la distribución. A medida que el valor del parámetro (g.l) aumenta la moda de la distribución aumenta, es decir, el máximo de la curva se desplaza hacia la derecha.

2) La función de distribución tiene una expresión matemática bastante complicada, razón por la cual el área acumulada bajo la curva, desde 0 hasta un valor d > 0 , está tabulada para diferentes grados de libertad, desde 1 hasta 45 o más según la tabla utilizada, y para diferentes valores percentiles :0,005 ; 0,01 ; 0,025 ; 0,05 ; 0,10 ; 0,25 ; 0,50 y sus complementarios. La tabla, del anexo 4 (A4), corresponde a los valores percentiles de distribuciones ji cuadrado con distintos grados de libertad. Cada línea se refiere a la distribución ji cuadrada con los grados de libertad indicados y donde cada columna corresponde a los percentiles *p* convencionales, ya mencionados

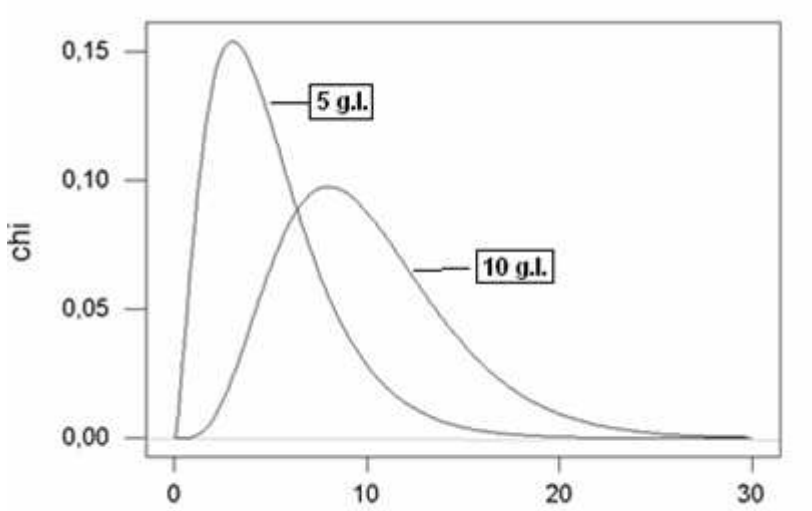


Figura 5.1. Gráficos de distribuciones ji cuadrada con 5 y 10 grados de libertad

Ejemplos 5.1

a) Determinar por tabla los valores percentiles indicados:

- $\chi^2_{0,05}(15) = 7,261$; $\chi^2_{0,90}(15) = 22,307$ los que se encuentran en la línea 15 columnas 0,05 y 0,90 respectivamente

- $\chi^2_{0,05}(9) = 3,325$; $\chi^2_{0,90}(9) = 14,684$ se buscan en las mismas columnas en línea 9. Observe que al aumentar los grados de libertad los valores percentiles son mayores concordante con la observación 1 anterior.

- $\chi^2_{0,05}(12) = 5,226$; $\chi^2_{0,95}(12) = 21,026$ los valores percentiles de la ji cuadrado son siempre positivos (ver figura 5.1) a diferencia de lo que ocurre en la normal estándar.

b) Obtener las probabilidades pedidas para el estadígrafo $D^2 = \chi^2(20)$.

De la línea 20 de la tabla se determina que:

- $P(D^2 < 10,85) = 0,05$ porque 10,85 es el percentil 0,05 de la distribución de $\chi^2(20)$.

- $P(D^2 > 28,41) = 1 - 0,90 = 0,10$, porque 28,41 corresponde al percentil 0,90.

- $P(9,59 \leq D^2 \leq 34,17) = 0,975 - 0,025 = 0,95$, porque 34,17 y 9,59 son los percentiles 0,975 y 0,025 respectivamente.

En el siguiente teorema se enunciarán, sin demostración, las propiedades reproductivas de la distribución ji cuadrada que son de interés.

Teorema 5.1.

1. Sean $\{D_i^2 = \chi^2(n_i)$, i $= 1, 2, 3.....k\}$ k variables ji cuadradas **independientes**, con n_i grados de libertad cada una,entonces la variable aleatoria $Y = \sum_{i=1}^{k} D_i^2$ tiene **distribución ji cuadrada** con $n = \sum_{i=1}^{k} n_i$ grados de libertad.

2. Sean $\{Z_i = N(0 \ , \ 1)$, i $= 1, 2, 3, ..., k\}$ k variables normales típicas **independientes**, entonces la variable aleatoria $Y = \sum_{i=1}^{k} Z_i^2$ tiene **distribución ji cuadrada** con k grados de libertad.

Observaciones.

1) El teorema establece que una suma de variables ji cuadradas independientes también tiene distribución ji cuadrada con grados de libertad la suma de los grados de libertad de cada una.
2) Además, demuestra que la variable aleatoria que resulta de sumar variables normales típicas independientes al cuadrado tiene distribución ji cuadrada y como consecuencia se deduce que **una normal típica al cuadrado** tiene distribución **ji cuadrada con un grado de libertad**.

Distribución t de Student.

Esta distribución se debe al Estadístico inglés William Sealey Gosset, químico de formación, alumno y colaborador de Karl Pearson, de quien se cuenta que publicó sus primeros trabajos bajo el seudónimo de *Student*, porque temía ser despedido si alguno de sus jefes, en la fábrica de cerveza Guiness donde trabajaba como químico, descubriera que realizaba investigaciones en estadística. La verdad es otra, pero lo importante es su contribución a la estadística.

En los inicios del siglo pasado Gosset estableció que el estadígrafo $t = \frac{\overline{X} - \mu}{\sqrt{S^2/n}}$ tiene distribución *t* con (n $-$ 1) grados de libertad, correspondiente a los de S^2.

Notación: $t = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \boldsymbol{t\ (n-1)}$; Notación percentil alfa : $\boldsymbol{t_\alpha (n-1)}$.

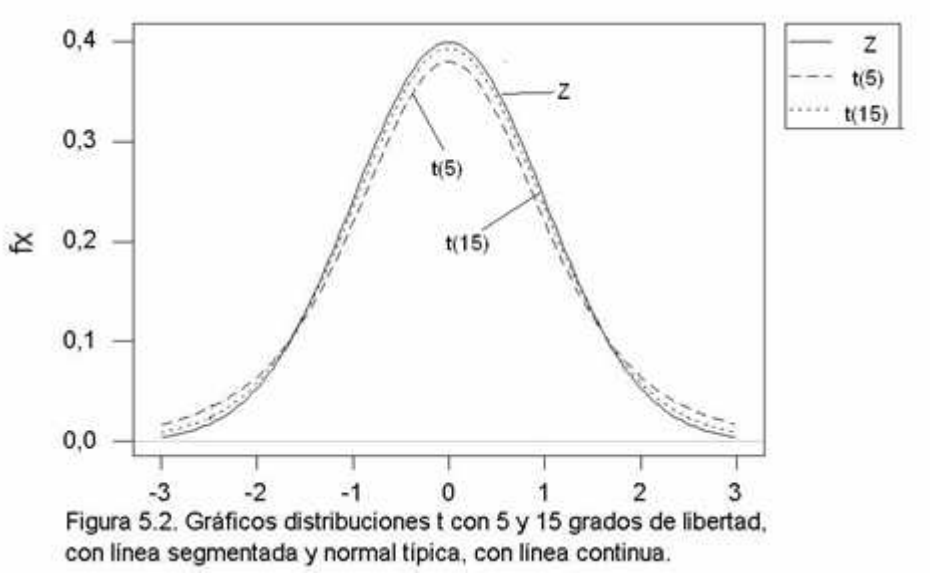


Figura 5.2. Gráficos distribuciones t con 5 y 15 grados de libertad, con línea segmentada y normal típica, con línea continua.

Observaciones.

1) Grados de libertad es el parámetro de la distribución, igual como ocurre con la ji cuadrada, pues de hecho esta distribución es consecuencia del cociente entre una normal estándar y la raíz aritmética de una ji cuadrada dividida por sus grados de libertad, ambas independientes entre sí.

2) La curva de la distribución *t de Student* es acampanada centrada en 0, similar a la normal estándar, pero con "colas más pesadas", o sea, encierran una mayor área, por lo que sus valores percentiles son mayores que los de Z, lo que implica mayor variabilidad. Esto parece intuitivamente razonable, porque se diferencia con el estadígrafo Z en que en el denominador en vez del **parámetro** σ^2 aparece la **varianza muestral** S^2 que es un estadígrafo. También se cumple que $\lim_{n\to\infty} t(n-1) = N(0\,,\,1)$, como se ilustra en la figura 5.2.

3) La función de distribución tiene una expresión matemática más complicada que la de Z, razón por la cual el área acumulada bajo la curva, desde 0 hasta un valor $t > 0$, está tabulada para diferentes grados de libertad, *n*, desde 1 hasta 90 o más según la tabla utilizada, y para diferentes valores percentiles : 0,75 ; 0,90 ; 0,95 ; 0,975 ; 0,99 ; 0,995 . Los valores percentiles complementarios solo se diferencian en el signo, pues son negativos, tal como ocurre en la distribución normal estándar. La tabla, del anexo 5 (A5), corresponde a percentiles de distribuciones t de Student, con distintos grados de libertad.

El uso de la tabla es similar al de la ji cuadrada con la diferencia que sólo aparecen los percentiles superiores debido a la simetría de la distribución, porque percentiles complementarios inferiores solamente cambian su signo a negativo.

Ejemplos 5.2

a) Determinar por tabla los valores percentiles complementarios indicados:

- $t_{0,90}(10) = 1,3722$; $t_{0,10}(10) = -1,3722$ los que se obtienen de la línea 10
- $t_{0,95}(5) = 2,0150$; $t_{0,05}(5) = -2,0150$ los que se obtienen de la línea 5.
- $t_{0,95}(24) = 1,7109$; $t_{0,05}(24) = -1,7109$ los que se obtienen de la línea 24.

Observe que al aumentar los grados de libertad los valores percentiles disminuyen, lo que se puede constatar al leer los valores hacia abajo en una misma columna. Para grados de libertad grandes, mayores a 90, los valores percentiles son bastante cercanos al de la normal típica como se puede verificar comparando con la última fila del cuadro 5.2.

b) Obtener las probabilidades pedidas para el estadígrafo $t = t\,(9)$:
De la línea 9 de la tabla se determina que:
- $P(t < 1,8331) = 0,95$ porque 1,8331 es el percentil 0,95 de la distribución de t
- $P(t > 1,3830) = 1 - 0,90 = 0,10$, porque 1,3830 corresponde al percentil 0,90
- $P(t \leq -0,7027) = 0,25$, porque -0,7027 es el percentil 0,25, complementario a 0,75
- $P(-1,3830 < t \leq 2,2622) = 0,975 - 0,10 = 0,875$
- $P(-2,2622 \leq t \leq 2,2622) = 0,975 - 0,025 = 0,95$, porque 2,2622 es el percentil 0,975

Distribución $\mathbb{F}$ de Snedecor-Fisher.

Esta distribución es conocida gracias al matemático y físico estadounidense George W. Snedecor quien la bautizó de este modo en reconocimiento al notable matemático, estadístico y genetista inglés Ronald A. Fisher, quien la había estudiado anteriormente en 1924 y con quien trabajaron en conjunto. La distribución es el resultado del cociente entre dos variables aleatorias independientes con distribución ji cuadrada, cada una dividida por sus correspondientes grados de libertad, *m* la del numerador y *n* la del denominador.

Si $U = \chi^2(m)$ y $V = \chi^2(n)$ con U y V independientes, entonces $F = \frac{U/m}{V/n}$ tiene distribución $\mathbb{F}$ con m y n grados de libertad en el numerador y denominador respectivamente.

Notación: $F = \mathbb{F}(m, n)$; Notación percentil alfa : $\mathbb{F}_{\alpha}(m, n)$.

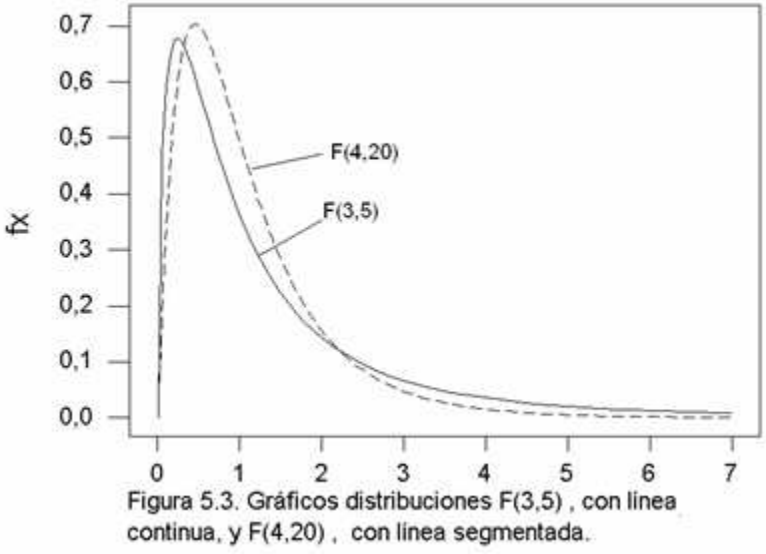


Figura 5.3. Gráficos distribuciones F(3,5) , con línea continua, y F(4,20) , con línea segmentada.

Observaciones.

1) Grados de libertad son los dos parámetros de la distribución .

2) La curva de la distribución parte de 0 y tiene una forma algo parecida a la de ji cuadrada, pero en este caso su moda se aproxima al valor 1 a medida que ambos grados de libertad aumentan. (Figura 5.3).

3) La función de distribución está tabulada para diferentes grados de libertad del numerador y denominador, y para diferentes valores percentiles : 0,90 ; 0,95 ; 0,975 ; 0,99 ; 0,995 . La tabla A6, del anexo 6, es una tabla de distribuciones $\mathbb{F}$. El uso de esta distribución, por lo general, es para los percentiles superiores. Si se necesitara algún percentil inferior se puede hacer uso de la siguiente relación $\mathbb{F}_{1\text{-}\alpha}(n\,,\,m) = 1/\mathbb{F}_{\alpha}(m\,,\,n)$.

4) Si $U = \frac{m*S_1^2}{\sigma_1^2} = \chi^2(m)$ y $V = \frac{n*S_2^2}{\sigma_2^2} = \chi^2(n)$, entonces por definición $F = \frac{U/m}{V/n} = \frac{m*S_1^2/m*\sigma_1^2}{n*S_2^2/n*\sigma_2^2} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \mathbb{F}(m\,,\,n)$. Este resultado será de gran utilidad y uso en desarrollos estadísticos futuros.

Ejemplos 5.2

Obtener los valores percentiles indicados (m se busca en la columna y n en la fila) :

- $\mathbb{F}_{0,95}(4\,,10) = 3,4780$ y $\mathbb{F}_{0,95}(10\,,\,4) = 5,9644$, son valores muy diferentes.

- $\mathbb{F}_{0,90}(5\,,\,5) = 3,4530$; $\mathbb{F}_{0,975}(5\,,\,5) = 7,1464.$

- $\mathbb{F}_{0,95}(3\,,\,12) = 3,4903$; $t_{0,05}(12\,,3) = 1/3,4901 = 0,2865$.

6. INFERENCIA ESTADISTICA PARA MEDIAS Y VARIANZAS

6.1 Introducción.

La inferencia estadística es una parte de la Estadística que comprende los métodos y procedimientos adecuados para deducir características de una **población** a partir de muestras aleatorias, en forma científicamente válidas, cuyo fin ***es obtener conclusiones*** respecto a ésta, sujetas a una **duda razonable** mediante la asignación de una **medida objetiva.**

La inferencia comprende dos aspectos. la estimación de parámetros y el contraste de hipótesis estadísticas.

6.2 Estimación de parámetros.

Un parámetro, como se recordará, representa un valor poblacional y por lo tanto es una constante. El valor de un parámetro se obtiene a través de un censo, lo que es posible de hacer cuando las poblaciones son finitas, pero en el caso de la inferencia el tipo de poblaciones que se estudian se consideran que son infinitas. En consecuencia la única vía de conseguir una *imagen* del parámetro es a través de muestras. Para que una muestra tenga validez estadística ésta debe ser aleatoria y simple, en los términos definidos en la unidad Distribuciones Muestrales. Una muestra aleatoria simple (m.a.s) permite obtener un *estimador* del parámetro de interés, esto es, un valor muestral o *estadígrafo* que estará "cercano" en alguna medida al valor del parámetro.

<u>Estimación puntual</u>.

Se llama **estimador puntual** de un parámetro a un estadígrafo que cumple con lo anterior. Sin embargo proporciona una imagen algo imprecisa del parámetro, puesto que una vez calculado el valor del estimador a partir de las observaciones muestrales, sólo se puede confiar en que éste esté "cercano" al del parámetro. Por ejemplo, si para estimar el peso promedio de una población de hombres adultos, una muestra aleatoria simple entrega una media $\overline{X}$ igual a 66,3 kg, la imagen que se puede asociar es que el verdadero peso promedio de las personas estará "alrededor" de ese valor ¿ cuán cercano?, imposible establecerlo.

Pueden existir muchos estimadores para un mismo parámetro, por lo tanto hay que establecer ciertos criterios que permita elegir de entre ellos al que sea el **mejor**, en el sentido de que tenga la mayor capacidad de entregar un valor cercano al de él.

Algunas propiedades que caracterizan a un buen estimador $\widehat{\theta}$ del parámetro θ se explican a continuación.

1° **Insesgamiento**, que consiste en que $E(\widehat{\theta}) = \theta$, lo que significa que e l valor "promedio" del estimador se distribuye alrededor del valor del parámetro θ.

2° **Eficiencia o precisión**, que consiste en tener la menor varianza entre los estimadores insesgados de θ, es decir, que de todos los estimadores $\widehat{\theta}$ que cumplan la propiedad anterior se debe preferir aquel cuya distribución tenga la menor variabilidad . De esta manera se asegura una alta probabilidad de que el valor de $\widehat{\theta}$ estará más cercano al de θ.

3° **Consistencia**, es decir, que en la medida que el tamaño de la muestra crezca el valor de $\widehat{\theta}$ estará cada vez más próximo al del parámetro θ. Esta es una propiedad asintótica.

4° **Suficiencia**, cuando el estimador utiliza toda la información relevante contenida en la muestra, de modo que ningún otro estimador pueda proporcionar información adicional para estimar al parámetro.

De los tres parámetros más importantes: μ , σ^2 y la proporción poblacional P, se puede establecer que $\overline{X}$, S^2 y $\widehat{P}$, respectivamente, son sus **mejores estimadores**, donde $\widehat{P}$ es la proporción muestral, ya que es demostrable que satisfacen los criterios anteriores.

Estimación por intervalos de confianza.

Es otra forma de estimación de parámetros, mucho más informativa que la puntual, pues permite establecer un rango de valores dentro del cual se encontraría el verdadero valor del parámetro, complementada con un nivel de seguridad o certeza de que esto sea cierto. Para construir intervalos de confianza es necesario partir de un intervalo de probabilidad $(1-\alpha)$ y disponer de una variable pivotal adecuada para el objetivo a conseguir. Un intervalo es de probabilidad si al menos uno de sus límites es una variable aleatoria o una función de ella. Una variable pivotal es un estadígrafo que debe incluir al parámetro a estimar, a su estimador y cuya distribución debe ser conocida y totalmente determinada.

Intervalo de confianza para la media de una población normal.

Existen dos casos a considerar, cuando la varianza poblacional es conocida y cuando esta varianza no es conocida.

Caso 1. Varianza poblacional σ^2 conocida.

En esta situación el único parámetro desconocido es μ el cual debe ser estimado puntualmente mediante $\overline{X}$, luego bajo la normalidad de la población la variable pivotal a utilizar es $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} = N(0,1)$. Un intervalo de probabilidad central $(1-\alpha)$ para la variable Z está dada por $P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1-\alpha$. Sustituyendo Z
$\Rightarrow P(-z_{1-\alpha/2} \leq \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha/2}) = 1-\alpha$ y despejando μ en la desigualdad
$\Rightarrow P(\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}) = 1-\alpha$, obteniéndose un intervalo de probabilidad para μ, porque sus dos límites son variables aleatorias que dependen del estimador $\overline{X}$. Sin embargo, una vez obtenida la muestra y calculado el valor de $\overline{X}$, el intervalo deja de ser aleatorio, pues sus límites serán constantes y en consecuencia no tiene asociada una probabilidad, transformándose en una **proposición**, cuyos valores son *verdadero* o *falso*, es decir, contiene o no a μ. Esta es la razón que explica por qué el intervalo obtenido se denomina de confianza con valor el de la probabilidad con que se construyó. Así

$$(\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \ \mu \ \leq \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n})$$

Intervalo del $100(1\text{-}\alpha)$% de Confianza para μ con varianza conocida.

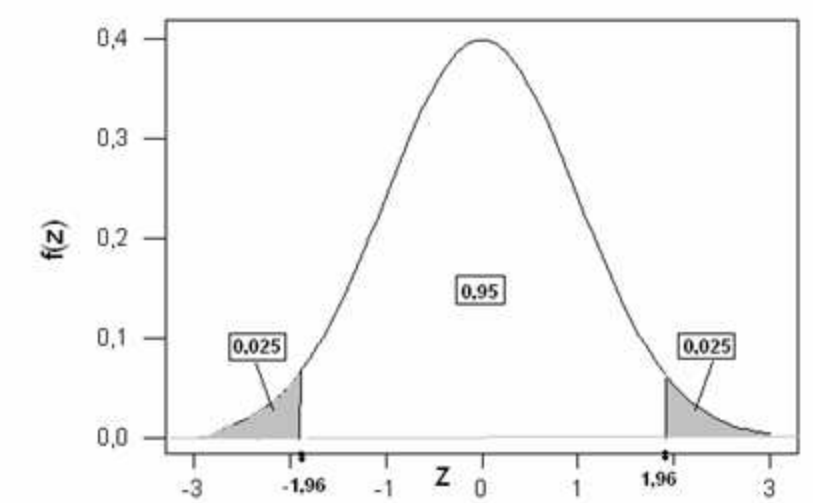


Figura 2.1. Intervalo de probabilidad 0,95 de una distribución normal típica.

Ejemplo 2.1.

Se desea estimar, mediante un intervalo de confianza del 95%, el rendimiento promedio de una nueva variedad de trigo cuya distribución se asume es normal y desviación típica de 12 qq/ha , o sea $X = N(\mu, 144)$. Para tal efecto se siembran 15 parcelas experimentales de 10x10 m . Sus rendimientos, expresados en qq/ha , fueron de 89,4 ; 92,8 ; 79,2 ; 82,6 ; 96,2 ; 65,6 ; 106,4 ; 86,0 ; 99,6 ; 69,0 ; 77,5 ; 58,8 ; 96,2 ; 80,9 ; 52,0.

Como este es un caso de varianza conocida, para construir el intervalo sólo se necesita calcular la media muestral, cuyo valor es 82,15 qq/ha, y determinar que $z_{0,975} = 1,96$ (fig. 2.1). Sustituyendo los valores en la expresión del recuadro anterior

$\Rightarrow$ ($82,15 - 1,96 * \sqrt{144/15} \leq \mu \leq 82,15 + 1,96 * \sqrt{144/15}$) al 95% de confianza

$\Rightarrow$ ($76,1 \leq \mu \leq 88,2$) al 95% de confianza. Se deduce, entonces, que con una certeza del 95%, el rendimiento promedio de la nueva variedad es de entre 76,1 y 88,2 qq/ha.

Caso 2. Varianza poblacional σ^2 desconocida.

En este caso los dos parámetros de la distribución normal son desconocidos y deben ser estimados por $\overline{X}$ y S^2. Debido a la normalidad de la población la variable pivotal a utilizar es $t = \frac{\overline{X}-\mu}{\sqrt{S^2/n}} = t(n-1)$. Ahora el intervalo de probabilidad $(1-\alpha)$ para la variable t está dada por $P(-t_{1-\alpha/2}(n-1) \leq t \leq t_{1-\alpha/2}(n-1)) = 1-\alpha$. Sustituyendo t

$\Rightarrow P(-t_{1-\alpha/2}(n-1) \leq \frac{\overline{X}-\mu}{\sqrt{S^2/n}} \leq t_{1-\alpha/2}(n-1)) = 1-\alpha$, despejando μ en la desigualdad

$\Rightarrow P(\overline{X} - t_{1-\alpha/2}(n-1)\sqrt{S^2/n} \leq \mu \leq \overline{X} + t_{1-\alpha/2}(n-1)\sqrt{S^2/n}) = 1-\alpha$, deduciéndose que

$(\overline{X} - t_{1-\alpha/2}(n-1)\sqrt{S^2/n} \leq \mu \leq \overline{X} + t_{1-\alpha/2}(n-1)\sqrt{S^2/n})$

Intervalo del $100(1-\alpha)\%$ de Confianza para μ con varianza desconocida.

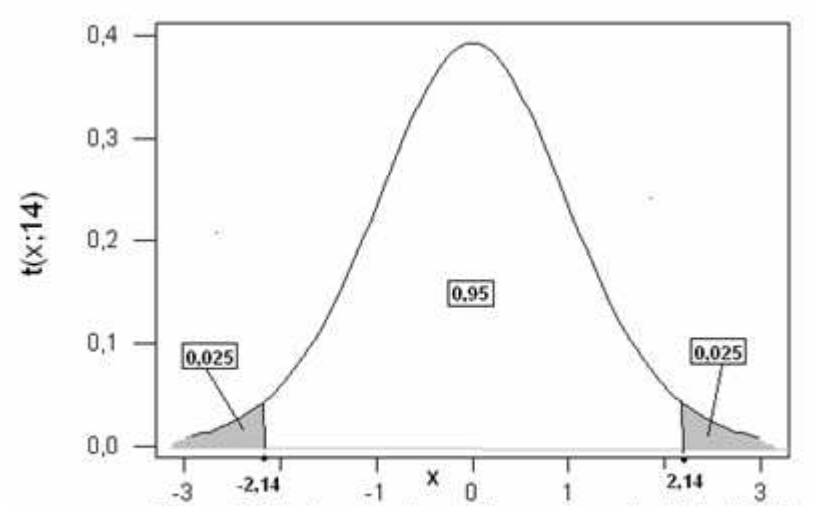


Figura 2.2. Intervalo de confianza de probabilidad 0,95 de una t de Student con 14 grados de libertad.

Ejemplo 2.2.

Asuma que en el mismo enunciado del ejemplo 2.1 no se tiene conocimiento de la variabilidad de los rendimientos de esta nueva variedad, es decir, no se conoce su varianza y que tanto la muestra como los valores muestrales se mantienen. Ahora, además, de obtener un estimador puntual para la media se necesita calcular el estimador de σ^2, S^2, mediante la fórmula $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{n-1}$, que con los datos anteriores resulta ser 243,0. Se necesita, también, el valor $t_{0,975}(14) = 2{,}1448$ (fig. 2.2), ya que ahora la distribución del estadígrafo es t *de Student*. Sustituyendo ($82,15 - 2,1448 * \sqrt{243/15} \leq \mu \leq 82,15 + 2,1448 * \sqrt{243/15}$) al 95% de confianza $\Rightarrow$ ($73,5 \leq \mu \leq 90,8$) al 95% de confianza. Se puede apreciar que esta estimación es más imprecisa que la obtenida con varianza conocida.

Intervalo de confianza para la varianza y desviación típica de una población normal.

Cuando la varianza es desconocida su estimador puntual es S^2 y una estimación por intervalo de confianza debe establecerse utilizando la variable pivotal $D^2 = \frac{(n-1)S^2}{\sigma^2}$ cuya distribución, se recordará es ji cuadrada con (n-1) grados de libertad y un intervalo central de probabilidad (1-α) para una ji cuadrada es

$P(\chi^2_{\alpha/2}(n-1) \leq D^2 \leq \chi^2_{1-\alpha/2}(n-1)) = 1-\alpha$, sustituyendo D^2

$\Rightarrow \; P(\chi^2_{\alpha/2}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}(n-1)) = 1-\alpha$, despejando σ^2

$\Rightarrow \; P(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}) = 1-\alpha$, luego se deduce

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right)$$

Intervalo del $100(1-\alpha)$% de Confianza para σ^2.

El intervalo de confianza para la desviación típica se obtiene tomando la raíz de los tres términos de la desigualdad.

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}\right)$$

Intervalo del $100(1\text{-}\alpha)$% de Confianza para σ.

Ejemplo 2.3.

Se aprovecharán los datos de los ejemplos anteriores para ejemplificar la estimación por intervalo de confianza de la varianza y desviación típica cuando estas son desconocidas. De los cálculos anteriores S^2 resultó ser igual a 243, luego $\left(\frac{14*243}{26,12} \leq \sigma^2 \leq \frac{14*243}{5,63}\right)$ $\Rightarrow (130,2 \leq \sigma^2 \leq 604,3)$ al 95% de confianza ya que $\chi^2_{0,025}(14) = 5,63$ y $\chi^2_{0,975}(14) = 26,12$ (fig. 2.3) y el intervalo para σ es $(\sqrt{130,2} \leq \sigma \leq \sqrt{604,3}) \Rightarrow (\ 11,4 \leq \sigma \leq 24,6)$ al 95% de confianza, luego al 95% de confianza el verdadero valor de la desviación típica poblacional es de entre 11,4 y 24,6 qq/ha .

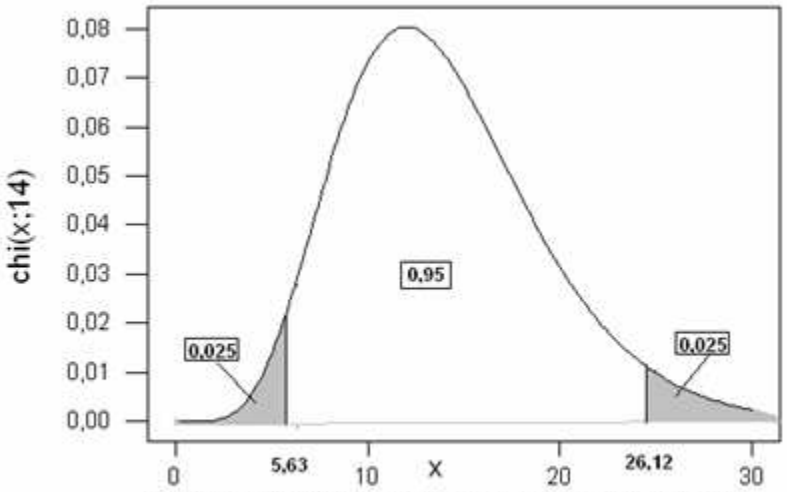


Figura 2.3. Intervalo de probabilidad 0,95 de una distribución chi cuadrado con 14 grados de libertad.

Intervalo de confianza para la diferencia de las medias de dos poblaciones normales.

La estimación se obtendrá a partir de muestras aleatorias *independientes* de $X_1 = N(\mu_1, \sigma_1^2)$ y $X_2 = N(\mu_2, \sigma_2^2)$ de tamaño n_1 y n_2 respectivamente, y se desea estimar $d = (\mu_2 - \mu_1)$. Su estimador $\hat{d} = (\overline{X}_2 - \overline{X}_1)$ tiene distribución normal, por ser una combinación lineal de $\overline{X}_1 = N(\mu_1, \sigma_1^2/n_1)$ y $\overline{X}_2 = N(\mu_2, \sigma_2^2/n_2)$, con $E(\hat{d}) = E(\overline{X}_2 - \overline{X}_1) = \mu_2 - \mu_1$ y $V(\hat{d}) = V(\overline{X}_2 - \overline{X}_1) = V(\overline{X}_2) + V(\overline{X}_1) = \sigma_1^2/n_1 + \sigma_2^2/n_2$, por lo tanto $\hat{d} = N(d, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ y en consecuencia $\frac{\hat{d}-d}{\sqrt{V(\hat{d})}} = \frac{(\overline{X}_2-\overline{X}_1)-(\mu_2-\mu_1)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} = N(0, 1)$.

En el caso más realista, de varianzas poblacionales σ_1^2 y σ_2^2 desconocidas, éstas deben ser estimadas por S_1^2 y S_2^2 respectivamente. El supuesto habitual en casos de 2 o más poblaciones es el de **homocedasticidad**, es decir, que todas las varianzas poblacionales son desconocidas e iguales, luego $\sigma_1^2 = \sigma_2^2 = \sigma^2$, donde σ^2 es la varianza común a ambas poblaciones y por lo tanto S_1^2 y S_2^2 son estimadores de σ^2, razón por la cual combinando

ambas muestras se obtiene el estimador S_p^2 , que corresponde a la media ponderada entre S_1^2 y S_2^2 respecto a sus grados de libertad, luego $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

Recuérdese que el denominador en el cálculo de una varianza corresponde a los grados de libertad de esa varianza muestral y en este caso es igual (n_1+n_2 -2). Sustituyendo σ_1^2 y σ_2^2 por su estimador S_p^2 se obtiene la varianza estimada de $\hat{d}$, $S_p^2/n_1 + S_p^2/n_2 = S_p^2(1/n_1 + 1/n_2)$.

Por lo tanto $\frac{(\overline{X}_2-\overline{X}_1)-(\mu_2-\mu_1)}{\sqrt{S_p^2(1/n_1+1/n_2)}}$ tiene distribución t de Student con (n_1+n_2-2) grados de libertad, porque se está usando una varianza estimada con esos grados de libertad.

El estadígrafo anterior corresponde a la variable pivotal a utilizar para obtener el intervalo de confianza para la diferencia de las medias. Se debe mencionar que, si no se cumpliera el supuesto de homocedasticidad, se tendría una variable pivotal cuya distribución no es exacta.

El intervalo de probabilidad $(1-\alpha)$ para la variable t está dada por:
$P(-t_{1-\alpha/2}(m) \leq t \leq t_{1-\alpha/2}(m)) = 1-\alpha$, con $m = n_1+n_2-2$. Sustituyendo t y despejando μ_2-μ_1 de la desigualdad se obtiene

$$((\overline{X}_2-\overline{X}_1) - t_{1-\alpha/2}(m)\sqrt{S_p^2(1/n_1 + 1/n_2)} \leq \mu_{21} \leq (\overline{X}_2-\overline{X}_1) + t_{1-\alpha/2}(m)\sqrt{S_p^2(1/n_1 + 1/n_2)}\,)$$

Intervalo del $100(1-\alpha)$% de Confianza para μ_2-μ_1, con varianzas desconocidas e iguales.

6.3 Contraste de hipótesis estadísticas.

El contraste de hipótesis, también denominado Prueba de Hipótesis o Docimasia de Hipótesis, corresponde a un conjunto de metodologías cuyo objetivo es verificar si un determinado parámetro toma uno o varios valores posibles de interés. También una prueba de hipótesis puede referirse a la distribución de poblaciones, todo ello evidentemente, a partir de muestras aleatorias.

Existen algunos conceptos básicos vinculados a una prueba de hipótesis y que se explicarán en lo que sigue.

Una **hipótesis estadística** es una proposición acerca de una característica poblacional, como puede ser su distribución o el valor o valores de sus parámetros, y que necesita ser probada. Como se verá, una hipótesis estadística **nunca** podrá ser aceptada libre de **toda** duda, pues siempre existirá un cierto nivel de incertidumbre. Una hipótesis respecto a un parámetro puede ser **simple,** si especifica un único valor del parámetro y **compuesta,** si especifica más de un valor del parámetro.

Una **prueba de hipótesis estadística** consta de **dos** hipótesis. Una denominada *hipótesis nula*, designada por H_0 , y la otra *hipótesis alternativa*, designada por H_1 o H_a. La hipótesis nula es la hipótesis *conservadora* que representa *lo conocido*, el *statu quo.* La hipótesis nula **debe ser una hipótesis simple**, y si se refiere a un parámetro debe especificar un **único** valor para éste. La hipótesis alternativa es la hipótesis que representa el *cambio*, lo que se *quiere probar*. Esta **puede ser una hipótesis simple o compuesta**. Por lo general, se consideran hipótesis alternativas compuestas. Una hipótesis alternativa compuesta puede ser de tres tipo:

1) Hipótesis alternativa bilateral, cuando es la negación de H_0

2) Hipótesis alternativa unilateral derecha, cuando plantea para el parámetro un valor **mayor** al especificado en H_0.

3) Hipótesis alternativa unilateral izquierda, cuando plantea para el parámetro un valor **menor** al especificado en H_0.

Ejemplos 3.1.

a) En un juicio a un individuo que supuestamente cometió un delito, las hipótesis nula y alternativa para un juez son, respectivamente, *Inocente* versus *Culpable*.

b) Un asesor económico aconseja a un productor de kiwi cambiarse a la viticultura porque resultará más rentable. El agricultor si quiere considerar seriamente la alternativa deberá reunir múltiples consejos e información al respecto y deberá plantearse las siguientes hipótesis nula y alternativas respectivamente: *mantenerse como productor de kiwi* versus *cambiarse a la viticultura*.

Los dos ejemplos anteriores se refieren a un ámbito no matemático-estadístico. Un ejemplo en el ámbito estadístico es el siguiente.

c) Un Instituto de Investigación afirma haber desarrollado una nueva variedad de trigo cuyo **rendimiento promedio** supera en 6 qq/ha los 72 qq/ha que rinde la variedad tradicional. Alguien que quiera verificar tal aseveración, debe plantearse las hipótesis $H_0 : \mu = 72$ versus $H_1 : \mu = 78$.

Una **prueba de hipótesis estadística** es una regla que consiste i) en *tomar la decisión* de **aceptar H_0**, cuando estadísticamente la muestra no entregue *evidencia suficiente* para decidir rechazarla o ii) en *tomar la decisión* de **rechazar H_0** si la *evidencia muestral* deja "una mínima duda" de que esa sea la decisión correcta. En resumen, una prueba de hipótesis es una regla de decisión que permite aceptar o rechazar una hipótesis nula, a partir de información muestral. Aceptar una hipótesis nula **no permite la conclusión que ésta sea verdadera**, así como rechazarla, **no permite la afirmación de que la hipótesis alternativa es verdadera. Nunca** es posible probar estadísticamente que una hipótesis nula es verdadera, pues se trata sólo de una cuestión de "credibilidad probabilística".

Ejemplo 3.2.

En el caso 3.1 c) el interesado debe diseñar una muestra aleatoria para reunir información sobre el rendimiento de la nueva variedad y una regla, por el momento arbitraria, podría ser que si se obtiene una media muestral "más cercana a 72" se acepta H_0 y por el contrario si ésta es "más cercana a 78" se rechaza H_0.

Nótese que la anterior es una perfecta regla de decisión, porque cualquier valor $\overline{X}$ que se obtenga, permitirá optar por una u otra hipótesis y además que la decisión debe basarse en un estadígrafo. Sin embargo **no es** una regla diseñada estadísticamente, como se verá posteriormente.

Se llama **región crítica** de una prueba de hipótesis a un conjunto RC que contiene a todos los valores del estadígrafo que conducen al rechazo de H_0.
En el ejemplo 3.2, la región crítica es $RC = \{\overline{X} / \overline{X} > 75\}$, pues para esos valores, $\overline{X}$ estará más cerca de 78 y la decisión será rechazar la hipótesis nula.

En toda prueba de hipótesis existe la posibilidad de cometer **dos tipos de errores**, uno al tomar la decisión de aceptar y el otro la de rechazar la hipótesis nula. Siempre está presente la posibilidad de cometer uno de ellos, pero obviamente el propósito es tomar todas las veces la decisión correcta y como ello no es posible hay que disminuir el *riesgo* de cometer errores de decisión y la manera de lograrlo consiste en mantener baja su posibilidad de ocurrencia.

Las posibles decisiones a tomar se muestran en el siguiente cuadro.

Hipótesis verdadera \ Decidir por	**H_0**	**H_1**
H_0	Decisión correcta	**Decisión errónea: error tipo I**
H_1	**Decisión errónea: error tipo II**	Decisión correcta

El cuadro muestra que en dos situaciones la decisión es la correcta y en otras dos la decisión es incorrecta, pero no existe certeza a que tipo corresponde la decisión tomada. Cuando se toma la decisión de **rechazar H_0, siendo esta la hipótesis *verdadera,* el error que se comete se denomina de tipo I.** Al tomar la decisión de **aceptar H_0, siendo esta la hipótesis *falsa,* el error que se comete se denomina de tipo II.** De los dos errores, el que provoca consecuencias más grave es el tipo I y por lo tanto la posibilidad de cometerlo debe ser más "pequeña". La posibilidad de cometer el error tipo II también importa, pero sus consecuencias son menos grave, razón por la cual debe ser mantenido en niveles de riesgo "razonables". Los niveles de riesgo de ambos errores se establecen en término de probabilidades, según las siguientes definiciones.

Definiciones.

1. La **magnitud del error tipo I** se designa por α, siendo $\alpha = \text{Prob}(\text{rech. } H_0 \text{ / } H_0 \text{verdadera})$.

2. La **magnitud del erro tipo II** se designa por β, donde $\beta = \text{Prob}(\text{aceptar } H_0 \text{ / } H_0 \text{falsa})$.

3. La **Potencia** de una prueba de hipótesis es la probabilidad de **rechazar una hipótesis nula que es falsa** y es igual a $\mathbf{1} - \beta$**.**

En el ejemplo 3.1 a) el juez *puede cometer* el error tipo I cuando decide **declararlo culpable** en circunstancia que el individuo es **realmente inocente**. El juez *puede cometer* el error tipo II si decide **declararlo inocente** cuando **realmente es culpable.** En cualquier otra situación el juez **toma la decisión correcta.** Del comentario anterior resalta que es **más grave** cometer el error tipo I, es decir, declarar culpable a un inocente. También es grave cometer el error tipo II, pero sus consecuencias son menos graves.

En el ejemplo 3.1 b) el agricultor cometería el error tipo I si se **cambia a la viticultura** y resulta que ésta **no es más rentable** que el kiwi. Es fácil apreciar que este error le trae un gran daño económico e incluso podría ser su ruina económica. El error tipo II lo comete si se **mantiene como productor de kiwi** y este resulta **menos rentable que la viticultura.** En este caso también habría un daño económico, en el sentido que perdió la oportunidad de hacer un buen negocio, pero su situación no cambia, sigue igual como estaba, lo que en economía se llama *costo de oportunidad*.

En las dos situaciones anteriores resulta claro que el error tipo I **debe** ser controlado mediante niveles de riesgo bajos que le den al juez o al inversionista "cierta seguridad de protección" contra este error. Por esta razón es que la probabilidad máxima de cometerlo, valor α, queda al arbitrio del interesado o investigador. Con el fin de tener valores comparativos de riesgo, en estadística se conviene en utilizar valores de α de 5% , 1% , 0,1% ó 10%, y no valores intermedios. En las situaciones comunes se ocupa el valor del 5%.

En una prueba de hipótesis se llama **nivel de significación** al valor que el investigador le asigna a α. El nivel se acostumbra a expresarlo en porcentaje. Si el nivel de significación de una prueba es del 1%, entonces α = 0,01.

El ejemplo estadístico 3.1 c) servirá para ver integralmente los conceptos anteriores. Si se quiere comprobar científicamente la aseveración del Instituto de Investigación, es necesario, entonces, realizar una prueba con las hipótesis: $H_0 : \mu = 72$ versus $H_1 : \mu = 78$. Asumiendo que ambas poblaciones se comportan normales, entonces según H_0 la nueva variedad híbrida tiene un comportamiento $N(72, \sigma^2)$, es decir, el mismo de la variedad en uso, mientras que bajo H_1 su comportamiento es $N(78, \sigma^2)$, por el momento la varianza no juega su papel, razón por la cual no se especificará su valor, aunque se supondrá igual en ambas poblaciones.

La figura 3.1 grafica la situación anterior, en donde la campana de la izquierda, X_0, muestra el comportamiento de la variedad híbrida cuando su rendimiento **no es** mejor que la tradicional y la de la derecha, X_1, cuando su rendimiento la supera en 6 qq/ha. Para obtener información que permita apoyar una u otra hipótesis, es necesario tomar una m.a.s. Al no conocer cual es la *real situación* de la nueva variedad, no se sabe si la muestra proviene de la primera o de la segunda de las distribuciones.

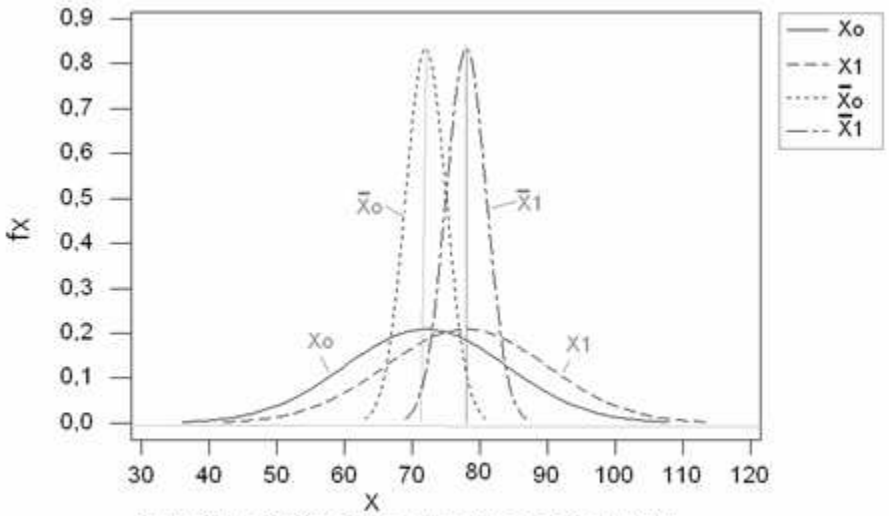


Figura 3.1. Distribuciones poblacionales bajo Ho(Xo), H1(X1) y de ambas medias de muestras tamaño 16.

Sin embargo, la decisión respecto a μ no se toma sobre la base del comportamiento de las poblaciones, sino del comportamiento del estadígrafo $\overline{X}$ estimador del parámetro, representado por las campanas más leptocúrticas, según sea H_0 o H_1 la hipótesis verdadera. En consecuencia la regla de decisión se establece en relación al comportamiento de $\overline{X}_0 = N(72, \sigma^2/n)$ y $\overline{X}_1 = N(78, \sigma^2/n)$, como lo muestra la figura 3.2, que representa las mismas dos campanas leptocúrticas de la figura anterior. La RC se establece en relación a un *valor crítico* (VC) *K*, expresándose en términos generales como $RC = \{\overline{X}/\overline{X} > K\}$, que según el criterio utilizado en el ejemplo 3.2, *K* = 75, éste se ubicaría justo en el punto de corte de las dos campanas de la figura 3.2. En esta situación el error tipo I y tipo II tendrán la misma probabilidad de ocurrir, correspondiendo al área sombreada a la derecha y a la izquierda de *K* respectivamente. Pero el área de la derecha debe tener la magnitud α, entonces la posición de *K* queda determinada por esta condición. Si el nivel de significación de la prueba es del 5%, *K*

debe estar más hacia la derecha, más cercano a 78, de forma tal que el área sombreada bajo la curva que grafica el comportamiento de la media muestral bajo la hipótesis nula H_0, área de la derecha, sea igual a 0,05. De esta manera la magnitud del error tipo II, valor de β, corresponde al área sombreada bajo la curva de la media muestral bajo la hipótesis alternativa H_1. Visualmente se aprecia que la magnitud de β es bastante mayor que la magnitud de α. Es fácil apreciar, que en esta misma situación, al disminuir β aumenta α y viceversa, por el hecho de tener que mover la posición de *K* hacia la izquierda o hacia la derecha respectivamente (fig. 3.2).

La única forma de disminuir β manteniendo fijo el valor de α, consiste en aumentar el tamaño muestral, es decir aumentando *n*. De esa forma se consigue que ambas curvas sean más leptocúrticas, o sea estén más concentradas alrededor de su media y por lo tanto el área de traslape entre ellas sea menor, como se aprecia en la figura 3.3, en la cual la distribución de las medias muestrales corresponde a muestras tamaño 25, mayor que en el caso anterior. Nótese que la posición de *K* se mueve hacia la izquierda, debido a que las áreas disminuyen y *K*, como se dijo, es el límite de un área del 5% bajo la curva $\overline{X}_0$. Un ejemplo numérico ayudará a aclarar estos conceptos.

Ejemplo 3.3.

Supongamos que $X = N(\mu, 144)$, es el comportamiento del rendimiento de la nueva variedad híbrida, del ejemplo 6.3.1 c), donde el valor de μ depende de cual hipótesis, H_0 o H_1, es la verdadera. Se asumió que la desviación típica del rendimiento es 12 qq/ha, ya que para los cálculos se necesitará de tal información. Si, como se hace frecuentemente, se fija arbitrariamente en 16 el tamaño de la muestra, se tendrá que $\overline{X}_0 = N(72, 9)$ y $\overline{X}_1 = N(78, 9)$, pues $\frac{\sigma^2}{n}$ = $\frac{144}{16}$ es 9. De esta manera el valor de *K* se determina asignando α = 0,05
$\Rightarrow$ Prob(rech. H_0 / H_0verdadera) = 0,05 $\Rightarrow P(\overline{X} > K / \mu = 72) = 0,05$
$\Rightarrow P(Z > \frac{K-72}{3}) = 0,05 \Rightarrow 1 - \phi(\frac{K-72}{3}) = 0,05 \Rightarrow \phi(\frac{K-72}{3}) = 0.95$
$\Rightarrow \frac{K-72}{3} = \phi^{-1}(0,95) \Rightarrow \frac{K-72}{3} = 1,645 \Rightarrow K = 76,9$. Con este valor se puede calcular la probabilidad de cometer el error tipo II: $\beta =$ Prob(aceptar H_0 / H_0falsa)
$\Rightarrow \beta = P(\overline{X} \leq K / \mu = 78) \Rightarrow \beta = P(Z \leq \frac{76,9-78}{3}) \Rightarrow \beta = \phi(-0,37) = 0,356$, que corresponde al área sombreada de la izquierda de la figura 3.2.

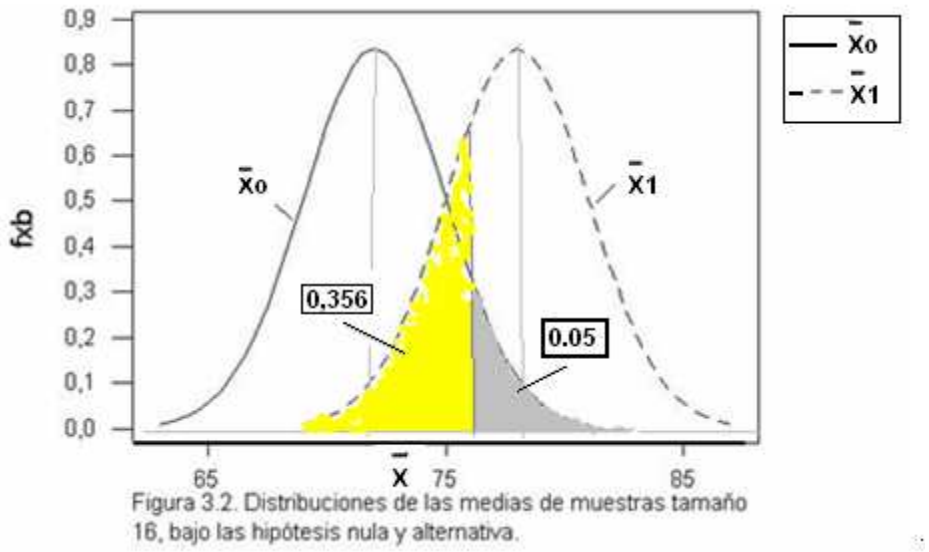


Figura 3.2. Distribuciones de las medias de muestras tamaño 16, bajo las hipótesis nula y alternativa.

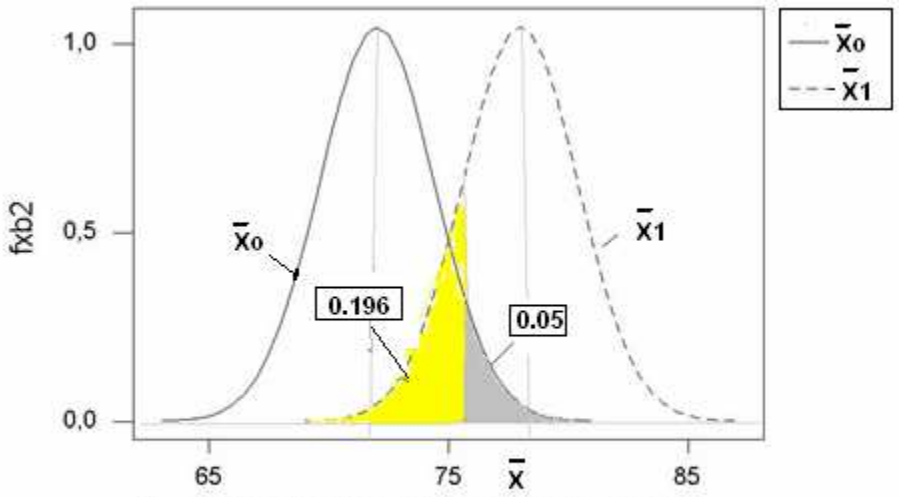


Figura 3.3. Distribuciones de las medias de muestras tamaño 25, bajo las hipótesis nula y alternativa.

Resumiendo, para un nivel de significación del 5% y un tamaño de muestra 16 el valor crítico *K* corresponde a 76,9 con una probabilidad del error tipo II de 35,6%, es decir, aproximadamente 7 veces el error tipo I. Si se aumenta el tamaño de la muestra a 25 se tendrá que $\overline{X}_0 = N(72,\ 5,76)$ y $\overline{X}_1 = N(78,\ 5,76)$, pues $\sigma^2/n =$ 144/25 es 5,76. Siguiendo los mismos pasos anteriores se determina que, ahora *K* toma el valor 75,9, más a la izquierda que antes, con una probabilidad de 19,6% para el error tipo II, casi 4 veces el de α, como se ilustra en la figura 3.3.

En el ejemplo anterior se planteó la relación entre el tamaño de muestra y la magnitud de los errores tipo I y tipo II como suele hacerse en la realidad, esto es, definir el nivel de significación de la prueba y decidir el tamaño de la muestra por consideraciones prácticas, con lo cual se pierde el control del error tipo II, por ello, esa **no es la forma científica** de hacerlo. El tamaño de la muestra es el resultado de decidir a-priori los valores aceptables para α y β, el que dependerá de lo que planteen las hipótesis nula y alternativa, esta última en términos de una hipótesis simple.

Ejemplo 3.4.

Se desea establecer el tamaño de muestra necesario para contrastar las hipótesis del ejemplo 3.1 c), $H_0: \mu = 72$ versus $H_1: \mu = 78$. Asumiendo que $X = N(\mu, 144)$ se tendrá que la distribución de las medias muestrales bajo la hipótesis nula y alternativa son $\overline{X}_0 = N(72, 144/n)$ y $\overline{X}_1 = N(78, 144/n)$. Entonces para valores α = 0,05 y β = 0,15, que corresponden a valores habituales, se tiene:

Prob(rech. H_0 / H_0verdadera) $= 0,05 \Rightarrow P(\overline{X} > K / \mu = 72) = 0,05 \Rightarrow P(Z > \frac{K-72}{\sqrt{144/n}}) = 0,05$

$\Rightarrow 1 - \phi(\frac{(K-72)\sqrt{n}}{12}) = 0,05 \Rightarrow \phi(\frac{(K-72)\sqrt{n}}{12}) = 0.95 \Rightarrow \frac{(K-72)\sqrt{n}}{12} = 1,645$ (1).

Prob(aceptar H_0 / H_0falsa) $= 0,15 \Rightarrow P(\overline{X} \leq K / \mu = 78) = 0,15 \Rightarrow P(Z \leq \frac{K-78}{\sqrt{144/n}}) = 0,15$

$\Rightarrow \phi(\frac{(K-78)\sqrt{n}}{12}) = 0,15 \Rightarrow \frac{(K-78)\sqrt{n}}{12} = -1,04$ (2)

(1) y (2) establecen un sistema para *K* y *n* que al dividir miembro a miembro (1) por (2) se obtiene: $\frac{K-72}{K-78} = \frac{1,645}{-1,04} \Rightarrow \frac{K-72}{K-78} = -1,58 \Rightarrow K = 75,6$. Sustituyendo en (1) $\frac{(75,6-72)\sqrt{n}}{12} = 1,645 \Rightarrow \frac{3,6\sqrt{n}}{12} = 1,645 \Rightarrow \sqrt{n} = 5,48 \Rightarrow n \geq 31$. En el cálculo de n **siempre** se debe aproximar hacia arriba, para no sobrepasar el valor de α. Entonces con un tamaño muestral de 31 o más se podría cometer un error máximo, tipo I ó tipo II, de 5% ó 15% respectivamente, al contrastar las hipótesis planteadas.

Esquema para contrastar hipótesis.

El método científico exige el cumplimiento de ciertas condiciones como son el planteamiento de hipótesis, un análisis lógico y crítico y una metodología válida para probar la hipótesis planteadas. Así, para probar hipótesis es necesario ceñirse a un esquema de 6 pasos que satisface tales exigencias y que se explican a continuación.

1° Se plantean las hipótesis nula, H_0, y la alternativa H_1. La hipótesis nula siempre corresponde a una hipótesis simple, ya que debe especificar **completamente** la distribución poblacional, bajo la cual se establece el estadígrafo de prueba y su distribución, la que debe ser conocida. La hipótesis alternativa especifica lo se quiere probar, que por lo general representa el cambio en relación a la hipótesis nula. Esta hipótesis puede ser simple o compuesta. Por lo general es una hipótesis compuesta, es decir, especifica infinitas distribuciones poblacionales alternativas.

2° Se debe elegir el nivel de significación de la prueba o valor de α, que se refiere al riesgo máximo de cometer el error tipo I, el que según se explicó anteriormente es el que provoca consecuencias más grave.

3° Se debe identificar el estadígrafo de prueba, el que debe tener características similares a la variable pivotal y cuya distribución debe ser conocida.

4° Se especifica la Región Crítica, *RC*, cuya construcción depende de la hipótesis alternativa, el valor de α y la distribución del estadígrafo de prueba.

5° Consiste en planificar la muestra aleatoria cuyas observaciones entregarán la evidencia que permitirá tomar la decisión de rechazar o aceptar la hipótesis nula. Para este propósito es necesario procesar los valores y obtener un valor calculado del estadígrafo de prueba o valor

muestral. A continuación se debe verificar si el valor, así calculado, pertenece o no la Región Crítica. Si pertenece, la decisión es **rechazar** la hipótesis nula, en caso contrario la decisión es **aceptarla o no rechazarla**. Aceptar la hipótesis nula debe interpretarse en el sentido que los datos no proporcionan evidencia suficiente para refutarla, lo que no es equivalente a concluir que lo que plantea la hipótesis nula es lo verdadero. Recuerde que es imposible establecer la certeza de que una hipótesis es verdadera a partir de una muestra. Al rechazar una hipótesis nula se debe concluir que con los datos muestrales es más *creíble* o probable lo que especifica la hipótesis alternativa, dado que, bajo la condición que la hipótesis nula es la verdadera, la probabilidad de obtener una muestra que proporcione los datos que nos conduce a la hipótesis alternativa resulta ser *pequeña*. Una probabilidad *pequeña* se refiere a que su valor es igual o menor al nivel de significación de la prueba de hipótesis cuyo valor es α.

6° En este paso se debe **redactar una conclusión** respecto al problema en es tudio, la que se deduce del análisis de los resultados realizados en la etapa anterior.

En cada uno de los siguientes tipos de pruebas de hipótesis sólo se indicarán los pasos 1, 3 y 4 que son específicos de cada una, puesto que los pasos 2, 5 y 6 son generales y tienen el mismo enunciado anterior.

Prueba de hipótesis para la media de una población normal**.**

Sea la población $X = N(\mu\,,\,\sigma^2)$ de la cual se toma una m.a.s. tamaño *n.*

1° Las hipótesis son:

$$H_0 : \mu = \mu_0 \text{ versus } H_1: \begin{cases} \mu \neq \mu_0 & \text{hipótesis bilateral} \\ \mu > \mu_0 & \text{hipótesis unilateral derecha} \\ \mu < \mu_0 & \text{hipótesis unilateral izquierda} \end{cases}, \ \mu_0 \in \Re$$

Existen dos casos a considerar:

Caso 1. Varianza poblacional σ^2 conocida.

3° En esta situación, al igual que para intervalos de confianza, el estadígrafo de prueba es $Z = \frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}} = N(0,1)$, bajo la hipótesis H_0.

4° La región crítica depende de lo establecido en los tres pasos anteriores y en particular de la hipótesis alternativa, por lo cual hay tres posibles *R.C.* asociadas a cada una de las tres hipótesis alternativa, con un z_c que resulta de los cálculos al sustituir $\overline{X}$ en el estadígrafo indicado en el paso anterior:

RC = { z_c / $z_c < -z_{1-\alpha/2}$ o $z_c > z_{1-\alpha/2}$ } región crítica bilateral

RC = { z_c / $z_c > z_{1-\alpha}$ } región crítica unilateral derecha

RC = { z_c / $z_c < -z_{1-\alpha}$ } región crítica unilateral izquierda

Obsérvese que la región crítica no se estableció $\overline{X} > K$, porque resulta más directa la forma anterior, para evitar tener que despejar $\overline{X}$, como se deduce de: $\frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}} > z_{1-\alpha}$, que al despejar se obtiene $\overline{X} > \mu_0 + z_{1-\alpha}\sqrt{\sigma^2/n}$, donde $K = \mu_0 + z_{1-\alpha}\sqrt{\sigma^2/n}$.

Ejemplo 3.5.

Se desea probar, al nivel del 5%, si una nueva variedad de trigo tiene mayor rendimiento que la variedad tradicional, actualmente en uso, cuyo rendimiento promedio se sabe es de 72 qq/ha con una desviación típica de 12 qq/ha. Con esta descripción se debe plantear la prueba a realizar, es decir, establecer los pasos 1 a 4 del esquema propuesto.

1) $H_0 : \mu = 72$ versus $H_1: \mu > 72$

2) Se fijará un nivel de significación del 5% (α = 0,05)

3) El estadígrafo de prueba, bajo la hipótesis H_0, es $Z = \frac{\overline{X}-\mu_0}{\sqrt{\sigma^2/n}} = N(0,1)$, suponiendo que los rendimientos se distribuyen normales, lo que habitualmente es verdadero, y por ser conocida la varianza poblacional.

4) La región crítica es unilateral derecha porque la hipótesis alternativa lo es, luego
RC = { z_c / $z_c > z_{0,95} = 1,645$}

5) Con el objetivo de realizar la prueba planteada, se siembran 10 parcelas experimentales de 10x10 m con semilla de la nueva variedad, obteniéndose una producción para cada una de 89,4 ; 92,8 ; 82,6 ; 96,2 ; 106,4 ; 86,0 ; 69,0 ; 77,5 ; 96,2 ; 80,9 qq/ha.
A partir de los datos se calcula que $\overline{X}$ = 87,7 y $z_c = \frac{87,7-72}{\sqrt{144/10}} = 4,14$ y como este valor pertenece a la *RC*, pues 4,14 > 1,645, entonces la decisión es rechazar H_0.

6) Basado en la evidencia proporcionada por la muestra aleatoria es posible concluir que la nueva variedad tiene un rendimiento superior a la tradicional, al nivel del 5%.

Observación.

En la conclusión es importante dejar constancia del nivel de significación de la prueba, porque es posible que la decisión de rechazar la hipótesis nula sea incorrecta, es decir, se puede estar cometiendo el error tipo I, cuyo valor **máximo** es el valor de α. Sin embargo en el ejemplo 3.5 , el verdadero valor del error tipo I, de haberse cometido, es mucho menor al 5%, debido a que z_c = 4,14 es bastante mayor que el valor crítico 1,645, valor límite de la región de rechazo, lo que indica que el z_c está muy al *interior* de la región crítica, lo que otorga mayor seguridad en no estar cometiendo un error en la decisión tomada.

Caso 2. Varianza poblacional σ^2 desconocida.

Las hipótesis son las mismas del caso 1, en consecuencia sigue el paso siguiente:

3° En esta situación el estadígrafo de prueba, bajo la hipótesis H_0, es $t = \frac{\overline{X}-\mu_0}{\sqrt{S^2/n}} = t(n-1)$, por lo cual de la muestra se debe obtener tanto el valor de $\overline{X}$ como de S^2.

4° Las regiones críticas con un t_c que resulta de los cálculos al sustituir $\overline{X}$ y S^2 en el estadígrafo indicado, son similares a las del caso 1, pero con valores percentiles de la t:

RC = { t_c / $t_c < -t_{1-\alpha/2}\ (n-1)$ o $t_c > t_{1-\alpha/2}(n-1)$} región crítica bilateral
RC = { t_c / $t_c > t_{1-\alpha}\ (n-1)$} región crítica unilateral derecha
RC = { t_c / $t_c < -t_{1-\alpha}\ (n-1)$} región crítica unilateral izquierda

Note que en ambos casos la región crítica bilateral es el complemento del intervalo de confianza, pues corresponde a la parte externa de éste.

Ejemplo 3.6.

Un productor de pollos Broiler afirma que los pollos que produce cumplen con una norma sanitaria que establece que la cantidad de hormonas que estos contengan no debe superar los 19 nanogramos. Un inspector sanitario decide probar tal afirmación sobre la base de 10 pollos.

El siguiente es el planteamiento de la prueba a realizar por el inspector, puesto que éste debe probar, hipótesis H_1, que el productor no cumple la norma.

1) $H_0 : \mu = 19$ versus $H_1: \mu > 19$

2) El inspector decide fijar un nivel de significación del 5% (α = 0,05)

3) El estadígrafo de prueba, bajo la hipótesis H_0, es $t = \frac{\overline{X}-\mu_0}{\sqrt{S^2/10}} = t(9)$, pues la varianza poblacional es **desconocida** y asumiendo que los contenidos de hormonas se distribuyen normales.

4) La región crítica es unilateral izquierda como la hipótesis alternativa , por lo tanto *RC* = { t_c / $t_c > t_{0,95}\ (9) = 1,8331$ } .

5) Para verificar la afirmación del productor el inspector sanitario toma una muestra aleatoria de 10 pollos del productor, obteniendo los siguientes contenidos de hormona, en nanogramos, en cada pollo: 18 ; 22 ; 21 ; 19 , 18 ; 17 ; 19 ; 20 ; 22 ; 20. De estos valores se obtiene que $\overline{X}$ = 19,6, S^2 = 2,94 y $t_c = \frac{19,6-19}{\sqrt{2,94/10}}$ = 1,10 , que al **no** pertenecer a la *RC* implica la decisión de aceptar H_0, o sea, no rechazarla.

6) La conclusión que obtiene el inspector es que la evidencia muestral no permite establecer que el productor no cumpla la norma.

Observaciones.

Con la decisión tomada por el inspector, el error susceptible de haberse cometido es el error tipo II, cuyo nivel no está explícito, pero está directamente vinculado al tamaño de la muestra y como la muestra es relativamente pequeña puede corresponder a una alta probabilidad. El valor de β puede calcularse a posteriori y en él se podría buscar una explicación de por qué *la prueba no fue capaz de rechazar* H_0. En este caso es *irrelevante* informar del valor α.

Prueba de hipótesis para las medias de dos poblaciones normales.

Sean las poblaciones $X_1 = N(\mu_1\ , \sigma_1^2)$, de la cual se toma una m.a.s. tamaño n_1 y $X_2 = N(\mu_2\ , \sigma_2^2)$, de la cual se toma una m.a.s. tamaño n_2.

1° Las hipótesis son: $H_0 : \mu_2 = \mu_1$ versus H_1: $\begin{cases} \mu_2 \neq \mu_1 & \text{hipótesis bilateral} \\ \mu_2 > \mu_1 & \text{hipótesis unilateral derecha} \\ \mu_2 < \mu_1 & \text{hipótesis unilateral izquierda} \end{cases}$

Es fácil deducir que las hipótesis anteriores se pueden replantear así:

$H_0 : \mu_2 - \mu_1 = 0$ versus $H_1: \begin{cases} \mu_2 - \mu_1 \neq 0 \\ \mu_2 - \mu_1 > 0 \\ \mu_2 - \mu_1 < 0 \end{cases}$, con tres casos a considerar.

Caso 1. Varianza poblacionales σ_1^2 y σ_2^2 conocidas.

Este es un caso poco usual, pero se tratará porque servirá de apoyo en la explicación de los casos 2 y 3. Las hipótesis nula y alternativa son comunes a los tres casos.

3° A partir de muestras aleatorias *independientes* de $X_1 = N(\mu_1, \sigma_1^2)$ y $X_2 = N(\mu_2, \sigma_2^2)$ de tamaño n_1 y n_2 respectivamente, el estimador de $(\mu_2 - \mu_1)$ es $(\overline{X}_2 - \overline{X}_1)$ cuya distribución es $N(\mu_2 - \mu_1, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ y $Z = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = N(0, 1)$, según lo establecido en la construcción del Intervalo de confianza para la diferencia de dos medias poblacionales. En consecuencia como bajo $H_0 : \mu_2 - \mu_1 = 0$, el estadígrafo de prueba es:
$Z = \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = N(0, 1)$.

4° Las regiones críticas asociadas son las mismas del Caso 1, para la media de una población normal con varianza conocida, esto es
$RC = \{ z_c / z_c < -z_{1-\alpha/2}$ o $z_c > z_{1-\alpha/2} \}$ región crítica bilateral
$RC = \{ z_c / z_c > z_{1-\alpha} \}$ región crítica unilateral derecha
$RC = \{ z_c / z_c < -z_{1-\alpha} \}$ región crítica unilateral izquierda

Caso 2. Varianzas poblacionales σ_1^2 y σ_2^2 desconocidas e iguales.

3°Este es el caso más usual, en donde σ^2, es la varianza común a ambas poblaciones, correspondiente al supuesto de **homogeneidad de varianzas** u **homocedasticidad** y el estadígrafo a utilizar es $t = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} = t(n_1 + n_2 - 2)$, tal como se utilizó anteriormente para construir el Intervalo de Confianza para la diferencia de dos medias y que bajo H_0 adopta la forma $t = \frac{\overline{X}_2 - \overline{X}_1}{\sqrt{S_p^2(1/n_1 + 1/n_2)}} = t(n_1 + n_2 - 2)$, donde se recordará que $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$.

4° Las regiones críticas asociada son:
$RC = \{ t_c / t_c < -t_{1-\alpha/2}\ (n_1 + n_2 - 2)$ o $t_c > t_{1-\alpha/2}(n_1 + n_2 - 2) \}$ región bilateral
$RC = \{ t_c / t_c > t_{1-\alpha}\ (n_1 + n_2 - 2) \}$ región unilateral derecha
$RC = \{ t_c / t_c < -t_{1-\alpha}\ (n_1 + n_2 - 2) \}$ región unilateral izquierda

Ejemplo 3.7.

Para determinar si el parasitismo disminuye la capacidad física de caballos para competencias, se evalúa el rendimiento de 20 caballos sin desparasitar, obteniendo un rendimiento promedio de 29,9 y una varianza de 15. A su vez se evalúa el rendimiento de 12 caballos desparasitados, obteniendo que su rendimiento promedio es de 32,4 con una varianza de 10. El rendimiento se mide en una escala cuyo máximo es 40. ¿Es posible

establecer, al nivel del 5 %, que el parasitismo afecta la capacidad física de caballos para competencias ?

El planteamiento de la prueba se efectúa en los pasos 1 a 4, para lo cual es necesario hacer algunos alcances. El rendimiento de ambas poblaciones se asume normal y se establece en términos de la media μ, así la población 1 será la de caballos *desparasitados (con tratamiento)* y la población 2 la de caballos *sin desparasitar (sin tratamiento)*. Entonces lo que se quiere probar es que el rendimiento promedio de la población 2 es *menor* que el de la población 1.

1) En consecuencia las hipótesis serán $H_0 : \mu_2 = \mu_1$ versus $H_1: \mu_2 < \mu_1$

2) Se utilizará α = 0,05

3) Como se trata de dos poblaciones con varianzas no conocidas, ya que la información del promedio y la varianza proviene de muestras, el estadígrafo de prueba es

t = $\frac{\overline{X}_2-\overline{X}_1}{\sqrt{S_p^2(1/n_1+1/n_2)}} = t(n_1+n_2-2)$ con $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

4) La región crítica es *RC* = { t_c / $t_c < -t_{0,95}\ (30) = -1,6973$} , unilateral izquierda

5) Según el enunciado los valores de la media y varianza muestrales son:
con tratamiento $\overline{X}_1 = 32,4$, $S_1^2 = 10$, $n_1 = 12$; sin tratamiento $\overline{X}_2 = 29,9$, $S_2^2 = 15$, $n_2 = 20$, de donde $S_p^2 = \frac{11*10+19*15}{30} = 13,2$ y $t_c = \frac{29,9-32,4}{\sqrt{13,2(1/12+1/20)}} = -1,88 \in RC \Rightarrow$ rechazar H_0.
6) Se puede concluir, a un nivel del 5%, que en base a la evidencia muestral el parasitismo disminuye la capacidad física de caballos para competencias.

Una forma más general de la prueba para comparar dos medias consiste en plantearse las hipótesis de que las diferencias entre las dos medias es una cantidad *d*, no necesariamente igual a 0. Replanteando las hipótesis y el estadígrafo, queda en los siguientes términos:

1º $H_0 : \mu_2 - \mu_1 = d$ versus $H_1: \begin{cases} \mu_2 - \mu_1 \neq d \\ \mu_2 - \mu_1 > d \\ \mu_2 - \mu_1 < d \end{cases}$, $d \in \Re$

3º $t = \frac{(\overline{X}_2-\overline{X}_1)-(\mu_2-\mu_1)}{\sqrt{S_p^2(1/n_1+1/n_2)}} = t(n_1+n_2-2)$, que bajo H_0 queda $t = \frac{(\overline{X}_2-\overline{X}_1)-d}{\sqrt{S_p^2(1/n_1+1/n_2)}} = t(n_1+n_2-2)$.
4º En el resto se procede igual al caso 2.

Caso 3. Varianzas poblacionales σ_1^2 y σ_2^2 desconocidas y distintas.

Corresponde al caso de heterogeneidad de varianza y es un caso en el cual no existe un estadígrafo de prueba con distribución exacta conocida y en consecuencia se debe recurrir a aproximaciones, alguna de las cuales se incluyen en los programas estadísticos computacionales. Uno de las aproximaciones más conocidas es el procedimiento de Smith-Satterthwaite. Otro procedimiento[(1)] consiste en calcular $t' = \frac{\overline{X}_2-\overline{X}_1}{\sqrt{S_1^2/n_1+S_2^2/n_2}}$ tal que:

---

(1) Métodos Estadísticos, Snedecor,G. y Cochran,W.; CECSA, 4°impresión, 1977.

i) t' tiene distribución aproximada $t\,(n-1)$, si $n_1 = n_2$ o ii) se compara t' con el valor crítico $t^* = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$, donde $t_1 = t\,(n_1\text{-}1)$ y $t_2 = t\,(n_2\text{-}1)$, con ponderadores $w_1 = S_1^2/n_1$ y $w_2 = S_2^2/n_2$, si $n_1 \neq n_2$.

Prueba de hipótesis para la igualdad de dos varianzas poblacionales.

Corresponde a la prueba para la homogeneidad o igualdad de dos varianzas.

1° Las hipótesis son $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$.

2° Se elige α de 5% o de 10%, según se cuente con una tabla $\mathbb{F}$ que tenga o no el percentil $1 - \alpha/2$.

3° El estadígrafo a utilizar es $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \mathbb{F}(m-1\,,\, n-1)$, deducido en la unidad de distribuciones muestrales, el que bajo la hipótesis H_0, pues al ser iguales σ_1^2 y σ_2^2 se cancelan, toma la forma $F = S_1^2/S_2^2 = \mathbb{F}(n_1 - 1\,,\, n_2 - 1)$, donde $(n_1\text{-}1)$ y $(n_2\text{-}1)$ son los grados de libertad de S_1^2 y S_2^2 respectivamente.

Generalmente las tablas de la distribución $\mathbb{F}$ están resumidas para los valores percentiles superiores, razón por la cual la prueba es conveniente realizarla en los siguientes términos:

Se calcula la razón $F = \frac{S_m^2}{S_n^2} = \mathbb{F}(m\,,\, n)$ ubicando en el numerador la varianza muestral mayor y en el denominador la menor, de modo que la razón sea mayor que 1.

4° La región crítica es $RC = \{\ F_c / F_c > \mathbb{F}_{1-\alpha/2}(m, n)\ \}$, siendo F_c el valor muestral del estadígrafo que resulta de sustituir los valores respectivos de S^2.

Ejemplo 3.8.

Una situación que se debe establecer previamente cuando las varianzas poblacionales son desconocidas es si estas son homogéneas, para de esa manera discriminar si la prueba se refiere al caso 2 o al caso 3. Esta prueba debe realizarse a-priori a la comparación de medias, pero en beneficio del desarrollo conceptual de la unidad se efectuará en este caso a-posteriori con los datos del ejemplo 3.7 en cuyo enunciado se establece que $S_1^2 = 10$ y $S_2^2 = 15$. El desarrollo es el siguiente:

1) $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$

2) α = 0,10 , pues se utilizará una tabla del 95% de la distribución $\mathbb{F}$.

3) el estadígrafo es $F = \frac{S_2^2}{S_1^2} = \mathbb{F}(19\ ,\ 11)$. La varianza de la muestra 2 va en el numerador, porque es la mayor.

4) La región crítica es bilateral, pero $RC = \{\ F_c / F_c > \mathbb{F}_{0,95}(19\,,\,11) = 2,66\ \}$, por limitaciones de la tabla utilizada.

5) $F_c = 15/10 = 1,5 \notin RC \Rightarrow$ aceptar H_0

6) Se concluye que las varianzas poblacionales son homogéneas, al nivel del 10%.

6.4 Comentarios sobre intervalos de confianza y pruebas de hipótesis.

En esta sección se analizarán algunas situaciones complementarias en relación a la estimación mediante intervalos de confianza, respecto a las pruebas de hipótesis y en particular al planteamiento de la hipótesis alternativa cuando se docima la media de una población, debido a que a veces se deben tener en cuenta ciertas consideraciones en relación al error tipo I.

Precisión , confianza y tamaño de muestra en intervalos de confianza.

Se llama **error de muestreo** a la diferencia d entre el valor muestral de un estimador y el del parámetro al cual estima. En términos estadístico: d $= \mid \widehat{\theta} - \theta \mid$. El error de muestreo es función del tamaño muestral, de la varianza y del valor percentil de la distribución de $\widehat{\theta}$. En el caso del intervalo de confianza para μ con varianza conocida d = $z_{1-\alpha/2}\sqrt{\sigma^2/n}$ y cuando la varianza poblacional no es conocida d = $t_{1-\alpha/2}(n-1)\sqrt{S^2/n}$. En otros términos el error de muestreo es igual a la mitad de la amplitud del intervalo de confianza.

Se llama **precisión** de una estimación, al grado de aproximación del valor muestral del estimador respecto al valor poblacional. Se mide en términos del error de muestreo, de modo que a menor tamaño del error de muestreo existe mayor precisión. Precisión no se debe confundir con *exactitud*, que cuando ésta se refiere a un cálculo depende del número de decimales o del instrumento de cálculo, o cuando se trata de una medición depende del instrumento con que se realiza la medición, ya que tiene que ver con la aproximación del valor calculado respecto a su valor real. La precisión es un término más estadístico y la exactitud es más ingenieril.

La precisión y el grado de confianza de un intervalo están relacionados a través del tamaño de la muestra, pues para una misma muestra a **mayor grado de confianza** se tiene una **menor precisión** y viceversa. La única forma de mantener la precisión aumentando el nivel de confianza o viceversa, consiste en *aumentar* el tamaño de la muestra. Algunos ejemplos ayudarán a conceptualizarlos.

Ejemplo 4.1.

En el ejemplo 2.1 se necesitaba estimar la media de una población normal de varianza 144 a partir de una muestra tamaño 15, resultando un promedio de 82,15 qq/ha y un intervalo del 95% de confianza para μ con límites 76,1 y 88,2 qq/ha. En este caso la precisión es de 6,05 qq/ha. Si con la misma muestra se construye un intervalo al 90% de confianza el error de muestreo es d = $1{,}645*\sqrt{144/15} = 5{,}1$ qq/ha. Se puede observar que se disminuyó el grado de confianza, pero aumentó la precisión. Si se aumenta la confianza al 99%, entonces d = $2{,}575*\sqrt{144/15}$ = 8,0 qq/ha. Deduzca que pasaría si se tratara de aumentar la confianza al 100%.

La forma científica de enfocar el problema consiste en determinar el tamaño de muestra necesario para una determinada precisión y nivel de confianza. Entonces, el planteamiento en el caso anterior debe ser, por ejemplo, "calcular el tamaño de muestra necesario para estimar la media poblacional con una confianza del 95% y una precisión de 3 qq/ha". Ahora se conoce

que d = 3, $z_{0,975}$ = 1,96, luego, 3 = 1,96$*\sqrt{144/n}$, despejando n se obtiene 61,47, pero como *n* tiene que ser un número natural se aproxima siempre hacia arriba, lo que implica *n* = 62.

La población, el parámetro, las hipótesis a contrastar y el tamaño de muestra en una prueba de hipótesis para una población.

Cuando se desea realizar una inferencia es importante tener claridad cual es la población y el o los parámetro de ella que se está investigando a partir de muestras aleatorias. Es frecuente que el concepto que se tiene de la población sea algo difuso y resulta que es un aspecto muy importante, porque las conclusiones se refieren a ella y sólo a ella y la muestra tiene que ser un subconjunto que la represente, luego la población debe estar definida en términos bien precisos.

Respecto a las hipótesis, un error frecuente es plantearlas para los estadígrafos en circunstancias que estos son variables aleatorias y por lo tanto la probabilidad de que ocurra un valor puntual es cero. Las hipótesis **siempre** se plantean para los parámetros y la hipótesis nula **siempre** es una hipótesis simple, pues el valor del parámetro especificado en ésta **determina** la distribución del estadígrafo de prueba que debe ser exacta.

El planteamiento de la hipótesis alternativa para un parámetro algunas veces puede generar dudas, pues depende de a cual decisión errónea se le quiere dar mayor protección, es decir, el planteamiento formal de una hipótesis está influida por la estructura de la probabilidad de una conclusión errada. El análisis de ciertos casos ayudarán a desarrollar esta idea.

Caso 1. Si un investigador desea probar que **tomar café aumenta el riesgo de cáncer gástrico**, las hipótesis a contrastar son: *tomar café aumenta el riesgo de cáncer gástrico* versus *tomar café **no** aumenta el riesgo de cáncer gástrico.* El punto es cuál debe ser la hipótesis nula y cuál la alternativa. Si se considera que lo conservador es considerar que tomar café no produce daño gástrico, entonces

H_0: *tomar café **no** aumenta el riesgo de cáncer gástrico*

H_1*: tomar café aumenta el riesgo de cáncer gástrico*

Se evaluará la consecuencia de tomar cada una de las posibles decisiones erróneas:

i) **si se acepta H_0** cuando la hipótesis alternativa es la verdadera, se está cometiendo el error tipo II, de probabilidad β y como la conclusión será que no hay riesgo al tomar café, la consecuencia del error es grave, porque se está poniendo en riesgo la salud en términos de un error que por lo general tiene valores de probabilidad más alto que el tipo I. Es decir al elegir plantear así las hipótesis, el error más grave que es el riesgo de contraer cáncer, no está siendo controlado adecuadamente.

ii) **si se rechaza H_0** cuando ésta es verdadera, se está cometiendo el error tipo I, de probabilidad α. La conclusión será que tomar café es riesgoso para la salud y la decisión será abstenerse de beber café. El **costo** es perderse la oportunidad de tomar café, especialmente si se es adicto al café, pero no hay riesgo para la salud. Si se permutan las dos hipótesis anteriores, ambos tipos de errores, también se permutan, verificándose que el riesgo para la salud queda protegido con el nivel de significación, como debe ser. Recuerde que el error de peores consecuencias es el tipo I.

Caso 2. Una Compañía Tabacalera afirma que la cantidad de nicotina que en promedio contiene, uno de sus tipos de cigarrillos, no excede de 2,5 mg. Un investigador que desea verificar tal aseveración debe optar por establecer sus hipótesis nula y alternativa. El investigador toma la opción que $H_0: \mu = 2,5$ versus $H_1: \mu > 2,5$ y va a realizar la prueba

con un tamaño muestral suficiente para tener un nivel de significación del 1% y un error tipo II de probabilidad $\beta = 0{,}15$.

Se analizará cual es la consecuencia de cada una de las dos decisiones erróneas.

i) **si se acepta H_0** cuando la hipótesis alternativa es la correcta, se está cometiendo el error tipo II cuya probabilidad es del 15% y decidiendo que la evidencia muestral no es suficiente para contradecir la afirmación de la Compañía, luego se estarían aceptando cigarrillos con exceso de nicotina, lo que sería muy perjudicial para la salud de los fumadores y con un alto nivel de riesgo.

ii) **si se rechaza H_0** siendo H_0 verdadera, es decir, H_1 falsa, se está cometiendo el error tipo I cuya probabilidad es del 1% y decidiendo erróneamente que los cigarrillos exceden los 2,5 mg de nicotina. En esta situación se está perjudicando a la Compañía con un nivel de riesgo del 1% muy inferior al 15% de riesgo que corre la salud de los fumadores.

Como evidentemente la salud de las personas es mucho más importante que el daño económico de la Compañía, el error tipo I debe proteger al consumidor y en consecuencia las hipótesis deben ser $H_0 : \mu = 2,5$ versus $H_1 : \mu < 2,5$. Ahora el fabricante se verá perjudicado con una probabilidad del 15% al aceptar H_0, pero la Compañía tiene una solución para esta situación, la cual consiste en financiar un análisis de los contenidos de nicotina en los cigarrillos en una muestra mucho mayor, con lo cual se consigue disminuir el valor de β.

Caso 3. Una agroindustria establece como norma de calidad que la fruta que envien los productores debe contener un porcentaje de frutos con daños por insectos de a lo más 6%. Si la partida contiene un porcentaje mayor será rechazada.

La decisión se tomará en base a una muestra de tamaño suficiente para tener $\alpha = 5\%$ y $\beta = 15\%$, siendo el valor del parámetro a probar una proporción o porcentaje P. La hipótesis alternativa a plantearse tiene dos posibilidades, proteger preferentemente a la agroindustria o proteger al productor. Si se considera como norma que se debe proteger al más débil las hipótesis deben ser H_0: $P = 0,06$ versus H_1: $P > 0,06$. De esta manera al rechazar H_0 cuando H_1 es falsa, se está cometiendo el error tipo I, que conduce a rechazar la partida cuando ésta cumple la norma, pero la probabilidad de este error es de sólo 5%. Por el contrario si la hipótesis alternativa fuera H_1: $P < 0,06$, al aceptar H_0 cuando ésta es falsa el error cometido es el tipo II, luego hay una probabilidad del 15% de rechazar una partida que cumple la norma, en vez del 5% anterior. En este caso la atención hay que ponerla en si la hipótesis alternativa debe plantear la aceptación o el rechazo de la partida de fruta, según cuál decisión errónea sea más grave.

Tamaño de muestra.

Se tomará el caso de la Compañía Tabacalera para explicar el procedimiento del cálculo del tamaño de muestra necesario para cumplir con valores pre establecidos para los errores tipo I y tipo II. Se optará por las hipótesis que protegen la salud de los fumadores, es decir, $H_0 : \mu = 2,5$ versus $H_1 : \mu < 2,5$. Pero para resolver el problema se debe tener información de la variabilidad del contenido de nicotina en los cigarrillos, así que supóngase que la desviación típica es de 0,5 mg y como la hipótesis alternativa **debe ser una hipótesis simple** se asumirá que $H_1 : \mu = 2,3$. Entonces el tamaño de muestra para $\alpha = 0,01$ y $\beta = 0,15$ se obtiene a partir del siguiente planteamiento.

$\alpha = P(\textit{rech } H_0 / H_0 \textit{ verdadera}) \Rightarrow 0{,}01 = P(\overline{X} < K / \mu = 2,5)$ (1)

$\beta = P(\textit{aceptar } H_0 / H_0 \text{ falsa}) \Rightarrow 0{,}15 = P(\overline{X} \geq K / \mu = 2,3)$ (2)

que con el supuesto que el contenido de nicotina en los cigarrillos tiene distribución normal

(1) $\Rightarrow P(\frac{\overline{X}-2{,}5}{0{,}5/\sqrt{n}} < \frac{K-2{,}5}{0{,}5/\sqrt{n}}) = 0,01 \Rightarrow \phi(\frac{K-2{,}5}{0{,}5/\sqrt{n}}) = 0,01 \quad \Rightarrow \phi(\frac{(K-2{,}5)\sqrt{n}}{0{,}5}) = 0,01 \ \ (3)$

(2) $\Rightarrow P(\frac{\overline{X}-2{,}3}{0{,}5/\sqrt{n}} \geq \frac{K-2{,}3}{0{,}5/\sqrt{n}}) = 0,15 \Rightarrow 1 - \phi(\frac{K-2{,}3}{0{,}5/\sqrt{n}}) = 0,15 \Rightarrow \phi(\frac{(K-2{,}3)\sqrt{n}}{0{,}5}) = 0,85 \ (4)$

$(3) \Rightarrow \frac{(K-2{,}5)\sqrt{n}}{0{,}5} = \phi^{-1}(0,01) \Rightarrow \frac{(K-2{,}5)\sqrt{n}}{0{,}5} = -2,33 \ \ (5)$

$(4) \Rightarrow \frac{(K-2{,}3)\sqrt{n}}{0{,}5} = \phi^{-1}(0,85) \Rightarrow \frac{(K-2{,}3)\sqrt{n}}{0{,}5} = 1,04 \qquad (6)$

El sistema de ecuaciones (5) y (6) tiene dos incógnitas que son *K y n*. Para eliminar *n*, se divide miembro a miembro (5)/(6), se obteniéndose $\frac{K-2{,}5}{K-2{,}3} = -2,24$, luego $K = 2{,}36$. Sustituyendo *K* en (6) $\Rightarrow \ \frac{(2{,}36-2{,}3)\sqrt{n}}{0{,}5} = 1,04 \ \Rightarrow 0,12\sqrt{n} = 1,04 \Rightarrow n = 76$. En consecuencia se debe analizar una muestra de 76 cigarrillos o más.

Observación.

El tamaño de muestra depende de las condiciones: de variabilidad poblacional reflejada en el valor de la desviación típica; del nivel de significación requerido; del valor de la potencia $(1-\beta)$ deseada y de la diferencia, $d = \mu_1 - \mu_0$, que se establece a partir de los valores de las medias en las hipótesis alternativa y nula respectivamente. Para esta última condición es necesario que la hipótesis alternativa sea una hipótesis simple y como en general las hipótesis alternativas son compuestas hay un tamaño de muestra asociado a cada valor de la diferencia *d.*

7. TEOREMA CENTRAL DEL LIMITE E INFERENCIAS PARA PROPORCIONES.

7.1 Muestras de tamaño pequeño.

Para muestras de tamaño pequeño las inferencias deben realizarse con la distribución exacta del estadígrafo de prueba, esto es, si la distribución poblacional es normal utilizando la distribución normal de la media muestral, si la distribución poblacional es binomial con la distribución binomial del estadígrafo, si la distribución poblacional es Poisson con la distribución Poisson del estadígrafo y así en otros casos.

En la unidad anterior la metodología para las inferencias se basan en el supuesto de normalidad poblacional, para de esta manera obtener estadígrafos o variables pivotales con distribución normal o t de Student o χ^2. Hay muchos casos en los cuales la normalidad poblacional no se cumple y en consecuencia se debe proceder con la distribución exacta, lo que trae algún grado de complicación por que las tablas de esas distribuciones son menos completas que la de la distribución normal típica. El siguiente es un ejemplo de este tipo.

Se sabe que un tipo de vacuna contra el distemper es alérgica en un 40% de los casos. Un laboratorio promueve una nueva vacuna tan efectiva como la anterior, aunque algo más cara, que es menos alérgica que la en uso. Para tal efecto se inoculan 20 perros para decidir sobre la afirmación del laboratorio. Las hipótesis son $H_0: P = 0,40$ versus $H_1: P < 0,40$ y sea X número de caninos de la muestra que presentan alergia producida por la vacuna, cuyos valores posibles son $0, 1, 2, ..., 19, 20$, en consecuencia la distribución es $X = Bin(20,\ 0,40)$, luego la regla de decisión debe diseñarse para una $RC = \{X / X < K\}$, donde K es un número natural . La cuestión es ¿cómo se determina el valor de K ?. La respuesta está en la distribución acumulativa de la binomial anterior, donde se observa que $P(X \leq 3\ ) = 0,0160$ y $P(X \leq 4) = 0,0510$, de modo que para un nivel de significación del 5%, la última probabilidad da aproximadamente ese valor y en consecuencia $K = 5$, pues recuérdese que $\alpha = Prob\{ X < 5 / P = 0,40\}$= 0,051.

Para establecer el valor de β es necesario fijar un valor alternativo simple para P. Supongamos que $H_1: P = 0,20$, entonces:
$\beta = Prob\{X \geq 5 / P = 0,20\} = 1 - Prob\{X \leq 4 / P = 0,20\} = 1 - 0,6296 = 0,3704$, es decir, el error tipo II es aproximadamente del 37%.

7.2 Teorema del Límite Central.

No obstante lo anterior, es posible validar la distribución normal como parte de la metodología estadística, tomando muestras de tamaño grande, situación que establece el Teorema Central del Límite, el que se puede enunciar así.

Sea X variable aleatoria con cualquier distribución, tal que $E(X) = \mu$ y $V(X) = \sigma^2$ *y* $\overline{X}$ *la media de una muestra tamaño n, entonces* $\overline{X} \rightarrow N(\mu, \sigma^2/n)$ *cuando* $n \rightarrow \infty$.

Consecuencias.

Del Teorema anterior se deduce que:
1) $\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \to N(0,1)$ cuando $n \to \infty$.
2) Cuando *n* es *suficientemente grande*, lo que para la mayoría de los casos ocurre si $n > 30$, se puede hacer uso de que $\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \approx N(0,1)$. Esta es una consecuencia importante, porque establece que basta tener muestras de tamaño mayor a 30 para que la distribución de la media muestral sea *prácticamente normal*, independientemente de cual sea la distribución poblacional.

7.3 Proporción Poblacional.

Sea A una característica de interés a estudiar en la población, la que inducirá una partición de ésta en dos subconjuntos: el de los individuos que poseen la característica y el de los individuos que no la poseen. Así en una población finita de tamaño *N* la proporción P de individuos que la poseen queda determinado por $P = \frac{\#A}{N}$. Según la ley de los grandes números $\lim_{N\to\infty} P = \lim_{N\to\infty} \frac{\#A}{N} = p$, que conceptualmente es la probabilidad de *A*, P(A). Esta probabilidad $p = P(A)$ se denominará en adelante **proporción poblacional** en poblaciones infinitas y se designará simplemente por $\boldsymbol{P}$.

El estimador de la proporción P se define como $\widehat{P} = \frac{X}{n}$, que corresponde a la proporción *muestral*, donde *X* es el número de individuos en la muestra que presentan la característica *A* cuya probabilidad de ocurrencia es *p*, en consecuencia la distribución de la variable aleatoria *X* es *Bin*(n, p), y a partir de ésta se puede deducir la distribución del estadígrafo $\widehat{P}$.

Distribución del estadígrafo $\widehat{P}$.

$E(\widehat{P}) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}\, np = p = P$, en consecuencia $\widehat{P}$ es un estimador insesgado de P.
$V(\widehat{P}) = V(\frac{X}{n}) = \frac{1}{n^2}V(X) = \frac{1}{n^2}\, np(1-p) = \frac{p(1-p)}{n} = \frac{PQ}{n}$, donde $Q = 1 - P$.
Luego la distribución de $\widehat{P}$ es *Binomial de media* P *y varianza* PQ/n.

Aproximación a la normal de la distribución de $\widehat{P}$.

Se recordará que una variable aleatoria binomial es generada mediante una suma de *n* variables Bernoulli, luego $X = \sum_{i=1}^{n} Y_i$ y dado que $\widehat{P} = \frac{X}{n} = \frac{\sum_{i=1}^{n} Y_i}{n} = \overline{Y}$, se establece que la proporción muestral es **la media** de variables Bernoulli y como consecuencia por el Teorema del Límite Central $\widehat{P} \to N(P, PQ/n)$ cuando $n \to \infty$. Se deduce, entonces, que cuando *n* es *suficientemente grande* $\widehat{P} \approx N(P, PQ/n) \Rightarrow \frac{\widehat{P}-P}{\sqrt{PQ/n}} \approx N(0, 1)$. En el caso de una proporción se considera que *n* es *suficientemente grande* si satisface la relación $nPQ > 4$, lo que indica que el valor de *n* depende del valor de P, como por ejemplo para $P = 0,5 \Rightarrow n*\frac{1}{2}*\frac{1}{2} > 4 \Rightarrow n > 16$, o sea en este caso se necesita un *n* de 17 o más. Para $P = 1/10 \Rightarrow n*\frac{1}{10}*\frac{9}{10} > 4 \Rightarrow n > 400/9$, es decir, se necesitaría un *n* de 45 o más.

Las figuras 3.1 , 3.2 y 3.3 ilustran como una distribución *Bin*$(n, 0,10)$ se aproxima a una distribución $N(0,1n, 0,09n)$. Ver también la figura 6.2 de la unidad 4.

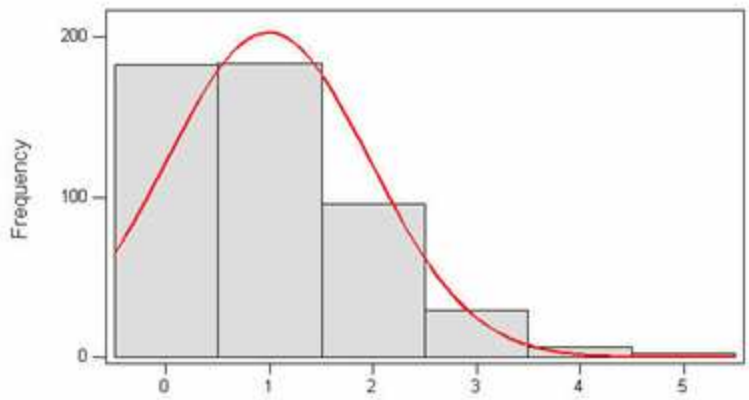


Figura 3.1. Histograma de una distribución binomial (10 , 0,10)

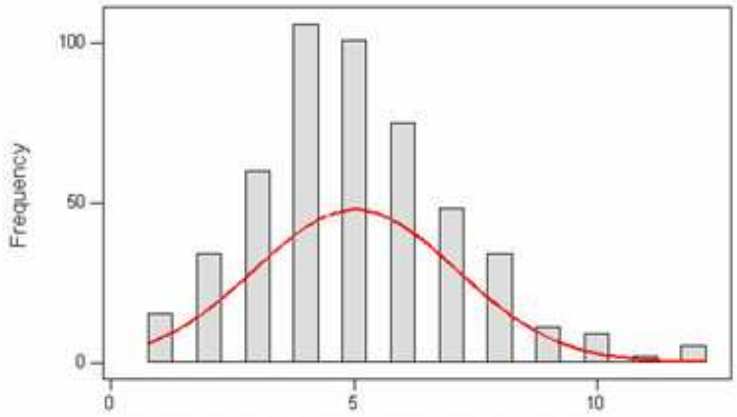


Figura 3.2. Histograma de una distribución Binomial (50 , 0,10)

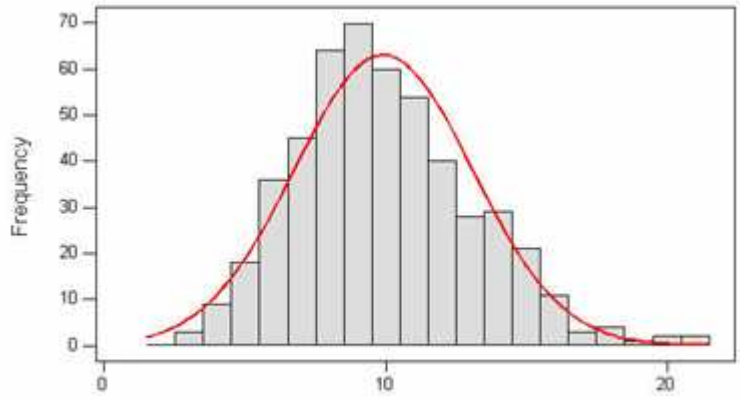


Figura 3.3. Histograma de una distribución Binomial (100 , 0,10)

En lo que sigue se desarrollará la inferencia para proporciones basada en muestras tamaño grande, utilizando $\frac{\widehat{P}-P}{\sqrt{PQ/n}} \approx N(0\,,\,1)$ aproximación establecida por el Teorema Central del Límite.

7.4 Intervalos de Confianza para Proporciones.

El desarrollo sigue un esquema similar al utilizado para intervalos de confianza para la media de distribuciones normales.

Intervalo de confianza para una proporción.

El estadígrafo $Z = \frac{\widehat{P}-P}{\sqrt{PQ/n}} \approx N(0\,,\,1)$, se utilizará como variable pivotal y dado que *Prob*$(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1-\alpha \quad \Rightarrow$ *Prob*$(-z_{1-\alpha/2} \le \frac{\widehat{P}-P}{\sqrt{PQ/n}} \le z_{1-\alpha/2}) \approx 1-\alpha$, pero ahora la probabilidad del intervalo es sólo aproximada. Despejando P en la desigualdad anterior se establece que *Prob*$(\widehat{P} - z_{1-\alpha/2}\sqrt{PQ/n} \le P \le \widehat{P} + z_{1-\alpha/2}\sqrt{PQ/n}) \approx 1-\alpha$. Sin embargo, como es P lo que se está estimando la $V(\widehat{P}) = PQ/n$ no es conocida, por lo cual se debe utilizar su estimador $\widehat{V}(\widehat{P}) = \widehat{P}(1-\widehat{P})/n = \widehat{P}\widehat{Q}/n$, pero aunque $\frac{\widehat{P}-P}{\sqrt{\widehat{P}\widehat{Q}/n}}$ debería tener aproximadamente una distribución t de Student por estar utilizando una varianza estimada, resulta que si *n* es grande $t(n-1) \approx N(0\,,\,1)$, luego por doble aproximación *Prob*$(\widehat{P} - z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n} \le P \le \widehat{P} + z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n}\ ) \approx 1-\alpha$, de donde

$$(\widehat{P} - z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n} \le P \le \widehat{P} + z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n}\ )$$

Intervalo del 100$(1-\alpha)$% aproximado de confianza para P

Ejemplo 4.1

Un organismo de defensa al consumidor examinó 100 latas de atún envasadas por cierta industria encontrando que 9 de ellas estaban en mal estado. En un intervalo de confianza del 95%, ¿ cuál es la proporción de latas en mal estado de la producción total de la industria?

Para el intervalo de confianza se requiere $\widehat{P} = \frac{9}{100} = 0,09$; $\sqrt{\widehat{P}\widehat{Q}/n} = \sqrt{\frac{0,09*0,91}{100}} = 0,029$ y $z_{0,975} = 1,96 \Rightarrow (0,09 - 1,96*0,029 \le P \le 0,09 + 1,96*0,029) \Rightarrow (0,033 \le P \le 0,147)$ al 95% aproximado de confianza. Puede apreciarse que el rango estimado va entre 3,3% y 14,7% de latas en mal estado, que es una estimación con poca precisión.

Mejor, entonces, es plantearse que si se desea tener una estimación con una precisión o error de muestreo menor al 3% y una confianza del 95% ¿ cuál debería ser el tamaño muestral requerido ? Como *n* resultará bastante mayor que 100, que es el tamaño de muestra ya utilizado, para una precisión de un $5,7$%, semi longitud del intervalo anterior, y recordando que el error de

muestreo en una distribución normal está dado por $z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n}$ se tiene que $z_{1-\alpha/2}\sqrt{\widehat{P}\widehat{Q}/n} <$ 0,03 $\Rightarrow 1,96*\sqrt{0,09*0,91}/\sqrt{n} < 0,03 \Rightarrow \sqrt{n} > 1,96*\sqrt{0,09*0,91}/0,03$
$\Rightarrow \sqrt{n} > 18,697 \Rightarrow n > 349,6$, luego $n \ge$ 350. Es decir, para ese nivel de precisión se necesitaría examinar por lo menos 350 latas seleccionadas al azar. Con ese tamaño de muestra se tendría una estimación de la verdadera proporción P de latas en mal estado, en un rango de $P \pm$ 0,03, es decir, una estimación con un 3% de error y una confianza del 95%.

Intervalo de confianza para la diferencia entre la proporción de dos poblaciones.

La estimación se obtendrá a partir de muestras aleatorias *independientes* tamaño n_1 y n_2 de cada población respectivamente, y se desea estimar (P_2 - P_1), mediante $(\widehat{P}_2$ - $\widehat{P}_1)$ cuyos valores característicos son:

i) $E(\widehat{P}_2 - \widehat{P}_1) = E(\widehat{P}_2) - E(\widehat{P}_1) = P_2 - P_1$

ii) $V(\widehat{P}_2 - \widehat{P}_1) = V(\widehat{P}_2) + V(\widehat{P}_1) = \frac{P_2Q_2}{n_2} + \frac{P_1Q_1}{n_1}$, cuyo estimador está dado por

$\widehat{V}(\widehat{P}_2 - \widehat{P}_1) = \frac{\widehat{P}_2\widehat{Q}_2}{n_2} + \frac{\widehat{P}_1\widehat{Q}_1}{n_1}$. Si los tamaños muestrales n_1 y n_2 son grandes, entonces $\frac{(\widehat{P}_2-\widehat{P}_1)-(P_2-P_1)}{\sqrt{\frac{\widehat{P}_2\widehat{Q}_2}{n_2}+\frac{\widehat{P}_1\widehat{Q}_1}{n_1}}} \approx N(0,1)$. El estadígrafo anterior corresponde a la variable pivotal a utilizar para obtener el intervalo de confianza para la diferencia de dos proporciones.

El intervalo de probabilidad $(1-\alpha)$ al sustituir la variable pivotal es como antes *Prob*$(-z_{1-\alpha/2} \leq \frac{(\widehat{P}_2-\widehat{P}_1)-(P_2-P_1)}{\sqrt{\frac{\widehat{P}_2\widehat{Q}_2}{n_2}+\frac{\widehat{P}_1\widehat{Q}_1}{n_1}}} \leq z_{1-\alpha/2}) \approx 1-\alpha$, que al despejar $P_2 - P_1$ se obtiene

$$((\widehat{P}_2 - \widehat{P}_1) - z_{1\text{-}\alpha/2}\sqrt{\frac{\widehat{P}_2\widehat{Q}_2}{n_2} + \frac{\widehat{P}_1\widehat{Q}_1}{n_1}} \leq P_2 - P_1 \leq (\widehat{P}_2 - \widehat{P}_1) + z_{1\text{-}\alpha/2}\sqrt{\frac{\widehat{P}_2\widehat{Q}_2}{n_2} + \frac{\widehat{P}_1\widehat{Q}_1}{n_1}})$$

Intervalo del 100$(1-\alpha)$% aproximado de confianza para $P_2 - P_1$

Ejemplo 4.2

Una industria de alimentos desea promover por TV un nuevo cereal. Una agencia de publicidad le asegura que un cierto comercial será igualmente efectivo en el estrato ABC1 como en los estratos C2 y C3, sin embargo la industria cree que por las características del comercial será menos efectivo en el C2-C3. Para verificar la hipótesis de la empresa se decide pasar por TV el comercial durante dos semanas en el horario de una teleserie de moda, al cabo de las cuales se tomarán muestras de espectadores fanáticos de la teleserie de ambos estratos socio-económicos, para verificar la retención del mensaje en cada uno. Terminado el periodo de prueba el resultado del muestreo indicó que recordaban el mensaje 90 personas de un total de 120 del estrato ABC1 y también otras 90 de un total de 150 del estrato C2-C3 ¿cuál es la diferencia entre la proporción de personas de cada estrato que recuerdan el comercial, en un rango del 95% de confianza ?

Sea P_1 el parámetro del estrato ABC1, cuyo estimador es $\widehat{P}_1 = \frac{90}{120} = 0,75$ y P_2 el parámetro del estrato C2-C1, cuyo estimador es $\widehat{P}_2 = \frac{90}{150} = 0,60$. Para construir el intervalo se requieren los valores $\widehat{P}_2 - \widehat{P}_1 =$ - $0,15$, $\sqrt{\frac{\widehat{P}_2\widehat{Q}_2}{n_2} + \frac{\widehat{P}_1\widehat{Q}_1}{n_1}} = \sqrt{\frac{0,60*0,40}{150} + \frac{0,75*0,25}{120}} =$ 0,056 y $z_{0,975} = 1,96$, luego ($-0,15 - 1,96*0,056 \leq P_2 - P_1 \leq$ - $0,15 + 1,96*0,056$) implica que ($-0,260 \leq P_2 - P_1 \leq -0,040$) al 95% aproximado de confianza. El intervalo obtenido establece que el porcentaje de retención es entre un 4% a un 26% superior en el estrato ABC1, dado que la diferencia es negativa y por lo tanto superior para P_1.

7.5 Contraste de hipótesis para proporciones.

El esquema es similar al de las pruebas de hipótesis para las medias poblacionales.

Prueba de hipótesis para la proporción de una población.

Es el caso en el cual la característica *A* produce dos subpoblaciones y se requiere probar que porcentaje representa la subpoblación con la característica A respecto al total.

1° Las hipótesis son $H_0 : P = P_0$ vs. H_1: $\begin{cases} P \neq P_0 & \text{hipótesis bilateral} \\ P > P_0 & \text{hipótesis unilateral derecha} \\ P < P_0 & \text{hipótesis unilateral izquierda} \end{cases}$, $0 \leq P_0 \leq 1$

2° el nivel de significación se determina con los criterios habituales

3° en esta situación, con *n* sufientemente grande, el estadígrafo de prueba, bajo la hipótesis H_0, es $Z = \frac{\widehat{P} - P_0}{\sqrt{P_0 Q_0 / n}} \approx N(0\,,\,1)$.

4° la región crítica corresponde a la de una distri bución normal típica, con un z_c que resulta de los cálculos al sustituir $\widehat{P}$ en el estadígrafo indicado en el paso anterior.

RC = { z_c / $z_c < -z_{1-\alpha/2}$ o $z_c > z_{1-\alpha/2}$ } región crítica bilateral

RC = { z_c / $z_c > z_{1-\alpha}$ } región crítica unilateral derecha

RC = { z_c / $z_c < -z_{1-\alpha}$ } región crítica unilateral izquierda

Ejemplo 5.1.

Se desea verificar si la multiplicación por estacas de cierta planta medicinal es viable, para lo cual debe enraizar a lo menos el 40% de las estacas, para lo cual se someterán a enraizamiento 140 estacas. El siguiente es el planteamiento para esta situación.

1) Las hipótesis son: $H_0 : P = 0{,}40$ versus $H_1 : P > 0{,}40$, pues el parámetro a probar es una proporción y la multiplicación por estacas sólo sería viable si la proporción de estacas que enraizan es mayor al 40%.

2) Se fijará un nivel de significación del 5%

3) El estadígrafo de prueba es $Z = \frac{\widehat{P} - P_0}{\sqrt{P_0 Q_0 / n}} \approx N(0\,,\,1)$, porque $n = 140$ es sufientemente grande.

4) Corresponde utilizar *R.C* = { z_c / $z_c > z_{0{,}95} = 1{,}645$ } .

5) Para probar las hipótesis anteriores se establecen 140 estacas en un medio para enraizamiento, verificándose, después de un tiempo, que de estas enraizan 60. Se calcula $\widehat{P} = \frac{60}{140} = 0,429$ y $z_c = \frac{0{,}429 - 0{,}40}{\sqrt{0{,}4 * 0{,}6 / 140}} = \frac{0{,}029}{0{,}0414} = 0,70 \notin RC \Rightarrow$ aceptar H_0.

6) La evidencia muestral no es concluyente para establecer que la multiplicación por estaca es viable.

Observaciones.

1) Una cuestión a plantearse es calcular el valor de la potencia de la prueba anterior que no permite rechazar H_0. El siguiente planteamiento resuelve esta situación:
$\alpha = Prob(\text{rech } H_0 / H_0 \text{ verdadera}) \Rightarrow 0,05 = Prob(\widehat{P} > K / P = 0,4) \Rightarrow \phi(\frac{K-0,4}{0,0414}) = 0,95$
$\Rightarrow \frac{K-0,4}{0,0414} = 1,645 \Rightarrow K = 0,468.$
$\beta = Prob(\text{aceptar } H_0 / H_0 \text{ falsa}) = Prob(\widehat{P} \leq K / P = 0,429) = \phi(\frac{K-0,429}{\sqrt{0,429*0,571/140}})$, sustituyendo el valor de K se obtiene $\beta = \phi(\frac{0,468-0,429}{\sqrt{0,429*0,571/140}}) = \phi(\frac{0,039}{0,0418}) = \phi(0,93) = 0,8238$, luego $1 - \beta = 0,1762$, es decir, la potencia es 17,6%, lo que es un valor muy bajo.

2) La otra forma de enfocar el problema, como se ha planteado antes, consiste en calcular el tamaño *n* suficiente para α del 5% y una potencia del 80%. El planteamiento implica
$\alpha = Prob(rech\ H_0 / H_0\ verdadera) \Rightarrow 0,05 = Prob(\widehat{P} > K / P = 0,4)$ (1)
$\beta = Prob(aceptar\ H_0 / H_0 \text{ falsa}) \Rightarrow 0,20 = Prob(\widehat{P} \leq K / P = 0,45)$ (2) , asumiendo $0,45$ como valor alternativo para P.
(1) $\Rightarrow P(\frac{\widehat{P} - 0,4}{\sqrt{0,4*0,6}/\sqrt{n}} > \frac{K-0,4}{0,49/\sqrt{n}}) = 0,05 \Rightarrow \phi(\frac{(K-0,4)\sqrt{n}}{0,49}) = 0,95$ (3)
(2) $\Rightarrow P(\frac{\widehat{P} - 0,45}{0,497/\sqrt{n}} \leq \frac{K-0,45}{0,497/\sqrt{n}}) = 0,20 \Rightarrow \phi(\frac{(K-0,45)\sqrt{n}}{0,497}) = 0,20$ (4)
(3) $\Rightarrow \frac{(K-0,4)\sqrt{n}}{0,49} = \phi^{-1}(0,95) \Rightarrow \frac{(K-0,4)\sqrt{n}}{0,49} = 1,645$ (5)
(4) $\Rightarrow \frac{(K-0,45)\sqrt{n}}{0,497} = \phi^{-1}(0,20) \Rightarrow \frac{(K-0,45)\sqrt{n}}{0,497} = -0,84$ (6)

Resolviendo el sistema de ecuaciones (5) y (6), se obtiene $K = 0,433$ que sustituyéndolo en (5) resulta $n = 597$, muy superior a la muestra de 140 estacas. Este tamaño se puede disminuir si se utiliza como proporción alternativa 0,42 o 0,43.

Prueba de hipótesis para las proporciones de dos poblaciones.

Sean X_1 , X_2 , P_1 y P_2 dos poblaciones y las respectivas proporciones en que está presente una misma característica.

1° las hipótesis son: $H_0 : P_2 = P_1$ versus H_1: $\begin{cases} P_2 \neq P_1 & \text{hipótesis bilateral} \\ P_2 > P_1 & \text{hipótesis unilateral derecha} \\ P_2 < P_1 & \text{hipótesis unilateral izquierda} \end{cases}$

las que se pueden replantear así:

$H_0 : P_2 - P_1 = 0$ versus H_1: $\begin{cases} P_2 - P_1 \neq 0 \\ P_2 - P_1 > 0 \\ P_2 - P_1 < 0 \end{cases}$

3° A partir de muestras aleatorias *independientes* de X_1 y X_2 de tamaño n_1 y n_2 respectivamente, suficientemente grandes, el estimador de $(P_2 - P_1)$ es $\widehat{P}_2 - \widehat{P}_1$ con $V(\widehat{P}_2 - \widehat{P}_1) = V(\widehat{P}_2) + V(\widehat{P}_1) = \frac{P_2 Q_2}{n_2} + \frac{P_1 Q_1}{n_1}$. Pero bajo H_0 se tiene que $P_2 = P_1 = P$, en consecuencia $\widehat{P}_1$ y $\widehat{P}_2$ son estimadores de la proporción común, por lo cual se utiliza como estimador de P la media ponderada $\widehat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$ y de Q a $\widehat{Q} = 1 - \widehat{P}$. Sustituyendo estos estimadores en la varianza anterior, se obtiene que $V(\widehat{P}_2 - \widehat{P}_1) = \widehat{P}\widehat{Q}(\frac{1}{n_1} + \frac{1}{n_2})$,

obteniéndose como estadígrafo de prueba $\frac{(\widehat{P}_2-\widehat{P}_1)-(P_2-P_1)}{\sqrt{\widehat{P}\,\widehat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{(\widehat{P}_2-\widehat{P}_1)}{\sqrt{\widehat{P}\,\widehat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}} \approx N(0,1)$, porque bajo H_0 $P_2 - P_1 = 0$.

4) Las regiones críticas asociadas son las mismas de los casos anteriores de distribuciones normales.
RC = { z_c / $z_c < -z_{1-\alpha/2}$ o $z_c > z_{1-\alpha/2}$ } región crítica bilateral
RC = { z_c / $z_c > z_{1-\alpha}$ } región crítica unilateral derecha
RC = { z_c / $z_c < -z_{1-\alpha}$ } región crítica unilateral izquierda

Ejemplo 5.2

Resultados observados con un nuevo medicamento utilizado para aliviar la tensión nerviosa llevan a pensar que éste es mejor que el que se prescribe comúnmente. Para probar la efectividad del nuevo medicamento, a un grupo de 100 adultos se les administra el medicamento tradicional y a otros 100 adultos se les administra el nuevo medicamento, sin que ellos sepan cual están recibiendo. Los resultados establecen que del primer grupo 59 sienten alivio, mientras que en los del segundo grupo 71 experimentan alivio. ¿ Con la información obtenida a través de los pacientes, puede concluirse al nivel del 1%, que el nuevo medicamento tiene mejor efecto que el tradicional ?

Se seguirá el esquema de 6 pasos, para lo cual P_1 es la proporción de pacientes que se alivian con el medicamento tradicional y P_2 la proporción de pacientes que se alivian con el nuevo.
1) $H_0 : P_2 = P_1$ versus $H_1: P_2 > P_1$, pues el nuevo medicamento será recomendado si cumple con que la proporción de pacientes que son aliviados es mayor que con el tradicional.

2) α = 0,01 , porque una decisión errónea es muy riesgosa.
3) Como n_1 y n_2 son suficientemente grande, el estadígrafo será Z = $\frac{\widehat{P}_2-\widehat{P}_1}{\sqrt{\widehat{P}\,\widehat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}} \approx N(0,1)$.
4) *RC* = { z_c / $z_c > z_{0,99} = 2,33$ }

5) Los estimadores son $\widehat{P}_1 = \frac{59}{100} = 0,59$, $\widehat{P}_2 = \frac{71}{100} = 0,71$ y el estimador común $\widehat{P} = \frac{59+71}{200} = 0,65$, luego $z_c = \frac{0,71-0,59}{\sqrt{0,65*0,35(\frac{1}{100}+\frac{1}{100})}} = 1,78 \notin RC \Rightarrow$ aceptar H_0.
6) Con la evidencia entregada por la muestra no puede establecerse, con un nivel de significación del 1%, que el nuevo medicamento sea más efectivo que el tradicional para aliviar la tensión nerviosa. Si las hipótesis se hubieran planteado con un nivel de significación del 5%, la conclusión sería distinta.

Observaciones.

1) El orden de los estimadores en el estadígrafo de prueba debe ser el mismo que el de los parámetros en las hipótesis, ya que si su orden se invierte, el valor z_c **cambiará de signo**, lo cual no será consecuente con la *RC* lo que podría llevar a una decisión equivocada. Si en el ejemplo anterior, para las mismas hipótesis, se planteara el estadígrafo Z = $\frac{\widehat{P}_1-\widehat{P}_2}{\sqrt{\widehat{P}\,\widehat{Q}(\frac{1}{n_1}+\frac{1}{n_2})}}$ $\Rightarrow z_c$ = - 1,78 $\notin$ *RC*, al nivel del 1%, decisión que coincide con la que corresponde, pero con

un valor mucho más alejado del valor crítico. Sin embargo al nivel del 5% lo que corresponde es rechazar H_0, pero en esta forma errónea aún se aceptaría H_0.

2) Es posible que las hipótesis se planteen correctamente cambiando el orden de los parámetros, en cuyo caso el orden de los estimadores en el estadigrafo también debe ser cambiado, pues en este caso la *RC* cambia. Si en el ejemplo, las hipótesis se plantearan equivalentemente H_0: $P_1 = P_2$ vs. H_1: $P_1 < P_2$, en el cual el signo de la desigualdad debe cambiarse, el estadígrafo ahora debe ser como en la observación anterior y la región crítica será *RC* = { z_c / $z_c < -z_{0,99} = -2,33$ } del tipo unilateral izquierda. El valor calculado correcto será z_c = - 1,78 $\notin$ *RC*, al nivel del 1%, pero al 5% si pertenecería a la región crítica, tal como sucede con el planteamiento original. Esta situación es totalmente simétrica a la desarrollada en el ejemplo.

3) En el caso de una hipótesis alternativa bilateral las dos observaciones anteriores no tienen efecto.

4) Estas observaciones también son válidas para la prueba de hipótesis para dos medias.

7.6 Contraste de hipótesis para dos o más proporciones.

Hay dos casos a tratar y ambas con pruebas basadas en la distribución *ji cuadrada*. Una es la prueba de *concordancia* y la otra es la prueba *de asociación o de independencia*.

Prueba de Concordancia para dos o más proporciones.

Esta es una generalización de la prueba para una proporción, cuya distribución es binomial, y se asocia a una distribución multinomial. Se puede considerar en el contexto de una partición de la población en *k* clases cada una de las cuales representa una proporción P_i de la población, de modo que $\sum_{i=1}^{k} P_i = 1$. Se trata de probar si la proporción de cada clase tiene o no ciertos valores reales específicos P_{i0}. Esta prueba tiene importantes aplicaciones en genética en relación a las leyes de Mendel. El esquema de prueba es el que sigue.

1° Las hipótesis son $H_0 : P_1 = P_{10}$, $P_2 = P_{20}$, ..., $P_k = P_{k0}$ versus H_1: $\exists\ P_i \neq P_{i0}$, con $\sum_{i=1}^{k} P_{i0} = 1$.

2° El nivel de significación α será el seleccionado por el investigador.

3° Esta prueba se realiza con las *frecuencias observadas* (o_i) de cada clase, obtenidas a partir de una muestra aleatoria tamaño *n* de la población. Para este propósito se debe calcular la *frecuencia esperada* (e_i) de cada clase, bajo lo que establece la hipótesis nula. Es necesario resaltar que la proporción de cada clase respecto al resto sigue una distribución binomial de parámetros *n* y P_i, luego el *valor esperado* de cada clase, bajo la hipótesis nula es $e_i = n * P_{i0}$. Es decir, lo que se espera es que la muestra se distribuya proporcionalmente en cada clase como establece H_0. Se deben cumplir las siguientes relaciones $\sum_{i=1}^{k} o_i = \sum_{i=1}^{k} e_i = n$.

El estadígrafo de prueba, para valores de *n* suficientemente grande y valores de $o_i \geq 4$, es $D^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$, cuya distribución es aproximadamente ji cuadrada con (*k - 1*) grados de libertad, cuya notación es $\chi^2\ (k-1)$.

4° La región crítica es del tipo unilateral derecha , lo que es usual para pruebas basadas en la distribución ji cuadrada, pues se rechazará si las diferencias entre lo observado y lo esperado son grandes, luego $RC = \{ D^2 / D^2 > \chi^2_{1-\alpha}(k-1) \}$.

Ejemplos 6.1

a) Se asegura que una mezcla de semillas para césped contiene tres variedades de pasto, *lolium perenne*, *lawn grass* y *festuca rubra* en proporciones de 20% , 50% y 30% respectivamente. Se desea corroborar tal información para lo cual se hace el siguiente planteamiento:

1) $H_0 : P_1 = 0,20$, $P_2 = 0,50$, $P_3 = 0,30$ versus H_1: $\exists\ P_i \neq P_{i0}$, donde la clase 1 es lolium, la clase 2 es lawn grass y la 3 corresponde a festuca.
2) Se usará α= 0,05

3) el estadígrafo de prueba a utilizar es $D^2 = \sum_{i=1}^{3} \frac{(o_i - e_i)^2}{e_i} \approx \chi^2\ (2)$.

4) La región crítica es $RC = \{ D^2 / D^2 > \chi^2_{0,95}(2) = 5,991 \}$.

5) Con el objeto de corroborar o rechazar la hipótesis nula se ponen a germinar 300 semillas de la mezcla. Días después se identifica la especie de cada brote y se cuenta por especie, obteniéndose la siguiente distribución: 70 brotes de lolium, 120 de lawn grass y 110 de festuca, que corresponden a las frecuencias observadas (o_i). Se deben calcular las respectivas frecuencias esperadas $e_i = 300 * P_i$, todo lo cual se resume en la siguiente tabla, a partir de la que se obtiene D^2:

var	*l.p*	*l.g*	*f.r*	Total
o_i	70	120	110	300
e_i	60	150	90	300

$\Rightarrow\ D^2 = \frac{(70-60)^2}{60} + \frac{(120-150)^2}{150} + \frac{(110-90)^2}{90} = 12,1 \in RC$

$\Rightarrow$ rechazar H_0.

6) Los datos obtenidos en la muestra de 300 semillas establecen que es muy improbable que sea verdadera la afirmación de que la proporción de las especies sea la especificada en la hipótesis nula, al nivel del 5%.

b) En genética en un cruce dihíbrido entre dos plantas heterocigóticas de guisantes, cada una con el genotipo *RrAa* y genes independientes, pueden producir uno de los tipos de gametos *RA* ó *Ra* ó *rA* ó *ra* , donde *R* representa el alelo dominante de la forma *redondeada*, *r* el alelo recesivo *rugoso*, *A* el alelo dominante de *color amarillo* y *a* el alelo recesivo de *color verde*. Según la Ley de Mendel, la segregación de caracteres independientes, *RA , Ra , rA y ra* se dan en la proporción 9:3:3:1. Para corroborar la ley anterior se analizaron 480 casos encontrándose la siguiente segregación fenotípica: 282 del tipo *RA* , 80 del tipo *Ra* , 95 del tipo *rA* y 23 del tipo *ra.* ¿ Los datos muestrales obtenidos entregan evidencia suficiente para contradecir la Ley de Mendel ?

1)Las hipótesis $H_0 : P_1 = 9/16,\ P_2 = 3/16\ ,\ P_3 = 3/16\ ,\ P_4 = 1/16$ versus H_1: $\exists\ P_i \neq\ P_{i0}$, donde las clases 1 , 2, 3 y 4 representan respectivamente a los tipos *RA* , *Ra* , *rA* y *ra.*

2) Se procederá con α= 0,05.

3) El estadígrafo es $D^2 = \sum_{i=1}^{4} \frac{(o_i - e_i)^2}{e_i} \approx \chi^2\ (3)$.

4) La región crítica es $RC = \{D^2/D^2 > \chi^2_{0,95}(3)\ = 7,815\}$.

5) La tabla resume la información

Fenotipo	**RA**	**Ra**	**rA**	**ra**	**Total**
o_i	282	80	95	23	480
e_i	270	90	90	30	480

$\Rightarrow D^2 = 3,56 \notin RC \Rightarrow$ aceptar H_0.

6) Los datos muestrales obtenidos no entregan evidencia suficiente que permitan refutar la Ley de Mendel.

c) Para establecer si existen o no diferencias entre productores lecheros respecto a su preferencia por 5 marcas de insumos, se realiza una encuesta cuyo resultado se resume en el siguiente cuadro

Marca	M_1	M_2	M_3	M_4	M_5
N°preferencias	28	25	35	39	28

1) Las hipótesis son $H_0 : P_1 = P_2 = P_3 = P_4 = P_5 = 1/5$ versus H_1: $\exists\ P_i \neq 1/5$.

2) Se elige α = 0,05.

3) El estadígrafo de prueba es $D^2 = \sum_{i=1}^{5} \frac{(o_i - e_i)^2}{e_i} \approx \chi^2\ (4)$.

4) La región crítica es $RC = \{D^2/D^2 > \chi^2_{0,95}(4)\ = 9,488\}$.

5) La tabla indica los valores observados y los esperados de donde resulta que $D^2 = 4,3 \notin RC \Rightarrow$ aceptar H_0.

Marca	M_1	M_2	M_3	M_4	M_5	**Total**
o_i	28	25	35	39	28	155
e_i	31	31	31	31	31	155

6) La evidencia muestral no es suficiente para establecer que las preferencias de los productores lecheros se incline por alguna de las marcas.

Observaciones.

1) En las tablas de clasificación simple con *k* categorías, como las que se utilizan en las pruebas de concordancia, el valor esperado sólo es necesario calcularlo para (***k - 1)*** de las categorías, pues la última resulta por diferencia, lo que explica los (*k - 1*) *grados de libertad* de la distribución ji cuadrada del estadígrafo de prueba.

2) La prueba de hipótesis para una proporción $H_0 : P = P_0$ versus la **hipótesis bilateral**
H_1: $P \neq P_0$, es equivalente a una prueba de concordancia para **dos** proporciones, cuyas hipótesis son $H_0 : P_1 = P_{10}$, $P_2 = P_{20}$ versus H_1: $P_1 \neq\ P_{10}$ y $P_2 \neq\ P_{20}$, donde $P_2 = 1 - P_1 = Q_1$.

Prueba de Independencia.

Consiste en determinar si existe o no asociación entre las categorías de dos variables cualitativas *A* y *B*, cuya estructura corresponde a una clasificación cruzada, denominada tabla de contingencia, con *a* categorías de *A* y *b* categorías de *B* lo que involucra *a* x *b* celdas o casillas. En esta tabla se distinguen dos distribuciones, las de filas o categorías de *A* y las de columnas o categorías de *B*, probabilidades estimadas por los valores muestrales y que por ubicarse en los márgenes se llaman *distribuciones marginales*.

$A\backslash B$	B_1	B_2	....	B_j	.....	**Distr.filas**
A_1						$p_{1.}$
A_2						$p_{2.}$
.....						.....
A_i				p_{ij}		$p_{i.}$
.....						.....
Distr.columnas	$p_{.1}$	$p_{.2}$	....	$p_{.j}$	.....	**1,0**

Tabla 6.1 Distribución conjunta y marginales de probabilidad

La suma, tanto de la distribución de filas como la de columnas, es igual a 1,0 por corresponder al total. Cada casilla contiene, en esta tabla, la probabilidad de ocurrencia conjunta de la categoría i de *A* y la categoría j de *B*. Además, se cumplen las siguientes igualdades $\sum_{i=1}^{a}\sum_{j=1}^{b} p_{ij} = 1{,}0$; $\sum_{i=1}^{a} p_{i.} = 1{,}0$; $\sum_{j=1}^{b} p_{.j} = 1{,}0$; $\sum_{j=1}^{b} p_{ij} = p_{i.}$; $\sum_{i=1}^{a} p_{ij} = p_{.j}$. Cuando la distribución marginal de A y de B son **independientes**, entonces $p_{ij} = p_{i.} * p_{.j}$, es decir, la probabilidad conjunta es el producto de las marginales, como ocurre con la distribución de vectores aleatorios discretos.

El esquema a seguir en la prueba es :

1º Las hipótesis son:

H_0 : Existe independencia entre las categorías de A y de B

versus H_1 : Existe asociación entre las categorías de *A* y de *B*.

2° Se fija el nivel α de la prueba

3° Al igual que en la prueba anterior ésta se realiza con las frecuencias observadas y esperadas por cada celda. Las *frecuencias conjuntas observadas* (o_{ij}), son las que se obtienen con la muestra aleatoria de la población, donde a cada individuo se les mide dos características, por ejemplo sexo y estado civil o condición del sellado de tarros de alimentos y turno en que se produjeron los tarros. La frecuencia esperada se obtiene con las probabilidades marginales como se muestra en la tabla 6.1. Sin embargo tales probabilidades son desconocidas, razón por la cual deben ser estimadas con los datos muestrales. Sea *n* el tamaño muestral, f_i y c_j las frecuencias marginales de filas y columnas respectivamente. Luego $\hat{p}_{i.} = f_i / n$ y $\hat{p}_{.j} = c_j / n$ son los estimadores de las frecuencias marginales. Las frecuencias conjuntas esperadas (e_{ij}) son obtenidas bajo la hipótesis H_0 de independencia, lo que implica que:

$e_{ij} = n * \hat{p}_{ij} = n * \hat{p}_{i.} * \hat{p}_{.j} = n * \frac{f_i}{n} * \frac{c_j}{n} = f_i * c_j / n.$

El estadígrafo de prueba a utilizar es $D^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$ cuya distribución, cuando *n* suficientemente grande y $o_{ij} \geq 4$, es aproximadamente ji cuadrada con $(a-1)*(b-1)$ grados de libertad, denotada por $\chi^2\ ((a-1)(b-1))$.

4° La región crítica es $RC = \{ D^2 / D^2 > \chi^2_{1-\alpha}((a-1)(b-1)) \}$

Ejemplo 6.2

En una encuesta a 500 productores de trigo se les consultó sobre su superficie sembrada y la tecnología empleada en su predio. Posteriormente fueron clasificados en tres categorías de tamaño y tres niveles de tecnología , dando origen a la siguiente información:

Tamaño\Nivel tecnológico	Bajo	Mediano	Alto
Pequeño	110	60	30
Mediano	70	60	50
Grande	20	40	60

¿ la información obtenida permite establecer, al nivel del 5%, que existe asociación entre el tamaño del predio y el nivel tecnológico de éste ?

1) Se plantean H_0 : Existe independencia entre el tamaño del predio y su nivel tecnológico y H_1 : Existe asociación entre el tamaño del predio y su nivel tecnológico.
2) Se fija el nivel α = 0,05.
3) El estadígrafo a utilizar es $D^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \approx \chi^2\ (4)$.
4) La región crítica es $RC = \{ D^2 / D^2 > \chi^2_{0,95}(4) = 9,488 \}$.
5) La tabla muestra las frecuencias observadas , esperadas(1) y los totales marginales.

Tamaño\Niv. tec.	**Bajo (1)**	**Mediano (2)**	**Alto (3)**	**Total fila (f_i)**
tipo frecuencia	***obs esp***	***obs esp***	***obs esp***	
Pequeño (1)	**110 *80***	**60 *64***	**30 *56***	**200**
Mediano (2)	**70 *72***	**60 *57.6***	**50 *50.4***	**180**
Grande (3)	**20 *48***	**40 *38.4***	**60 *33.6***	**120**
Total columna (c_j)	**200**	**160**	**140**	**500**

El valor de $D^2 = \frac{(110-80)^2}{80} + \frac{(60-64)^2}{64} + \frac{(30-56)^2}{56} + \frac{(70-72)^2}{72} + \ldots\ldots + \frac{(60-33.6)^2}{33.6} = 60,9$, pertenece claramente a la región critica, lo que lleva a rechazar la hipótesis nula.
6) Con la información aportada por la muestra se debe concluir que el nivel tecnológico está asociado al tamaño del predio, al nivel del 5%.

Nótese que con los 4 valores esperados calculados (en el pie de página) basta, porque los restantes salen por diferencia con las frecuencias marginales que están determinadas por las frecuencias observadas obtenidas en la muestra. Este argumento explica los *4 grados de libertad* de la distribución.

Observaciones.

1) Si las dos variables categóricas son de dos niveles cada una, lo que da origen a una tabla de contingencia 2 x 2, su distribución es ji cuadrada con 1 grado de libertad. En este caso se debe realizar una corrección, denominada *de Yates por continuidad* para variables discretas y

(1)El cálculo de las frecuencias esperadas se realiza según la fórmula $e_{ij} = f_i * c_j / n$, por la cual
$e_{11} = 200*200/500 = 80$; $e_{12} = 200*160/500 = 64$; $e_{21} = 180*200/500 = 72$; $e_{22} = 180*160/500 = 57,6$

que consiste en que el estadígrafo sea $D^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|o_{ij} - e_{ij}| - 0{,}5)^2}{e_{ij}} \approx \chi^2\ (1)$. Esta correción es conservadora, pues el valor calculado corregido es menor que el sin corregir, lo que trae como consecuencia que en los casos en que el valor sin corregir está rechazando al límite la hipótesis nula, con el valor corregido puede que ésta no se rechace.

2) Cuando en una tabla de contingencia la muestra se toma determinando a-priori la frecuencia marginal de filas o columnas, a diferencia de lo que sucede si es la muestra la que determina estas frecuencias, el desarrollo de la prueba se sigue realizando en los términos ya explicados, pero algunos autores sugieren un cambio, sutil, en el planteamiento de las hipótesis y la denominan Prueba de Homogeneidad, pues la hipótesis nula establecería que " la proporción de individuos en cada columna (fila) es igual para cada fila (columna)", es decir, que la distribución porcentual es la misma en todas las columnas (filas), dependiendo si son los totales marginales de filas (columnas) los que se establecen a-priori. De esta manera se dice que se está estableciendo si las categorías de *A* (*B*) son homogéneas en relación a las categorías de *B* (*A*). Como ilustración se utilizará el ejemplo 6.2. en el que a-priori se determina que la encuesta se le aplicará a 100 productores grandes (G), 150 medianos (M) y 250 pequeños (P), entonces la proporción G : M : P es 100 : 150 : 250, o sea, 2 : 3 : 5 y se plantea si esta proporción se da en los tres niveles tecnológicos. Si así fuera, se concluye que las tres categorías de tamaño de productores es *homogénea* en relación a su nivel tecnológico.

3) Puede establecerse que la prueba de hipótesis para dos proporciones $H_0 : P_1 = P_2$ versus la hipótesis alternativa **bilateral** $H_1 : P_1 \neq P_2$, es totalmente equivalente a una prueba ji cuadrada de una tabla de contingencia 2 x 2, para el mismo nivel de significación.

EJERCICIOS Y PROBLEMAS A RESOLVER

I. ESTADISTICA DESCRIPTIVA

1. Represente gráficamente de dos maneras diferentes la información del número de cajas exportadas de las siguientes especies y concluya cuál gráfico es más ilustrativo.

Especie	N°de cajas(miles)
Uva blanca	185
Uva negra y rosada	157
Pómaceas	215
Carozos	139

2. Las causas más frecuentes de atención en caninos en una clínica veterinaria de la comuna de Santiago en dos épocas del año se presenta a continuación:

Causa	N°atenciones Verano	N°atenciones Invierno
Neumonía	15	48
Gastritis	55	58
Enteritis	50	41
Parasitismo	60	52
Distemper	24	56
Dermatitis	8	4
Traumatismos	20	20

a) Construya un gráfico de sectores circulares por cada época de atención
b) Construya un gráfico para comparar las causas de atención, sin considerar la época, que sirva para destacar la moda.
c) Construya un gráfico en que resalte las causas más importantes en verano y en invierno.
d) Construya otro gráfico en que se puedan comparar las épocas por causa en el cual se destaque la época en la cual es más crítico el distemper, así como la gastritis.

3. En una encuesta a dueñas de casa de Ñuñoa y de San Miguel sobre las tres frutas más consumidas en su hogar durante el año, se obtuvo la siguiente información:

Fruta	Ñuñoa	San Miguel
Uva de mesa	20	16
Duraznos	22	12
Manzanas	17	24
Peras	12	12
Naranjas	10	18
Kiwis	27	10
Guindas	12	8

a) Interprete correctamente y en forma precisa el significado de los números 10 y 18 en naranjas.
b) Represente estos datos en un gráfico adecuado que destaque las preferencias en cada comuna

c) Construya otro gráfico que permita la comparación adecuada entre las comunas y responda ¿ en cuál comuna se consume más uva y en cuál se consume más pera ?. No se deje guiar por los valores absolutos.

4. En una encuesta a 600 productores de trigo se les consultó sobre la superficie sembrada y la tecnología empleada en su predio. Posteriormente fueron clasificados en tres categorías de tamaño y tres niveles de tecnología , dando origen a la siguiente información:

Tamaño\Nivel tecnológico	Bajo	Mediano	Alto
Pequeño	182	85	33
Mediano	68	60	72
Grande	20	41	39

a) Construya un gráfico que permita comparar adecuadamente nivel tecnológico según tamaño. ¿Qué conclusión es posible obtener y por qué?
b) Construya un gráfico adecuado para comparar tamaño según nivel tecnológico ¿Qué conclusión se obtiene?

5. La tabla muestra la distribución de 340 plantas enfermas que fueron sometidas a uno de los cuatro tratamientos curativos A , B , C y D, de acuerdo a su condición después de finalizado el tratamiento:

Tratam.\Condición	Mejor	Igual	Peor
A	13	43	14
B	34	28	38
C	22	18	10
D	35	31	54

Construya gráficos en que se puedan comparar los resultados por tratamiento:

a) En valores absolutos
b) En valores porcentuales
c)¿ Cuál gráfico resulta más adecuado para la comparación y por qué?
d)¿ Cómo conclusión cuál tratamiento resulta más efectivo? Justifique.

6. La información de la tabla corresponde a la producción de carne de ganado bovino(en miles de ton.), por categoría, durante 5 años en un matadero de Santiago:

Año	**Novillos**	**Vacas**	**Bueyes**	**Vaquillas**	**Terneros(as)**
97	90	67	13	60	12
98	97	74	14	64	9
99	94	81	17	70	7
2000	114	85	20	73	6
2001	123	90	21	77	8

a) Construya un gráfico lineal que muestre la producción de carne por categoría
b) Muestre la información anterior mediante un gráfico de barras agrupadas por categoría.
c) ¿Cuál de los dos gráficos resulta más ilustrativo y fácil de interpretar para efecto de comparar entre los años?

7. Los embarques de frambuesas frescas a Europa y USA , durante 8 semanas, en miles de cajas, se resume en la tabla a continuación:

Destino \ Semana	1	2	3	4	5	6	7	8
USA	34	80	48	59	49	83	47	62
EUROPA	10	14	20	27	25	30	13	8

Construya un gráfico adecuado:
a) Que muestre las cajas totales embarcadas
b) Que muestre comparativamente los embarques semanales por destino

8. La tabla especifica la natalidad y mortalidad por cada 1000 habitantes entre 1950 y 1995:

Año	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Natalidad	25.0	23.7	21.3	18.9	16.9	17.9	19.5	23.6	24.6	25.0
Mortalidad	13.2	13.0	11.7	11.3	10.6	10.8	10.6	9.6	9.3	8.5

a) Represente los datos mediante gráficos adecuados, de tres formas diferentes, uno de tipo lineal. ¿Cuál es más clarificador ?
b) ¿Cómo ha sido comparativamente la evolución de la natalidad y de la mortalidad en el tiempo?
c) ¿Qué conclusión puede obtenerse respecto al crecimiento poblacional?

9. Identifique y clasifique las siguientes variables según sean nominales , ordinales , discretas o continuas: (Ind. Piense en como graficaría cada una de ellas. Lo que se pone en el eje *X* es la variable).

a) Procedencia de los vacunos llegados al matadero de Lo Valledor
b) Producción total agropecuaria total durante 2001 por regiones
c) Número de lechones por raza en un criadero de cerdos
d) Número de atenciones diarias por distemper en una clínica veterinaria durante un año calendario
e) Ingreso per cápita de los países de America Latina en el año 2000
f) Número de alumnos por asignatura del ciclo básico
g) Número de asignaturas inscritas por los alumnos de Agronomía durante un semestre académico
h) Temperaturas registradas en una estación meteorológica durante las 24 horas
i) Proporción de manzanas producidas en un huerto por calibre
j) Diámetro de las manzanas cosechadas en un huerto
k) Cantidad de alumnos ingresados a la carrera de Agronomía con puntajes superiores a 700 puntos en los años 1997 , 1998 , 1999 , 2000 y 2001
l) Proporción de plantas sanas y enfermas en un vivero por especie
m) Producción de salmones por países durante 2010

10. En una encuesta a 750 familias se obtuvo la información del número de hijos de cada una de ellas, resumida en la siguiente tabla:

n°hijos	0	1	2	3	4	5	6	7	8
n°familias	40	140	220	160	85	45	25	20	15

a) ¿Cuántas familias tienen 4 hijos?
b) ¿Qué % de familias tiene 3 hijos?
c) ¿Cuántas familias tienen a lo más 3 hijos?
d) ¿ Qué % de familias tiene más de 4 hijos?
e) ¿Qué % de familias tiene 1 ó 2 hijos?
f) Calcule e interprete la media, mediana, moda y desviación estándar del número de hijos por familia.
g) ¿Cuáles de las medidas anteriores resulta más comparativa?

11. Se cuenta el número de arañitas rojas en 50 hojas de un manzano seleccionadas aleatoriamente, obteniéndose los siguientes datos:

8 6 5 3 3 4 0 2 4 5 0 6 5 2 4 6 7 1 4 3 7 6 5 3 0
4 6 2 1 0 3 5 5 4 3 1 1 2 0 6 4 1 3 2 8 4 5 6 2 3

Clasifique los datos en una tabla de frecuencias de variable discreta y resuelva los siguientes puntos:
a) ¿Qué porcentaje de hojas están sanas?
b) ¿Cuántas hojas tuvieron 4 arañitas? ¿qué % representa?
c) ¿Qué % de hojas tuvo a lo más 4 arañitas?
d) ¿Qué % de hojas tuvo más de 5 arañitas?
e) Calcule e interprete las siguientes medidas: rango ; promedio ; moda ; mediana ; desviación estándar.
f) Justifique que medidas permiten una mejor descripción de los datos anteriores.
g) Represente gráficamente los datos, utilizando gráfico de varas y otro de "tallo y hoja" ¿cuál resulta más ilustrativo?

12. El número de preguntas correctamente respondidas por 140 alumnos en una prueba de diagnóstico de Estadística fueron:

42 32 13 18 23 44 41 18 15 25 35 28 17 28 42 51 50 21 27 36
68 84 75 82 68 90 62 88 76 93 73 79 88 73 60 93 71 59 85 75
61 65 75 87 74 62 95 78 63 72 66 78 82 75 94 77 69 74 68 60
46 38 89 21 75 35 60 79 23 31 39 42 27 97 78 85 76 65 71 55
55 80 63 57 78 68 62 76 53 74 66 67 73 81 52 63 76 75 85 47
13 18 23 44 41 18 15 25 66 78 82 75 94 77 69 74 89 21 75 35
57 78 68 62 76 31 39 42 27 97 78 46 38 89 21 75 35 41 18 15

a) ¿ Por qué conviene clasificar estos datos en intervalos, siendo la variable discreta? Clasifíquelos usando 7 intervalos de igual amplitud y a base de la tabulación responda las preguntas a continuación. Compare contando los datos.
b) ¿Cuántos estudiantes obtuvieron menos de 61 ptos?
c) ¿Cuántos estudiantes obtuvieron más de 75 puntos?
d) ¿Qué % de estudiantes obtuvo entre 50 y 70 puntos?
e) Calcule, interprete y compare la media , la mediana
g) Calcule e interprete Q_1 , Q_3 y P_{95}
h) Calcule la varianza y la desviación estándar de los puntajes obtenidos ¿Qué tipo de información entregan estas dos medidas?
i) Confeccione con estos datos un diagrama de "tallo y hoja" y un "boxplot"

13. La tabla corresponde a la clasificación de los pesos de 250 manzanas Granny seleccionadas al azar de la producción de un huerto:

Peso(gr)	**f_i**
$120 \leq X < 135$	15
$135 \leq X < 150$	33
$150 \leq X < 165$	40
$165 \leq X < 180$	45
$180 \leq X < 195$	50
$195 \leq X < 210$	42
$210 \leq X \leq 225$	25
TOTAL	**250**

a) Calcule la media y mediana de los pesos e interprete estos valores
b) Calcule e interprete la varianza , desviación estándar y C.V de los pesos
c) Construya el histograma y el polígono de frecuencias
d) Calcule e interprete P_{10} y P_{75}
e) ¿Que % de las manzanas pesa menos de 140 gr?
f) ¿Cuántas de las 250 manzanas pesan más de 200 gr?
g) ¿Qué % de las manzanas tienen pesos entre $\mu \pm \sigma$?
h) ¿Entre qué pesos está comprendido el 90% central de las manzanas?

14. La información corresponde al peso en kg de 400 lechones destetados a las 3 semanas de edad.

Peso(kg)	**f_i**
$4,1 \leq X < 4,5$	55
$4,5 \leq X < 4,9$	40
$4,9 \leq X < 5,3$	35
$5,3 \leq X < 5,7$	30
$5,7 \leq X < 6,1$	25
$6,1 \leq X < 6,5$	45
$6,5 \leq X \leq 6,9$	50
$6,9 \leq X \leq 7,3$	55
$7,3 \leq X \leq 7,7$	65
TOTAL	**400**

a) Represente gráficamente y con las medidas adecuadas la información y justifique la elección de las medidas
b) ¿Qué puede decir de la variabilidad de los pesos al destete de estos lechones?
c) Si los lechones que pesan menos de 5 kg deben ser sometidos a dieta especial ¿qué porcentaje de ellos están en esta condición?
d) ¿Cuántos de los 400 lechones pesarán entre 5,5 y 7,0 kg?
e) Si se deben seleccionar los 150 lechones de mayor peso ¿a partir de qué peso deben ser elegidos?
f) ¿Cuántas de las 250 manzanas pesan más de 200 gr?
g) ¿Es posible suponer con esta muestra que la población tiene distribución normal?

15. Calcule el promedio ponderado de un alumno que obtuvo en un ramo las siguientes calificaciones con sus correspondientes ponderaciones:

Notas	Ponderación
4,5	1
3,2	2
5,4	3
5,0	2

16. Un inversionista posee tres tipos de acciones A , B y C en proporción $3:7:5$ ¿cuál es su ganancia promedio por acción si la ganancia de las acciones tipo A , B y C son \$250 , \$380 y \$170 respectivamente ?

17. Un grupo de 90 estudiantes , cuyo peso promedio es de 66,47 kg , viaja distribuido en dos buses A y B. Se sabe que el peso promedio de los estudiantes del bus A es 67,70 kg y el peso promedio de los del bus B es 65,40 kg ¿cuántos estudiantes viajan en cada bus ?

18. En una empresa el sueldo promedio de sus empleados es de \$225.000. La empresa decide mejorar sus sueldos reajustándolos en un 12% más una bonificación fija por trabajador de \$ 22.500 ¿cuál es el nuevo sueldo promedio de los trabajadores de la empresa ?

19. En un predio se determinó el porcentaje de animales enfermos y el número de cabezas por raza , los que se resumen en la tabla:

Raza	% de enfermos	n°de cabezas
Hereford	2,5%	1200
Angus	3,4%	800
Charolais	5,0%	2400

a) Calcule el número de animales enfermos por raza
b) Calcule el promedio simple del porcentaje de animales enfermos en el predio
c) Calcule el porcentaje total de animales enfermos en el predio.
d) ¿Cuál de los dos porcentajes es el real ?

20. Durante un mes los siguientes ingredientes de una ración tuvieron la variación de precios que se indican:

Ingredientes	% variación	costo ingrediente
Maíz	10	15
Cebada	-6	5
Heno	-8	4
Afrechillo	5	6
Harina pescado	7	9
Otros	12	3

a) Calcule la variación promedio en el mes, sin considerar el costo de los ingredientes
b) Calcule la variación promedio en el mes , considerando el costo de los ingredientes
c) ¿Cuál de los valores representa **bien** la variación en el costo de la ración ?

21. Un enfermo obtuvo los siguientes resultados en 3 exámenes :A= 50,35; B= 5,48; C= 0,03
Se sabe que estas pruebas en individuos sanos se caracteriza por los siguientes valores:

Examen	Promedio	Desv. est.
A	45,20	3,432
B	5,31	0,574
C	0,02	0,003

¿En cuál de los tres exámenes tiene **peor** resultado el enfermo, si valores altos son malos?

22. Se deben reponer pantallas de monitores de computador para lo cual se consulta a dos fabricantes. El primero produce pantallas con una duración media de 18250 horas y una desviación estándar de 450 horas ; el segundo produce pantallas con una duración media de 18780 horas y una desviación y una desviación estándar de 1950 horas. Si el costo de ella es similar ¿ cuál marca de pantalla recomendaría y por qué ?

23. Se midió el peso de los huevos de 300 gallinas ponedoras Leghorn alimentadas con una dieta X, mientras que otras 200 se alimentaron con la misma dieta más un aditivo vitamínico, todas de la misma edad, obteniéndose la siguiente información resumida:

Dieta X : 56 ± 12
Dieta X + vitamina: 59 ± 8

a) ¿Le parece adecuada como está expresada la información?
b) ¿Qué comentario le merece la comparación del efecto de ambas dietas en el peso de los huevos?

24 Tres atletas A, B y C a ser seleccionados para el Inter-Universitario marcaron los siguientes tiempos en 5 ensayos de los 100 metros planos.

A: 11,1 ; 11,0 ; 11,8 ; 15,8 ; 11,1
B: 11,3 ; 11,4 ; 11,5 ; 11,6 ; 11,4
C: 10,9 ; 11,0 ; 11,8 ; 11,7 ; 11,6

a) ¿ Basándose en medidas de posición y dispersión, a cuál atleta seleccionaría y por qué ?
b) Confeccione un boxplot con esta información ¿la conclusión es la misma?

25. ¿Qué porcentaje de las observaciones de una población queda comprendida entre la percentila 32 y la percentila 68 ?

26. La producción diaria de leche, en litros , obtenida por 7 productores son 1.000, 500, 800, 2.000, 1.350, 950, 23.500.
Calcule la producción promedio diaria del conjunto de los productores y explique por qué no es representativa. ¿ Cuál medida sería más representativa ?

27. Si Ud. tuviera que decidir la compra de sólo un tipo de hamburguesa de vacuno , cerdo , pollo o pavo para una "hamburguesa party" con un grupo de 30 amigos ¿ en qué medida estadística se basaría para tomar la decisión de que tipo comprar, si el precio no es relevante ?

28. Ud. como Jefe de Producción de una empresa agroindustrial está estudiando producir un nuevo concentrado de fruta donde tiene 3 posibilidades de saborizante : "suave" , "medio" , "intenso". Para ello prepara muestras de las tres situaciones y la da a degustar en Supermercados. ¿ En qué medida estadística basaría su decisión de cual saborizante utilizar en el concentrado?

29. Si la producción agropecuaria en una cierta región creció en 30% entre 1995 y 1998 y disminuyó en el mismo porcentaje entre 1998 y 2001 ¿son iguales la producción agropecuaria en 2001 y 1995 ? Explique porcentualmente.

30. De la tabla de frecuencia del problema 4 , ¿ cuál es el :
a) % de predios tamaño mediano?
b) % de predios con nivel tecnológico alto?
c) % de predios de nivel tecnológico alto y de tamaño pequeño?

d) % de predios con nivel tecnológico alto de tamaño pequeño?

31. El siguiente cuadro corresponde a la distribución de edades de los padres en un colegio.

Mujeres / Hombres	20-25	25-30	30-35	35-40	40-45	45-50
20-25	5	3	0	0	0	0
25-30	8	10	2	0	0	0
30-35	2	7	12	4	0	0
35-40	0	8	18	12	2	0
40-45	0	0	4	3	6	7
45-50	0	0	3	5	7	15

a) ¿qué porcentaje de las madres tienen entre 30 y 35 años?
b) ¿qué porcentaje de los hombres tienen edades entre 40 y 50 años?
c) calcule los promedios de edades de hombres y mujeres y **comente** cual es la diferencia de edades entre los padres y las madres.

32. Si las 250 manzanas del problema 13 es una "muestra representativa" de la producción del huerto y éste produce 75 ton. , obtenga una estimación del número de cajas exportable de 20 kg de este huerto , si se sabe que el peso de las manzanas de exportación pesan entre 160 gr. y 200 gr. y que se produce un 8% de descarte por diferentes motivos.

Problemas área de la salud

1.. El cuadro resume la frecuencia de 260 pacientes aquejados de un tipo de gripe, que fueron sometidos a uno de los tratamientos A, B o C, y su condición después del tratamiento.

Trat\ Condición	Mejor	Igual	Peor	Total
A	**42**	**54**	**24**	**120**
B	**33**	**15**	**12**	**60**
C	**32**	**28**	**20**	**80**

a) Represente gráficamente la información anterior con el objeto de mostrar cuál de los tratamientos produce una mayor mejoría. Tenga en cuenta la diferencia de frecuencia en cada tratamiento.
b) Concluya cuál de los tratamientos es más efectivo para aliviar la gripe.

2. En un estudio sobre las condiciones de salud en dos comunas marginales del Norte y del Sur de la RM, se inspeccionaron 500 y 400 niños de entre 5 y 10 años respectivamente en cada población, en relación al número de quistes de Giardosis en fecas, cuyos datos se resume en la tabla a continuación.

Número quistes	frec.*N*	frec.*S*
0	**35**	**75**
1	**70**	**120**
2	**105**	**60**
3	**135**	**45**
4	**80**	**40**
5	**55**	**35**
6	**20**	**25**
TOTAL	**500**	**400**

a) **interprete correctamente** el significado de la frecuencia 120 de S.
b) construya un gráfico que muestre comparativamente la situación de ambas comunas.
c) basándose en medidas estadísticas y el gráfico, discuta cuál es la situación comparativa entre ambas comunas.
d) Calcule la mediana y los valores percentiles 5 (5%) y 90 (90%) de cada distribución
e) ¿A qué porcentaje de los niños se les detectó más de tres parásitos?.

3.. En dos poblaciones A y B los pesos promedios de guaguas al nacer y su correspondiente desviación estándar son 2515 ± 40 gr para la población A y 2630 ± 380 gr para la población B.
a) ¿ En cuál población los pesos al nacer son más homogéneos y por qué ?
b) ¿En cuál de las dos poblaciones es más probable encontrar una guagua que pese al nacer menos de 2130 gr ? Suponga que los pesos al nacer distribuyen normal.

4..Con el fin de constatar el sobrepeso en mujeres de estatura media como factor de riesgo del cáncer de mama, a 420 mujeres a las que se les detectó la patología se les registró su peso, lo que se resume en la siguiente tabla.

Peso(kg)	f_i
$41 \leq X < 46$	8
$46 \leq X < 51$	22
$51 \leq X < 56$	35
$56 \leq X < 61$	38
$61 \leq X < 66$	45
$66 \leq X < 71$	53
$71 \leq X \leq 76$	66
$76 \leq X \leq 81$	85
$81 \leq X \leq 86$	68
TOTAL	**420**

a) Represente gráficamente y obtenga el peso promedio de las mujeres con cáncer de mama
b) ¿Qué puede decir de la variabilidad de los pesos de las mujeres con cáncer de mama?
c) Si las mujeres que pesan menos de 53 kg son de peso normal, las que pesan entre 53 y 68 kg tienen sobre peso y las de peso superior a 68 son obesas ¿cuál es la proporción de mujeres en cada una de las categorías?
d) ¿Es posible concluir con esta muestra que la población de pesos de mujeres con cáncer de mama tiene distribución normal?
e) Concluya una posible relación de la obesidad en mujeres como factor de riesgo del cáncer de mama.

II. PROBABILIDADES

1. Determine el espacio muestral S más reducido para los siguientes experimentos:
a) lanzar una moneda y observar todos los resultados posibles
b) examinar sucesivamente tres plantas y observar todos los resultados posibles en cuanto a su condición de sana
c)examinar sucesivamente tres plantas y observar el número de plantas sanas
d) lanzar un par de dados y observar los puntos obtenidos
e) observar la temperatura a las 14 hras. , todos los días de un año
f) en la cosecha de un manzano Granny medir el peso de cada manzana
g) medir el diámetro polar de un kiwi

2. Para cada una de los espacios de probabilidad $(S\,,\,P)$, determine si P es una probabilidad ***bien definida***:
a) $S=\{a\,,\,b\,,\,c\,,d\}$, tal que $P(\{a\})=1/6\,;\,P(\{b\})=1/5\,;\,P(\{c\})=1/3$ y $P(\{d\})=3/10$
b) $S=\{1\,,\,2\,,3\}$, tal que $P(\{1\,,\,2\})$ = 2/5 y $P(\{3\})$ = 3/5
c) $S=\{1\,,\,2\,,\,3\,,\,4\,,5\}$, tal que $P(\{1\})$ = 3/20 ; $P(\{2\,,\,3\})$ = 1/4 ; $P(\{3\})$ = 1/10 ; $P(\{1\,,\,3\,,4\})$ = 3/5

3. Sea S = { 2 ,3 , 5 , 8 } y sea P una función de probabilidad bien definida en S. Encuentre:
a) P(3) si P(2) = 1/3 , P(5) = 1/6 , P(8) = 1/9
b) P(2) y P(3) si P(5) = P(8) = ¼ y P(2) = 2 P(3)
c) P(5) si P({2,3}) = 2/3 , P({2 , 8}) = 1/2 y P(2) = 1/3

4. Sean A , B eventos de un espacio muestral S , tal que P(A) = 3/8 ; P(B) = 2/5 ; P(A ∩ B) = 1/4. Calcule la probabilidad:
a) P(A ∪ B) b) P(A') c) P(A' ∩ B) d) P(A' ∩ B') e) P(A ∪ B')
f) que ocurra A y B g) que ocurra A o B o ambos
h) que ocurra A pero no ocurra B i) que ocurra A o B pero no ambos

5. En cierto lugar hay 16 plantas de las cuales10 están en buen estado, 4 en regular estado y 2 en mal estado.
i) Al seleccionar aleatoriamente una planta ¿cuál es la probabilidad que ésta:
a) esté en buen estado b) no esté en mal estado c)no esté en buen estado

ii) al seleccionar aleatoriamente 2 plantas ¿cuál es la probabilidad que :
a) ambas estén en buen estado? b) ambas estén en mal estado?
c) al menos una esté en buen estado? d) a lo más una esté en mal estado?
e) exactamente una esté en mal estado? f) ninguna esté en mal estado?
g) ninguna esté en buen estado? h) las dos estén en igual estado?

6. De 15 semillas se sabe que hay 10 que producen flores rojas y 5 flores blancas. Se seleccionan 5 semillas al azar y se ponen a germinar ¿cuál es la probabilidad que :
a) ninguna sea de flores blancas? b) una exactamente sea de flor blanca?
c) sean 3 rojas y 2 blancas? d) las 5 sean del mismo color?
e) al menos una sea de flor roja? f) a lo más dos sean de color blanco?

7. Se lanzan dos dados ¿cuál es la probabilidad de obtener:
a) un par de seis b) sólo un seis c) al menos un seis
d) doce puntos e) cinco puntos f) siete puntos

8. De un conjunto de 9 cartas numeradas del 1 al 9 se eligen al azar dos simultáneamente ¿cuál es la probabilidad que:
a) una sea par y la otra impar? b) la suma de los puntos sea par?

9. En un grupo hay 15 hombres de los cuales 8 tienen 21 años cumplidos y 10 mujeres de las cuales 6 son menores de 21 años. Se eligen dos personas al azar ¿cuál es la probabilidad que:
a) ambas tengan 21 años cumplidos? b) ambos sean del mismo sexo?
c) sean de distinto sexo y menores de 21 años?

10. En un departamento universitario de 20 académicos 7 tienen postgrado y ganan a lo menos $1300000, 11 ganan menos de $1300000 y 8 no tienen postgrado y ganan menos de $1300000. Se seleccionan dos académicos al azar para una comisión, ¿ cuál es la probabilidad de que :
a) ambos tengan postgrado ?
b) ambos no tengan postgrado y ganen menos de $1300000 ?
c) ambos ganen menos de $1300000 y uno tenga postgrado pero el otro no?
d) ambos no tengan postgrado, si ambos ganan menos de $1300000?

11. Un club formado por 30 parejas de casados va a elegir un presidente y luego un secretario ¿cuál es la probabilidad que:
a) ambos sean hombres? b) sean de sexo opuesto? c) sea elegido un matrimonio?

12. Para ensayo se requieren ubicar al azar, en 9 parcelas dispuestas en hileras, 3 tratamientos
(A , B , C) , cada uno repetido 3 veces ¿Cuál es la probabilidad que:
a) queden ubicados en el orden A,B,C,A,B,C,A,B,C ?
b) queden los tres tratamientos con sus tres repeticiones en parcelas adyacentes?

13. Una línea de embalaje está alimentada por dos subsistemas A y B. Se ha determinado que P(A falle) = 0,2 ; P(sólo B falle) = 0,15 ; P(A y B fallen juntos) = 0,08. Calcule la probabilidad:
a) que sólo A falle b) que A falle si B ha fallado

14. En una universidad el 25% de los estudiantes falla en Cálculo , el 30% falla en Química y el 10% falla en ambas asignaturas. ¿Cuál es la probabilidad que un alumno cualquiera:
a) falle en Cálculo si ya reprobó Química b) falle en Química si reprobó Cálculo
c) falle en Cálculo o en Química d) no repruebe ninguna

15. Los 16 huertos de una localidad se clasificaron en términos del sistema de riego en *tecnificado (T)* o *surco (S)* y de su tamaño en *mediano (M)* o *pequeño (P).* Se encontraron que 10 huertos son de tamaño pequeño; 8 riega por surco ; 3 de tamaño pequeño y riego tecnificado. (IND. Con los datos confeccione una tabla de 2x2)

Se necesita realizar una encuesta en la localidad para lo cual se deben seleccionar 3 huertos al azar ¿Cuál es la probabilidad de que los 3 huertos seleccionados:
a) sean de tamaño mediano y tengan riego tecnificado ?
b) sean de tamaño pequeño si riegan por surco ?

16. De una caja que contiene 3 fichas azules , 5 blancas y 4 rojas.
i) si se extraen sucesivamente , sin remplazo, 2 fichas. Calcule la probabilidad que:
a) las dos sean rojas b) sean blanca y roja c) se elijan en el orden roja-blanca

ii) si se extraen sucesivamente , sin remplazo, 3 fichas. Calcule la probabilidad que se:
a) extraigan en el orden roja-blanca-azul b) elija una de cada color sin importar orden

17. De los dígitos 2 , 3 , 5 , 6 , 7 , 9 se seleccionan aleatoriamente tres en forma sucesiva,sin sustitución ¿cuál es la probabilidad que:
a) el número obtenido sea menor que 400? b) el número obtenido sea múltiplo de 5?
c) el número obtenido sea par menor que 400 ?

18. En un laboratorio se encuentran 10 plantas de las cuales 6 están sanas. Se examinan las plantas una a una.
i) ¿Cuál es la probabilidad que al examinar las dos primeras:
a) ambas estén sanas? b) ambas estén enfermas?
c) una esté sana y la otra enferma? d) la 2ª esté sana si la 1ª estaba sana?

ii) Si las plantas son examinadas hasta ubicar la 4ª planta enferma , ¿cuál es la probabilidad que la cuarta planta enferma se detecte al examinar:
a) la cuarta planta? b) la quinta planta? c) la décima planta?

19. Probar que P(A/B) $>$ P(A) $\Rightarrow$ P(B/A) $>$ P(B)

20. Sean dos sucesos tales que P(A) = 0,4 y P(A $\cup$ B) = 0,7 . Determine P(B) de modo que los sucesos A y B sean :
a) mutuamente excluyentes b) independientes

21. Al lanzar un par de dados , uno de color blanco y el otro de color rojo. Sea X el resultado del dado blanco e Y el resultado del dado rojo. Sean los sucesos A $= \{(x, y)/x + y = 5\}$ y $B = \{(x, y)/x > y\}$. Calcule la probabilidad de que suceda:
a) A y B b) A ó B c) A, si sucedió B d) B, si sucedió A

22. La probabilidad que en un packing una manzana tenga defectos por golpe de sol es 1/20 y la probabilidad que tenga machucones es 1/8. ¿Cuál es la probabilidad que una manzana tenga:
a) ambos defectos? b) defectos? c) sólo uno de los defectos?

23. La probabilidad que un durazno presente el desorden fisiológico A es 2/5 , que presente el desorden fisiológico B es 1/2 y que presente el desorden fisiológico C es 1/3. ¿ Si los desórdenes A,B y C se presentan en forma independiente, cuál es la probabilidad que un durazno seleccionado aleatoriamente:
a) presente los tres tipos de desórdenes fisiológicos?
b) presente al menos uno de estos desórdenes fisiológicos?
c) presente **sólo uno** de los desórdenes anteriores?

24. En un procesamiento agroindustrial la probabilidad que un producto se contamine con una bacteria A es 0,1 y la probabilidad que se contamine con otra bacteria B es 0,05. ¿Cuál es la probabilidad que en uno de estos procesos el producto se contamine?

25. Si en la elaboración de concentrado de tomate éste se contamina con una bacteria X con probabilidad del 2%, con otra bacteria Y con probabilidad del 5% y con **al menos una** de las dos bacterias con probabilidad del 5,5% , ¿cuál es la probabilidad que el concentrado:
a) no esté contaminado?
b) esté contaminado con **ambas** bacterias?
c) esté contaminado sólo con Y?
d) se contamine con la bacteria X, si está contaminado con la bacteria Y ?
e) ¿según la información anterior la contaminación con X e Y son independientes?

26. Una bolsa A contiene dos fichas rojas numeradas 1 y 2 , respectivamente , y dos fichas blancas numeradas 3 y 4. Otra bolsa B contiene 3 fichas blancas numeradas 5 , 6 y 7 , respectivamente y tres fichas azules numeradas 8 , 9 y 0. Se extraen aleatoriamente dos fichas de cada bolsa , ¿cuál es la probabilidad que:
a) las cuatro sean de igual color?
b) la suma de puntos de cada bolsa sea igual?

27. Demuestre que si A y B son sucesos independientes , entonces también lo son A' y B' y A y B'.

28. Una especie produce semillas de flores de color rojo , blanco y amarillo en porcentajes del 60% , 30% y 10% respectivamente. Los porcentajes de no germinación se sabe que son del 7% , 2% y 4% respectivamente.¿Cuál es :
a) el porcentaje de germinación de esta especie?
b) la proporción de plantas de cada color que se obtendrá en un almácigo?

29. En un vivero un 4% de las plantas de una procedencia A y un 1% de las plantas de otra procedencia B supera los 60 cm. y se sabe que un 60% de las plantas proviene de B. Se selecciona una planta al azar y se verifica que mide 73 cm ¿cuál es la probabilidad que provenga de B?

30. En un viñedo se plantan vides de tres procedencias A , B y C en proporciones del 25% , 50% y 25% respectivamente. La probabilidad que estas vides estén produciendo a los 2 años son respectivamente 0,1 ; 0,2 y 0,4 respectivamente.
a) ¿Cuál es la proporción de vides que estarán produciendo a los 2 años ?
b) ¿Si una planta elegida al azar no está produciendo a los 2 años ,cuál es la probabilidad que provenga de C ?

31. En un vivero una planta puede estar sana o tener una enfermedad A con probabilidad 0,25 u otra enfermedad B con probabilidad 0,35. Al estar sana la probabilidad que **no presente** marchitez en las hojas es 0,9 , al tener la enfermedad A **presenta** marchitez en las hojas con probabilidad 0,70 y al tener la enfermedad B **presenta** marchitez con probabilidad 0,60. ¿Cuál es la probabilidad :
a) que al examinar 5 plantas al azar estén todas sanas?
b) que al examinar 5 plantas al azar haya al menos una tenga la enfermedad A?
c) que una planta cualquiera **no presente** marchitez en las hojas?
d) que una planta esté **sana** , si **presenta** marchitez en las hojas?

32. 400 predios agrícolas de la VII región se clasificaron según su Nivel Tecnológico (Alto (A), Medio (M), Bajo (B)) y Tamaño (pequeño (p) y mediano (m)) . La siguiente tabla indica el número de predios en cada categoría.

Nivel \ Tamaño	pequeño	mediano	Total
Alto	30		50
Medio		50	
Bajo			150
Total		170	

Complete la tabla y calcule la probabilidad de que al elegir un predio al azar éste sea :
a) de Nivel (A) b) de Tamaño (m) c) de Nivel (M) y Tamaño (p)
d) de Nivel (B) o de Tamaño (m) e) no tenga Nivel (B)
f) de Tamaño (p) y no tenga Nivel (B)

33. En una ciudad se publican tres periódicos A, B y C. Una encuesta indicó las probabilidades de que los ejecutivos de una empresa de esa ciudad lean alguno de tales periódicos:
$P(A) = 0,25$, $P(B) = 0,3$, $P(C) = 0,20$, $P(A \cap B) = 0,1$, $P(A \cap C) = 0,12$, $P(B \cap C) = 0,08$ y $P(A \cap B \cap C) = 0,06$ ¿Cuál es la probabilidad de que un ejecutivo cualquiera:
a) no lea ningún periódico? b) lea sólo uno de los periódicos?
c) lea el periódico A o el B? d) lea a lo más uno de los periódicos?

34. En cierta comunidad, la probabilidad de que una familia tenga televisor es 0,64, una máquina lavadora es 0,55 y que tenga ambos artefactos es 0,35. Se selecciona una familia al azar, ¿cuál es la probabilidad de que :
a) no tenga máquina lavadora? b) solamente tenga televisor?
c) no tenga televisor o no tenga máquina lavadora.
d) no tenga televisor ni máquina lavadora.
e) solamente tenga televisor o solamente tenga máquina lavadora.

35. La probabilidad de que un vendedor de tractores, venda por lo menos tres tractores en un día es 0,2. ¿Cuál es la probabilidad de que venda 0, 1 o 2 tractores en un día?

36. En una caja de manzanas de exportación la probabilidad de que haya al menos una manzana mala es 0.05 y de que haya al menos dos malas es 0.01. ¿Cuál es la probabilidad de que la caja contenga :
a) ninguna manzana mala ? b) exactamente una manzana mala ?
c) a lo más una manzana mala ?

37. Un estudio determinó que la probabilidad de que un hombre casado vea un cierto programa de televisión es 0,4 , de que su mujer lo vea es 0,5 y la probabilidad de que el hombre vea el programa, dado que su esposa lo ve es 0,7. ¿ Cuál es la probabilidad de que :
a) una pareja de casados vea el programa ?
b) una mujer casada vea el programa, sabiendo que su esposo lo ve ?
c) solamente uno de ellos vea el programa ?
d) ninguno de los cónyuges vea el programa ?

38. En una empresa el 25% de los empleados son profesionales, el 15% de los empleados llega atrasado y el 10% es profesional y llega atrasado.
Confeccione una tabla de doble entrada con los datos anteriores (IND. Una categoría tiene que ver con si es *profesional* y la otra con la *puntualidad*).
Si se selecciona un empleado al azar, ¿cuál es la probabilidad de que éste:
a) llegue atrasado o sea profesional ?
b) sea profesional y no llegue atrasado ?
c) llegue atrasado, si resulta ser profesional ?
d) no sea profesional, si no llega atrasado ?

39. Sean los sucesos A y B tales que $P(A) = 0,25$, $P(A / B) = 0,5$ y $P(B / A) = 0,25$.
¿Cuáles de las siguientes proposiciones son verdaderas?
i) A y B son sucesos mutuamente excluyentes
ii) $P(A' / B) = 0.75$
iii) $P(A / B) + P(A / B') = 1$

40. La probabilidad de que en cierta ciudad llueva un día del año seleccionado aleatoriamente, es 0,25. El pronóstico local del tiempo atmosférico es correcto el 60% de las veces en que el pronóstico es de lluvia, y el 80% de las veces que se hace otro pronóstico.
a) Determine la probabilidad de que el pronóstico sea correcto en un día seleccionado al azar.
b) Si en un día determinado el pronóstico es correcto, determine la probabilidad de que ese día sea lluvioso.

Respuestas:
4. 21/40 ; 5/8 ; 3/20 ; 19/40 ; 17/20 ; 1/4 ; 21/40 ; 1/8 ; 11/40 **5 i)** 5/8 ; 7/8 ; 3/8 ; **ii)** 3/8 ; 1/120 ; 7/8 ; 119/120 ; 7/30 ; 91/120 ; 1/8 ; 13/30 **6.** 84/1001 ; 350/1001 ; 400/1001 ; 253/3003 ; 3002/3003 ; 834/1001
7. 1/36 ; 5/18 ; 11/36 ; 1/36 ; 1/9 ; 1/6 **8.** 5/9 ; 4/9 **9.** 11/50 ; 1/2 ; 7/50 **11.** 29/118 ; 30/59 ; 1/59
13. 0,12 ; 8/23 **14.** 1/3 ; 2/5 ; 9/20 ; 11/20 **16. i)** 1/11 ; 10/33 ; 5/33 **ii)** 1/22 ; 3/11 **17.** 1/3 ; 1/6 ; 1/10
18. i) 1/3 ; 2/15 ; 8/15 ; 5/9 **ii)** 1/210 ; 2/105 ; 2/5 **20.** 0,3 ; 0,5 **21.** 1/18 ; 17/36 ; 2/15 ; 1/2
22. 1/160 ; 17/160 ; 13/80 **24.** 0,145 **26.** 1/30 ; 2/45 **28.** 94,8% ; 58,9% , 31,0% y 10,1%
29. 3/11 **30.** 0,225 ; 6/31 **32.** 1/8 ; 17/40 ; 3/8 ; 4/5 ; 5/8 ; 9/20 **33.** 0,49 ; 0,33 ; 0,45 ; 0,82
34. 0,45 ; 0,29 ; 0,65 ; 0,16 ; 0,49 **36.** 0,95 ; 0,04 ; 0,99 **37.** 0,35 ; 7/8 ; 0,20 ; 0,45
38. 0,30 ; 0,15 ; 2/5 ; 14/17 **39.** ninguna **40.** 0,75 ; 0,20

III. DISTRIBUCIONES DE PROBABILIDAD

1. Una variable aleatoria $(v.a.)$ discreta X tiene por *función de cuantía , p(x$_i$):*

$$p(x_i) = \begin{cases} 1/8 \text{ si } x_i = 5 \\ 3/8 \text{ si } x_i = 8 \\ 1/2 \text{ si } x_i = 10 \end{cases}$$ Calcule:

a) P($X = 5$) b) P($X = 3$) c) P($X > 8$) d) P($X \geq 8$)

2. Para la variable aleatoria número de hijos varones en una familia de 5 hijos , obtenga la *función de distribución* ***p(x$_i$)*** y mediante ella calcule la probabilidad que una familia tenga:
a) exactamente 2 hijos varones b) ningún hijo varón c) más de 3 hijos varones
d) a lo más 3 hijos varones e) al menos un hijo varón.

3. En un conjunto de semillas de una especie floral hay 5 que corresponden a flores rojas, 3 a flores blancas y 4 a flores amarillas. Sea X la v.a. que especifica el número de semillas rojas obtenidas al seleccionar al azar 5 semillas:
a) obtenga la distribución de probabilidad de X b) calcule P($X = 3$)
c) calcule P($1 \leq X \leq 4$) d) calcule P($1 < X \leq 4$)
e) calcule P($1 \leq X < 4$) f) calcule P($1 < X < 4$)
g) calcule P($X < 3$) h) calcule P($X \geq 2$)

4. Una caja contiene 4 fichas rojas y 6 blancas. **Veinte** fichas son elegidas con remplazo. Si X es el número de fichas rojas elegidas , obtenga la distribución de X y calcule la probabilidad de obtener:
a) exactamente 8 fichas rojas b) ninguna ficha roja c) al menos una ficha roja

5. Una planta de kiwi de un vivero tiene una probabilidad de 0,8 de estar sana. Se seleccionan 10 plantas al azar , obtenga la distribución del número de plantas sanas y calcule la probabilidad de seleccionar:
a) 8 sanas b) ninguna sana c) todas sanas d) al menos una sana.

6. Para cada una de las variables discretas de los problemas anteriores calcule su *esperanza matemática* y su *varianza.*

7. Para cada una de las variables discretas de los problemas anteriores obtenga su función de distribución acumulativa ($f.d.a$) y recalcule las probabilidades pedidas a partir de ella.

8. La v.a. continua X: altura de un quillay en un bosque juvenil, tiene una *f.d.p*. dada por:

$$\text{f(x)} = \begin{cases} \frac{1}{2} - \frac{1}{8}x & \text{si } 0 \le x \le 4 \\ 0 & \text{para otros valores} \end{cases}$$

1) de acuerdo a esta distribución ¿son más frecuentes árboles altos o bajos en este bosque?
2) Calcule la probabilidad que un quillay de este bosque tenga altura:
a) entre 1 y 2 metros b) mayor que 3 metros c) menor o igual que 1,5 m
d) mayor que $\frac{1}{2}$ m y menor o igual a 2,5 m

9. La v.a. X: rendimiento de un cultivo,en qq. por cada 1000m^2, tiene:

f.d.p $f(x)\begin{cases} \frac{1}{45}(36 - x^2) & \text{si } 3 \le x \le 6 \\ 0 & \text{p.o.v} \end{cases}$

a) ¿cuál es la probabilidad que el cultivo rinda entre 4 y 5 qq. ?
b) ¿cuál es la probabilidad que el cultivo rinda más de 4,5 qq ?
c) ¿cuál es el rendimiento promedio de este cultivo ?
d) ¿qué tan homogéneo es el rendimiento de este cultivo ?

10. Una v.a. continua X: longitud de raíz principal de plántulas de nectarines toma valores entre 2 y 8 y tiene una *f.d.p*. de la forma $a(x+3)$, donde a es una constante a determinar . Calcule :
a) el valor de a b) P($3 < X < 5$) c) P($X \ge 4$) d) P(| $X - 5$| $< 0{,}5$)

11. Una v.a. continua X toma valores entre -2 y 1 tiene *f.d.p.* de la forma ax^2. Calcule:
a) el valor de a b) P($X < 0$) c) P($X \ge \frac{1}{2}$) d) P(-1 $\le X < \frac{1}{2}$)

12. Una v.a. X tiene una función de distribución acumulativa ($f.d.a.$)

$$F(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{3}{4}x^2 - \frac{1}{4}x^3 & si\ 0 \le x \le 2 \\ 1 & si\ x > 2 \end{cases}$$, calcule :

a) P($X < 1$) b) P($\frac{1}{2} < X \le \frac{3}{2}$) c) P($X \ge \frac{1}{2}$) d) P($X > \frac{3}{2}$)

13. Para cada una de las variables continuas de los problemas anteriores calcule su *esperanza matemática* y su *varianza.*

14. Para cada una de las distribuciones de los problemas anteriores obtenga la $f.d.a.$, $F(x)$, y **recalcule** las probabilidades pedidas.

15. Sea X v.a con *distribución uniforme* [-2 , 3].
a) obtenga la *f.d.p* , $f(x)$
b) obtenga la *f.d.a* ,$F(x)$
c) calcule P(-1 $< X \le \frac{3}{2}$) y P($0 < X \le \frac{5}{2}$). Compare y explique el por qué de la coincidencia.

16. Encuentre el valor esperado y la varianza para cada una de las siguientes variables aleatorias a) X: número de caras obtenidas al lanzar 5 monedas
b) X: suma de puntos obtenidos al lanzar 2 dados
c) X , con $f.d.p.$ $f(x) = 6x(1-x)$ si $0 \leq x \leq 1$

17. Reconozca la distribución, los parámetros y especifique la función de cuantía $p(x_i)$ de las siguientes v.a. , **justificando cada vez**, y calcule para cada una de ellas $E[X]$ y $V[X]$:
a) X_1: n° de plantas enfermas encontradas al examinar 25 plantas si la probabilidad de enferma es $1/5$
b) X_2 : n° de lesiones en hoja de tabaco, causadas por un virus que provoca en promedio 2 lesiones por hoja
c) X_3: n° de manzanas rojas obtenidas al seleccionar al azar 20 manzanas, **con sustitución**, de una caja que contiene 6 manzanas rojas, 4 manzanas jaspeadas y 2 manzanas verdes.

18. Para cada una de las variables , discretas o continuas, definidas en los problemas anteriores, calcule:

a) $E[2X+1]$; $V[2X+1]$ b) $E[3-X]$; $V[3-X]$

19. Si un día **no llueve** un contratista gana 5 UF y **si llueve** en el día pierde 1,5 UF . ¿ Cuál es su ganancia esperada en los meses de Otoño-Invierno si la probabilidad de lluvia un día cualquiera es de 0,3 ?

20. En un juego se puede ganar $ 50.000 con probabilidad 0,2 , ganar $ 20.000 con probabilidad 0,4 y en caso contrario perder una cierta cantidad de dinero. ¿ cuál es la cantidad de dinero que se debe perder para que el *juego sea justo* ?

21. La función $p(x_i)$ representa la probabilidad de un productor de obtener repollos según calidad:

$$p(x_i) = \begin{cases} 1/6 \text{ primera} \\ 1/2 \text{ segunda} \\ 1/4 \text{ tercera} \\ 1/12 \text{ desecho} \end{cases}$$

Si la ganancia por unidad es $ 150 para primera , $ 105 para segunda , $ 75 para tercera y $ 9 para desecho , calcule la ganancia esperada por el productor por unidad producida.

22. La hoja de la planta de tabaco pierde valor en la medida que el número de lesiones en su hoja sea mayor. Por experiencia se sabe que X: *número de lesiones por hoja* , tiene distribución :

x_i	0	1	2	3	4
$p(x_i)$	1/3	1/4	1/6	1/6	1/12

La ganancia por hoja de un agricultor depende del número de lesiones x_i , según la función $g(x) = 48 - 14x + x^2$.
Calcule la ganancia promedio por hoja del agricultor.

23. En un árbol se determinan las variables $\boldsymbol{X}$ **:** n° de i nsectos/hoja ; $\boldsymbol{Y}$: n° depredadores/hoja . La tabla define la distribución de probabilidad conjunta $p(x_i\ ,\ y_j)$.

$\boldsymbol{Y} \setminus \boldsymbol{X}$	0	1	2	3
0	0,03	0,09	0,08	0,30
1	0,11	0,09	0,06	0,04
2	0,19	0,01	0,00	0,00

a) obtenga las probabilidades marginales $p(x_i)\ ,\ p(y_j)$ e interprételas
b) ¿cuál es la probabilidad que con 2 predadores haya 3 insectos/hoja?
c) ¿ cuáles son los dos sucesos que tienen mayor probabilidad de ocurrir?
d) ¿cuál es la probabilidad que una hoja esté sana?
e) calcule una medida de asociación entre n° de ins ectos y n° de depredadores e interprétela
f) ¿son el n° de insectos/hoja y el n° de depredado res v.a. independientes? **Justifique**

24. En un packing se trabaja en dos turnos. Sea $X:$ n° de veces que falla semanalmente la correa transportadora en turno 1 e Y : n° de veces que falla semanalmente la correa transportadora en turno 2. $X\ e\ Y$ son variables aleatorias independientes y las siguientes son las distribuciones marginales de $X\ e\ Y$:

$$\text{p(x}_\text{i}\text{)} = \begin{cases} 0,50 \ \ si\ x_i = 0 \\ 0,20 \ \ si\ x_i = 1 \\ 0,30 \ \ si\ \ x_i = 2 \end{cases} \qquad \text{p(y}_\text{j}\text{)} = \begin{cases} 0,20 \ \ si\ \text{y}_\text{j} = 0 \\ 0,70 \ \ si\ \text{y}_\text{j} = 1 \\ 0,10 \ \ si\ \ \text{y}_\text{j} = 2 \end{cases}$$

a) ¿cuál es la probabilidad que la correa transportadora durante una semana cualquiera falle al menos una vez en ambos turnos?
b) ¿ en cuál de los dos turnos falla más en promedio la correa transportadora ?

25. Las siguientes tablas corresponden a la distribucion de ***X e Y*** , número de fallas de dos correas transportadoras en un packing, y se sabe que ambas funcionan **independientemente.**

$$p(x_i) = \begin{cases} 0,40 \quad \text{si}\ \ x_i = 0 \\ 0,30 \quad \text{si}\ \ x_i = 1 \\ 0,20 \quad \text{si}\ \ x_i = 2 \\ 0,10 \quad \text{si}\ \ x_i = 3 \end{cases} \qquad p(y_j) = \begin{cases} 0,30 \ \ \text{si}\ \ y_j = 0 \\ 0,35 \ \ \text{si}\ \ y_j = 1 \\ 0,20 \ \ \text{si}\ \ y_j = 2 \\ 0,15 \ \ \text{si}\ \ y_j = 3 \end{cases}$$

a) ¿cuál es la probabilidad que, durante un mes, ambas correas transportadora fallen una vez?
b) ¿ cuál es la probabilidad que durante un mes ambas correas **no** fallen?
c) ¿ cuál es la probabilidad que durante un mes una de las correas **no** falle y la otra falle **al menos una vez** ?
d) ¿cuál es la probabilidad que durante un mes **al menos** una de las correas falle?
e) ¿ cuál de las dos correas falla más en **promedio**?
f) determine la $V[\ X - Y]$
g) aplicando correctamente las propiedades calcule $E[\ 2X - 3Y + 5]$ y $V[2X - 3Y + 5]$

26. Sean X e Y v.a **independientes** con $p(x_i) = \binom{50}{x_i}(0,3)^{x_i}(0,7)^{50-x_i}\ ,\ x_i = 0, 1, 2, ..., 50$ y $p(y_j) = \frac{e^{-12}\,12^{y_j}}{y_j\,!}\ ,\ y_j = 0, 1, 2, 3, ...$ respectivamente. Calcule :

a) $E(3X - 2Y - 3)$ b) $V(3X - 2Y - 3)$

27. Sean X e Y v.a. continuas con una función de densidad conjunta de probabilidad:

$$f(x,y) = \begin{cases} \frac{1}{18}xy^2 & \text{si } 0 \le x \le 2\,,\ 0 \le y \le 3 \\ 0 & p.o.v \end{cases}$$

a) calcule P($X < 3/2\,,\ Y \le 2$)
b) calcule P($X > Y$)
c) obtenga las funciones de distribución marginales $g(x)\,,\ h(y)$
d) calcule $E[X]$ y $E[Y]$
e) calcule $Cov(X,Y)$
f) ¿son X e Y v.a. independientes?

28. Un aserradero que procesa madera de Pino y Eucaliptus estableció la siguiente función de densidad conjunta para la proporción de madera con nudos de Pino (X) y de Eucaliptus (Y):

$$f(x\,,\,y) = \begin{cases} \frac{2}{3}(x+2y) & \text{si } 0 \le x \le 1\,,\ 0 \le y \le 1 \\ 0 & p.o.v \end{cases}$$

a) obtenga las funciones de densidad marginales de X e Y y explique su significado de acuerdo al enunciado
b) calcule la probabilidad de obtener menos del 25% de madera de Pino **con nudo** y más del 70% de madera de Eucaliptus **sin nudo**
c) calcule la probabilidad de obtener a lo menos el 50% de madera de Pino y de Eucaliptus **sin nudo**.
d) calcule la probabilidad de obtener entre el 20% y el 80% de madera de Pino **con nudo**
e) ¿cual es el % esperado de madera de Pino **con nudo** ?
f) ¿cuál es el % esperado de madera de Eucaliptus **sin nudo** ?
g) ¿son **X e Y** v.a. independientes? **Justifique matemáticamente**

29. Dos variables aleatorias independientes X e Y tienen distribuciones dadas por $f(x)$ y $f(y)$ respectivamente.

$$f(x) = \begin{cases} \frac{2}{5}(x+2) & \text{si } 0 \le x \le 1 \\ 0 & \text{para otros valores} \end{cases} \quad ; \quad f(y) = \begin{cases} \frac{3}{4}(2y - y^2) & \text{si } 0 \le y \le 2 \\ 0 & \text{para otros valores} \end{cases}$$

a) ¿cuál es la probabilidad que Y tenga valores mayores que 1 ?
b) ¿ es homogéneo el comportamiento de X ?
c) ¿ cuál es la $P(X > \frac{1}{2}$ e $Y < \frac{1}{2})$?
d) calcule $E\left[X^2 + 2X + 5\right]$
e) calcule $V\left[7 - 3X - 2Y\right]$

Respuestas:

1. 1/8 ; 0 ; 1/2 ; 7/8 **2.** 0,3125 ; 0,03125 ; 0,1875 ; 0,8125 ; 0,96875 **3.** 35/132 ; 770/792 ; 595/792 ; 735/792 ; 560/792 ; 546/792 ; 596/792 **4.** 0,17971 ; 0,00004 ; 0,99996 **5.** 0,30199 ; 0 ; 0,10737 ; aprox. 1
8. 5/16 ; 1/16 ; 39/64 ; 5/8 **10.** a = 1/48 ; 7/24 ; 3/4 ; 1/6 **11.** a =1/3 ; 8/9 ; 7/72 ; 1/8
12. 1/2 ; 11/16 ; 31/32 ; 5/32 **16. a)** 5/2 ; 5/4 **b)** 7 ; 103/18 **c)** 1/2 ; 1/20 **19.** UF 3,95
20. $ 45.000 **21.** $ 97 **23. e)** Cov(X,Y) = -0,693 , ρ = -0,704 **27.** 1/6 ; 8/45 ; g(x) = $\frac{1}{2}$x , h(y) = $\frac{1}{9}y^2$; E(X) = 4/3 , E(Y) = 9/4 , Cov(X,Y) = 0 ; si. **28. b)** 77/800 ; **c)** 1/8 ; **d)** 3/5 ; **e)** 55,6% ; **f)** 38,9%

IV. DISTRIBUCION NORMAL

1. Sea $\boldsymbol{Z} = \boldsymbol{N}(0,1)$, calcule:
a) $P(Z < -1,58)$ b) $P(Z < 1,86)$ c) $P(-0,63 < Z \leq 0,84)$
d) $P(-1,22 \leq Z < -0,72)$ e) $P(Z > 0,93)$ f) $P(Z \geq -0,55)$

2. Calcule $\boldsymbol{a}$ tal que:
a) $P(Z < a) = 0,2332$ b) $P(Z \leq a) = 0,7448$
c) $P(Z \geq a) = 0,6913$ d) $P(Z > a) = 0,05$

3. En la asignatura de Estadística las notas tuvieron una media de 4,5 y una desviación estándar de 0,4 , mientras que en Botánica las notas tuvieron una media de 5,8 y una desviación estándar de 0,8. El alumno Veas obtuvo 4,8 en Estadística y 6,0 en Botánica ¿en cuál de las dos asignaturas el alumno Veas tuvo un rendimiento más destacado?

4. Sea la $v.a.$ $\boldsymbol{X} = N(12\ ,\ 4)$, calcule:
a) $P(X < 15)$ b) $P(X \leq 11)$ c) $P(X > 14)$ d) $P(13 < X \leq 14,5)$
e) $P(9 \leq X < 10,5)$ f) $P(10 \leq X < 14)$ g) $P(X > 10,5)$ h) $P(X < 12)$

5. Sea una $v.a.$ $X = N(10\ ,\ 25)$, calcule el valor de $\boldsymbol{a}$ si:
a) $P(X < a) = 0,0314$ b) $P(X \leq a) = 0,7820$
c) $P(X > a) = 0,4772$ d) $P(X \geq a) = 0,6528$

6. Se establece que las calificaciones en un examen de portulación tiene distribución normal con media 73 y desviación estándar 8 ¿cuál es la probabilidad que un alumno seleccionado al azar haya obtenido:
a) a lo más 60 puntos? b) entre 65 y 89 puntos? c) más de 80 puntos?

7. Si el número de alumnos que rinde el examen de postulación anterior es 640 ¿cuántos tendrán:
a) menos de 55 puntos? b) entre 65 y 81 puntos? c) más de 90 puntos?

8. Se asume que la distribución de pesos de manzanas Granny (en gr) , en un huerto , tiene distribución $N(160\,,625)$.

i) ¿Qué proporción de las manzanas del huerto pesa:
a) entre 145 y 190 gr? b) menos de 120gr? c) más de 200 gr?

ii)¿Cuál es el peso:
a) máximo del 10% de las manzanas **más pequeñas** del huerto?
b) mínimo del 20% de las manzanas de **mayor calibre** de este huerto?

iii) Si se cosechan al azar 1200 de estas manzanas ¿cuántas pesarán:
a) entre 150 y 200 gr? b) menos de 100 gr? c) más de 180 gr?

9. Sea $X = N(50\,,\,100)$, encuentre:
a) los valores a y b que limitan el 90% central de las observaciones
b) *c* tal que $P(X < c) = 0,20$ c) *d* tal que $P(X \geq d) = 0,10$

10. Si el 15% de las manzanas de menor tamaño del huerto anterior se destinan a producción de pulpa , el 30% de las manzanas de mayor tamaño va a exportación y el resto a consumo interno ¿cuáles son los rangos de pesos de las manzanas según su destino?

11. Si $X = N(20\,,\,\sigma^2)$, encuentre el valor de σ^2 si $P(X > 28) = 0,0228$

12. En una agroindustria se envasa un producto en tarros cuyo contenido neto tiene distribución normal, con desviación estándar de 5 gr. Si el 2,5% de los tarros tiene un peso mayor de 259,8 gr ¿cuál es el peso promedio de su contenido?

13. En un huerto de kiwi de la VI región se obtuvo que la distribución del diámetro polar, en mm, en la temporada 2001 fue $N(58, 16)$ y en la temporada 2002 fue $N(60, 36)$:
a) en un **mismo gráfico** comparativo muestre la situación de **ambas temporadas (indique puntos importantes en el gráfico)**
b) ¿en cuál de las dos temporadas hubo mayor porcentaje de frutos con diámetro polar **menor** a 54 mm. ? **Justifique** numéricamente.

14. Sea una población $X = N(\mu, 144)$, calcule :
a) $P(X < 3X - 44)$ si $\mu = 10$ b) μ si $P(X > 38) = 0,0668$

Respuestas:

1. 0,0571 ; 0,9686 ; 0,5352 ; 0,1246 ; 0,1762 ; 0,7088 **2.** -0,73 ; 0,66 ; -0,50; 1,645 **3.** En Estadística
4. 0,9332 ; 0,3085 ; 0,1587 ; 0,2029 ; 0,1309 ; 0,6826 ; 0,7734 ; 0,5000 **5.** 0,7 ; 13,9 ; 10,3 ; 8,05
6. 0,0526 ; 0,8185 ; 0,1922 **7.** 8 alumnos ; 437 alumnos ; 11 alumnos
8. 61,06% ; 5,48% ; 5,48% ; 128 gr, 181 gr ;721 manz ; 10 manz ; 254 manz **9.** 33,6 y 66,4 ; 41,6 ; 62,8 **11.** 16
12. 250 gr. **14.** 0,1587, 20.

V. DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

1. Reconozca la distribución , los parámetros y especifique la función de cuantía $p(x_i)$ de las siguientes $v.a.\,X$, **justificando cada vez** , y calcule para cada una de ellas $\boldsymbol{E(X)}$ y $\boldsymbol{V(X)}$ **:**
a) X : n° de caras obtenidas al lanzar 15 veces una moneda **no cargada**
b) X: n° de caras obtenidas al lanzar 15 veces una moneda **cargada** tal que la probabilidad de cara es 0,10.
c) X: n° de aciertos a un blanco en una serie de 25 disparos, si la probabilidad de acierto es $3/4$.
d) X: n° de fichas rojas obtenidas al realizar 20 extracciones , **con sustitución**, de una caja que contiene 6 fichas rojas y 4 fichas blancas
e) X: n° de lesiones en hoja de tabaco , por un virus que causa en promedio 3 lesiones por hoja
f) X: n° de bacterias en un cc de un medio de cultivo , cuyo promedio es 24

2. Sea la $v.a.\,X$ n° de caras obtenidas al lanzar 15 veces una moneda **no cargada**. Calcule:
a) $P(X \leq 5)$ b) $P(X \geq 10)$ c) $P(3 \leq X \leq 8)$
d) $P(3 < X \leq 8)$ e) $P(3 \leq X < 8)$ f) $P(3 < X < 8)$

3. Sea la v.a. X : n° de plantas sanas al examinar 25 plantas en un vivero, si la probabilidad de que una planta esté sana es 0,75. Calcule:
a) $P(X = 18)$ b) $P(X \leq 15)$ c) $P(X > 20)$ d) $P(18 \leq X \leq 22)$

4. En una oficina trabajan 50 personas. Se ha establecido que mensualmente existe una probabilidad de 0,10 que a una persona se le hagan descuentos indebidos en sus sueldos. ¿Cuál es la probabilidad que en un mes determinado se tengan:
a) más de 5 personas con descuentos indebidos?
b) sólo 3 personas con descuentos indebidos?
c) a lo más 2 personas con descuentos indebidos?

5. El número de camiones que llega a un centro de acopio es una variable de Poisson. Si el número promedio es de 10 camiones en el día , entre las 7 A.M y las 5 P.M .¿Cuál es la probabilidad de que:
a) en el día lleguen a lo menos 7 camiones ?
b) en el lapso de una hora no lleguen camiones?
c) en un lapso de media hora llegue al menos un camión ?

6. Supongáse que en una planta agroindustrial ocurren de manera aleatoria e independiente 2 accidentes por semana. ¿Cuál es la probabilidad de que:
a) en una semana no ocurran accidentes? b) en una semana ocurran 2 accidentes ?
c) que en una semana ocurra 1 accidente y en la siguiente ocurran 3 ?
d) en dos semanas sucesivas ocurran accidentes ?

7. Estudios sobre un nuevo insecticida biológico ha establecido que la probabilidad de sobrevivencia de la mosca de la fruta a este insecticida es de 0,0002, Se aplica este insecticida en un huerto con una población de 1000 moscas de la fruta. ¿Cuál es la probabilidad de que:
a) no sobreviva ninguna mosca ? b) sobreviva al menos una mosca ?
c) sobrevivan 2 moscas ?

Respuestas:

1. Bin (15 ; 0,5) ; **Bin** (15 ; 0,9) ; **Bin** (25 ; 0,75) ; **Bin** (20 ; 0,6) ; ***P*** (3) ; ***P*** (24)
2. 0,15088 ; 0,15088 ; 0,69269 ; 0,6788 ; 0,49631 ; 0,48242 **3.** 0,16541 ; 0,14944 ; 0,21374 ; 0,6944
4. 0,38388 ; 0,13856 ; 0,11173 **5.** 0,86986 ; 0,36788 ; 0,39347 **6.** 0,13534 ; 0,27067 ; 0,04884 ; 0,74764
7. 0,81873 ; 0,18127 ; 0,01637

VI. DISTRIBUCIONES EN MUESTRAS ALEATORIAS Y ESTIMACIONES

1. La distribución de los pesos de una población de adultos hombres es $X = N(64\,,\,36)$. Se eligen aleatoriamente muestras de 9 individuos de tal población. Calcule:
a) $P(61 \leq \overline{X} \leq 66)$ b) $P(\overline{X} \leq 59)$ c) $P(\overline{X} \geq 67,5)$

2. En un proceso de selección los estudiantes obtuvieron un promedio de 570 puntos con una desviación típica de 40. Se sabe que los puntajes se distribuyen normales. Se toma una muestra aleatoria de 25 estudiantes ¿cuál es la probabilidad de obtener un puntaje promedio de la muestra:
a) entre 560 y 585 ptos. b) menor a 565 ptos. c) mayor a 590 ptos.

3. La distribución de las alturas de personas de una población A es normal con media 173 cm y desviación típica 12 cm. y en otra población B es normal de media 175 cm y una desviación típica de 8 cm. De ambas poblaciones se toma una muestra aleatoria de 16 personas ¿en cuál de las dos muestras hay una mayor probabilidad de obtener una altura promedio mayor a 180 cm.? **Justifique.**

4. Se sabe que en un criadero el peso , en kg , de un cerdo tiene distribución normal con media 80 y varianza 16. Se toma una m.a.s. de 25 cerdos del criadero.

i) ¿Cuál es la probabilidad de obtener un peso promedio de los 25 cerdos :
a) entre 79,0 y 81,5 kg ? b) mayor que 82,3 kg? c) menor que 78,5 kg?

ii) Calcule el valor de:
a) *a* y *b* equidistante de μ tal que $P(a \leq \overline{X} \leq b) = 0,95$
b) *c* tal que $P(\overline{X} \leq c) = 0,15$ c) *d* tal que $P(\overline{X} \geq d) = 0,05$

5. Sea X el rendimiento por hectárea de una variedad de trigo cuya distribución se sabe que es $X = N(72\,,36)$. Se siembran con la variedad 9 parcelas de 5 x 10 m , distribuidas al azar, en un sector A de un fundo y en otro sector B , también distribuidas al azar, se siembran 16 parcelas iguales a las anteriores , y se les mide el rendimiento , proyectado a la ha. (es decir el rendimiento de la parcela amplificado por 200. ¿por qué?).
En un mismo gráfico muestre comparativamente la distribución de la media $\overline{X}$ del sector A respecto a la del sector B.

6. Si se toman **dos** muestras aleatorias de una **misma** población $X = N(\mu\,,\sigma^2)$, la primera de tamaño 10 y la segunda de tamaño 20, obteniéndose los promedios $\overline{X}_1$ = 34 y $\overline{X}_2$ = 37 ¿cuál promedio $\overline{X}_1$ o $\overline{X}_2$ es estimador de la media poblacional ? ¿ Cuál de los dos es ***mejor*** estimador de μ, es decir , tiene la menor varianza ?

7. La distribución de los pesos de recién nacidos en un hospital A es normal con media de 2260 gr y desviación típica 200 gr.. En otro hospital B la distribución de los pesos es también normal con media 2300 gr y desviación típica 120 gr.
De ambos hospitales se toma una m.a.s. de 16 recién nacidos.¿En cuál de las dos muestras hay mayor probabilidad de obtener un peso promedio de recién nacidos mayor a 2370 gr? **Justifique estadísticamente.**

8. Con el fin de generar una referencia para detectar malformaciones craneanas en guaguas de 12 meses de edad de se midió la variable X: perímetro craneano de guaguas normales a los 12 meses de edad y se asume que X distribuye normal. La medición en 15 guaguas dió los siguientes valores en centímetros:
45; 47; 48; 46; 42; 49; 44; 47; 50; 46; 43; 48; 45; 49; 44
a) estime en forma puntual la media y la varianza poblacional del perímetro craneano
b) obtenga estimaciones para μ mediante intervalos del 90% ; 95% y 99% respectivamente, si se conoce que $X = N(\mu\,,\,0,25)$
c) obtenga estimaciones para μ mediante intervalos del 90% ; 95% y 99% respectivamente, si en este caso **no se conoce** el valor de la varianza poblacional $\boldsymbol{\sigma^2}$
d) compare los intervalos del 90%, 95% y 99% obtenidos en b) y c) y obtenga conclusiones respecto a la precisión y confianza.

9. Se sabe que los pesos de cerdos de una población tiene distribución normal de media μ y varianza σ^2.

i) Se elige una m.a.s. de 9 cerdos :
a) determine μ si se sabe que σ^2 = 36 y que P($\overline{X} \leq 92$) = 0,0668
b) determine σ si se sabe que μ = 91 y que P($\overline{X} > 92$) = 0,2266

ii) Si se sabe que $X = N(95\,,\,36)$
a) calcule P($\overline{X} \geq 97$) , para la muestra tamaño 9.
b)¿cuál debe ser el tamaño de la muestra para que se cumpla que P($\overline{X} \geq 97$) $< 0{,}05$?

10. Una máquina envasadora de pulpa de manzana está ajustada para que envase en promedio 240 gr con una desviación estándar de 5 gr. Periodicamente se seleccionan 16 tarros al azar para verificar si la máquina está funcionando correctamente. La máquina se somete a ajustes si el promedio de la muestra resulta inferior a 237 gr. ¿ Si la máquina está envasando correctamente, cuál es la probabilidad que sea sometida a ajuste erróneamente?

11. Se sabe que el rendimiento (en qq/ha) de una nueva variedad de trigo tiene distribución $X = N(\mu, 144)$.
a) Se toma una m.a.s. tamaño 16 de X. Calcule $P(\overline{X} - \mu > 3)$.
b) ¿Cuál debería ser el tamaño de la m.a.s para que con una probabilidad del 95% la media muestral difiera del rendimiento promedio real en menos de 4 qq/ha ?

12. Sea la v.a. *t* con distribución *t de Student* con los *gl* indicados. Determine:
a) *a* tal que P($t < a$) $= 0{,}05$, t con 14 *g.l* b) *b* tal que P($t > b$) $= 0{,}005$, t con 20 *g.l*
c) *c* tal que P($t < c$) $= 0{,}025$, t con 8 *g.l* d) t tal que P(-t $\leq t \leq$ t) $= 0{,}95$, t con 18 *g.l*

13. Se sabe que el estadígrafo ***s*** tiene distribución ***t de Student*** con 15 *g.l* . Calcule:
a) P($s < 0{,}6912$) b) P($s >$ 2,6025) c) P($-1{,}3406 \leq s \leq 2{,}1315$)

14. Sea la v.a. ***D*** con distribución *chi cuadrado* (χ^2) , con los *g.l* indicados .Determine:
a) *a* tal que P($D < a$) $= 0{,}05$, *D* con 12 *g.l* b) *b* tal que P($D > b$) $= 0{,}005$, *D* con 23 *g.l*
c) *c* tal que P($D < c$) $= 0{,}025$, *D* con 9 *g.l*
d) *d* y *e* tal que P($d \leq D \leq e$) $= 0{,}95$ central, *D* con 15 *g.l*

15. Se sabe que el estadígrafo ***D*** tiene distribución χ^2 con 10 *g.l* . Calcule:
a) P($D <$ 3,247) b) P($D >$ 6,737) c) P(4,865 $\leq D \leq$ 18,307)

16. El rendimiento X de una variedad de maíz se conoce que tiene distribución $X = N(\mu\,,\,\sigma^2)$.
Con el fin de estimar μ se siembran 10 parcelas con la variedad de maíz , obteniéndose los siguientes rendimientos a la cosecha:

48 , 50 , 62 , 36 , 45 , 70 , 56 , 40 , 52 , 44

a) obtenga un rango del 95 % de confianza para el verdadero valor de la media b) obtenga una estimación para σ con una confianza del 90 %.

17. Un investigador desea estimar el contenido de Ca en frutos de nectarines , para lo cual selecciona aleatoriamente una muestra de estos obteniendo los siguientes valores:

10 ; 8,9 ; 9,7 ; 10,8 ; 11,0 ; 10,9 ; 9,5 ; 10,7 ; 8,3 ; 9,0 .

a) construya un intervalo del 95% de confianza para la media del contenido de Ca en los nectarines
b) construya un intervalo del 95% de confianza para la varianza del contenido de Ca en los nectarines

18. Se sabe que los aumentos en peso de corderos durante un periodo de 25 días tiene distribución $N(\mu\,,\,\sigma^2)$. Una muestra aleatoria de corderos tuvo las siguientes ganancias de peso a los 25 días:

9 ; 11 ; 12 ; 14 ; 15 ; 16 ; 19 ; 21 ; 24 ; 29 ; 17 ; 20

a) construya un intervalo de confianza del 95% para la varianza de las ganancias de pesos
b) ¿basado en los resultados de la muestra, con una confianza del 95%, puede establecerse que la **población de corderos** gana en promedio 20 kg a los 25 días?

19. La variable aleatoria X representa el peso (en kg) de pollos broiler en un criadero, cuya distribución está dada por:

$$f(x) = \begin{cases} \frac{3}{10}(3x - x^2) \text{ si } 1 \leq x \leq 3 \\ 0 \qquad p.o.v \end{cases} ,$$

Si de la población de pollos del problema anterior se toman muestras aleatorias tamaño 4 y se calcula el peso promedio de los cuatro pollos ¿cuál es la media y la varianza de **esta media muestral ?**

20. Se sabe que la cantidad residual de hormonas X en pollos Broiler tiene distribución normal con media 20 ppm y desviación típica de 4 ppm.
a) ¿ Cuál es la probabilidad de que un pollo cualquiera contenga más de 25 ppm de hormonas?
b) ¿ Cuál es la cantidad *máxima* residual de hormonas del 20% de pollos que contienen menos?
c) ¿ Cuál es la probabilidad de que en una **muestra aleatoria** de 9 pollos se obtenga una **media** de entre 19 ppm y 21 ppm de hormonas?
d) ¿ Cuál debería ser el **nuevo** tamaño de la muestra si se necesita una probabilidad de a lo más 5% de que la media obtenida sea menor a 19 ppm?

Respuestas:
4. i) 0,8643 ; 0,0020 ; 0,0301 **ii)** a = 78,43 , b = 81,57 ; c = 79,17 ; d = 81,31
9. i) 95 ; 4 **ii)** 0,1587 ; $n > 24$ **10.** 0,0082 **11.** 0,1587 ; $n \geq 35$ **12.** -1,7613 ; 2,8453 ; -2,306 ; 2,1009
13. 0,75 ; 0,01 ; 0,875 **14.** 5,226 ; 44,181 ; 2,700 ; 6,262 ; 27,488 **15.** 0,025 ; 0,75 ; 0,85 **18.** 9/5, 3/50

VII. PRUEBAS DE HIPOTESIS PARA LA MEDIA DE DISTRIBUCIONES NORMALES

1. Explique en que consiste, ayudándose con un gráfico, los errores tipo I y II de una prueba de hipótesis.

2. El contenido de proteínas de un alimento para ganado debe ser de a lo menos 200 g. por kg. Ante la sospecha de que la máquina dosificadora no está funcionando adecuadamente se lleva a cabo una inspección. En relación al planteamiento anterior, explique como el inspector puede cometer:
a) un error tipo I y cómo es posible controlar este error
b) un error tipo II y cómo es posible controlar este error.

3. ¿Cuál es la relación entre el nivel de significancia de una prueba y el error de tipo I ?

4. Formule las hipótesis nula y alternativa para probar la tesis médica que ***tomar más de 2 tazas de café al día aumenta el riesgo de cáncer gástrico***. Discuta en términos de las probabilidades de errores tipo I y tipo II con cuál de las posibles hipótesis alternativas se corre mayor riesgo respecto a la salud de los bebedores de café, si el valor de β es bastante mayor que α.

5. Un alimento para ganado debe contener 200 g de proteina en promedio, con una desviación típica de 24 g por kg. Ante la sospecha que la máquina esté dosificando **menos** del promedio es necesario realizar una inspección para lo cual se seleccionan 16 envases de 5 kg y a cada uno se les mide la cantidad de proteina por kg.

Al nivel del 5% **calcule** la probabilidad de que el inspector cometa el error tipo II , si la máquina está envasando un promedio de 185 g por kg.

6. En cada uno los siguientes casos establezca la distribución a utilizar, la Región Crítica , efectúe la prueba de hipótesis y obtenga conclusiones, si el supuesto es que la población tiene distribución $X = N(\mu\,,\,\sigma^2)$:

a) $H_0 : \mu = 27$ vs $H_1 : \mu \neq 27$; $\overline{X} = 30$, S = 4 , n = 25
b) $H_0 : \mu = 98,6$ vs $H_1 : \mu > 98,6$; $\overline{X} = 99,1$, $\sigma =$ 1,5 , n = 30
c) $H_0 : \mu = 3,5$ vs $H_1 : \mu < 3,5$; $\overline{X} = 2,8$, S = $0,6$, n = 18
d) $H_0 : \mu = 382$ vs $H_1 : \mu \neq 382$; $\overline{X} = 358$, $\sigma = 58$, n = 12
e) $H_0 : \mu = 57$ vs $H_1 : \mu > 57$; $\overline{X} = 61$, S = 12 , n = 36

7. Formule las hipótesis nula y alternativa para probar:
a) si un nuevo sistema de embalaje reduce el tiempo de este proceso, que actualmente es de 12,5 minutos, en al menos 2 minutos.
b) que una nueva tecnología de fabricación, produce ampolletas cuya duración promedio es por lo menos 6000 horas mayor que las tradicionales.

8. Para una población $X = N(\mu,\, 16)$ se necesita probar las hipótesis simples $H_0 : \mu = 20$ vs H_1: $\mu = 18$.
a) ¿ cuál es el valor del error tipo II para un un error tipo I de un 5%, si se seleccionó una muestra tamaño 25 para probar las hipótesis anteriores ?
b) en una figura muestre las distribuciones de las variables asociadas a la situación planteada, indique correctamente , con un decimal si es necesario, los valores de posición de las distribuciones, el valor K que limita la Región Crítica y **marque claramente** en la figura el error Tipo I y II.

9. Sea $X = N(\mu,\, 16)$ y las siguientes hipótesis $H_0 : \mu = 70$ vs $H_1 : \mu = 68$.
a) Se toma una muestra aleatoria de X, cuyos valores resultan ser: 73, 62, 75, 64, 72, 67, 74, 65.
¿Qué conclusión se obtiene con la $R.C = \{\overline{X}/\overline{X} < 68,5\}$?
b) Identifique y marque claramente en un gráfico los dos tipos de errores posibles de cometerse con $R.C = \{\overline{X}/\overline{X} < 68,5\}$
c) ¿cuál sería el tamaño de muestra mínimo y el valor de β para α = 0,05, si la Región Crítica es $R.C = \{\overline{X}/\overline{X} < 68,5\}$?

10. Asúmase que la residualidad (persistencia) de un insecticida tiene distribución normal con desviación típica σ = 2,5. Se sabe que el insecticida en uso tiene una residualidad media de 30 días. Otro laboratorio promueve otro insecticida con las mismas características, pero dicen que tiene una mayor residualidad . En un ensayo con el objetivo de verificar tal afirmación , una m.a.s. tamaño 12 dio como resultado un promedio de 32 días como duración del efecto del insecticida. ¿Puede establecerse , al nivel del 5% , que el nuevo insecticida tiene un efecto residual de mayor duración?

11. En el envasado de concentrado de tomate una máquina funcionando correctamente debe envasar en promedio 245 gr. , con una desviación típica de 6 gr por tarro. Un técnico con el fin de verificar si la máquina está funcionando correctamente toma una muestra aleatoria de tarros de la línea de envasado y mide su contenido. Los valores que obtuvo fueron: 232 ; 235 ; 249 ; 241 ; 233 ; 247 ; 244 ; 246 ; 241 ; 248 ; 245 ; 243

a) ¿los resultados de la muestra anterior **son suficiente**, al nivel del 5%, para que se detenga el funcionamiento de la máquina y sea ésta sometida a reparaciones, si se considera **más grave** detener erróneamente el funcionamiento de la máquina ?

b) ¿Cuál debería ser el tamaño de muestra mínimo necesario, para un nivel de significación del 5% y un error tipo II del 15%, para una hipótesis alternativa simple $\mu = 242$ gr?

12. Supóngase que una planta procesadora de alimentos establece que el nivel residual de insectida que estos contengan al llegar a la industria no debe superar los 5 ppm. Una partida de tomates es inspeccionada para ver si cumple la norma , tomándose una muestra al azar de 8 tomates , obteniéndose la siguiente información :

$\sum X_i = 37{,}6$ $\sum X_i^2 = 178$ X_i : contenido insecticida tomate "i"

¿ Los resultados de la muestra permiten concluir que la partida no cumple la norma, al nivel del 5%, si se considera más grave perjudicar al productor?

13. El gerente de producción de una exportadora frutícola desea saber si una nueva línea de embalaje reduce los tiempos actuales , que en promedio es de 14 minutos. El gerente deicde comprar la nueva línea si esta reduce los tiempos en al menos un 15% respecto a la línea actualmente en uso. Para decidir la compra solicita los tiempos logrados en 20 procesos de embalaje con la nueva línea. Los datos obtenidos y enviados al gerente son:
9,8 , 10,4 , 10,6 , 9,6 , 9,7 , 9,9 , 10,9 , 11,1 , 9,6 , 10,2 , 10,3 ,9,6 , 9,9 , 11,2 , 10,6 , 9,8 , 10,5 , 10,1 , 10,5 , 9,7. ¿Con los datos obtenidos, cuál es la decisión que debe tomar el gerente?

14. Se sostiene que con una nueva dieta para cerdos, cuyo objetivo es disminuir la grasa en cerdos, la cantidad promedio por kg de carne es a lo más de 100 gr. Se decide realizar un ensayo en el cual se **alimentarán** 10 cerdos con la **nueva dieta.**

a) ¿ Cuál es la variable asociada al problema y el parámetro de interés ?

b) Especifique las hipótesis, justifique la hipótesis H_1 planteada, **plantee correctamente** el estígrafo de prueba con su distribución y la región crítica correspondiente.

c) Una vez terminado el proceso de engorda, se faenan los cerdos obteniéndose los siguientes contenidos de grasa (en gr) por cada kg : 98 , 90 , 96 , 105 , 97 , 89 , 107 , 93 , 95 , 102.

¿ es posible establecer que con la nueva dieta se logra reducir la cantidad de grasa en cerdos, al nivel del 5% ?

d) ¿qué error es susceptible de estarse cometiendo en la decisión tomada en c) y cuál es su magnitud?

e) construya un intervalo de confianza del 95% para el promedio de grasa por kg con la nueva dieta

f) construya un intervalo de confianza del 95 % para la desviación típica del contenido de grasa.

15. Para controlar arañita roja en paltos se utiliza un acaricida el cual debe aplicarse solamente cuando el promedio de arañitas por hoja supera a 3,0. Con el fin de tomar una decisión de si es el momento de aplicar, un Agrónomo se propone realizar una Prueba de Hipótesis.

Por registros históricos se sabe que la desviación típica de arañitas por hoja es 0,64.

a) Explique cuál es la población en estudio en este problema, la variable asociada y el(los) parámetro(s) de interés ?

b) Especifique las hipótesis a plantear, justificando la hipótesis H_1 a probar, especifique **correctamente** el estadígrafo de prueba con su distribución y la región crítica correspondiente.

c) Para efecto del fin anterior el Agrónomo toma hojas de 10 árboles seleccionados al azar obteniendo los siguientes valores por hoja en cada árbol:

2,5 ; 3,9 ; 2,9 ; 3,9 ; 4,1 ; 4,0 ; 2,7 ; 4,2 ; 2,6 ; 2,8

¿de acuerdo a la información obtenida en la muestra, qué decisión debe tomar el Agrónomo?

d) explique y justifique en cual de los errores es posible estar incurriendo en la decisión obtenida por el Agrónomo.

16. Para satisfacer los requirimientos de exportación de uva de mesa la cantidad residual de sulfuroso **no** debe exceder el valor 0,69 en promedio.

Se afirma que un nuevo tipo de generador de sulfuroso para cajas de exportación permite satisfacer este requerimiento. Se aplica el generador a 10 cajas de uva de exportación y al final del periodo de almacenamiento se les mide la cantidad residual de sulfuroso , obteniéndose los siguientes valores: 0,8 ; 0,5 ; 0,8 ; 0,4 ; 0,6 ; 0,4 ; 0,7 ; 0,5 ; 0,4 ; 0,7

¿Qué conclusión es posible obtener respecto a si el nuevo generador satisface los requerimientos de exportación?

17. Un fabricante de cigarrillos sostiene que el contenido promedio de nicotina de los cigarrillos marca VC **no excede** los 2,5 mg., con una desviación estándar de 0,6 mg

Si una muestra aleatoria de 15 cigarrillos de la marca VC dio un promedio de 2,8 mg ¿qué puede concluirse de la aseveración del fabricante, al nivel del 5%, si se debe proteger la salud de las personas?

18. Se cree que una nueva tecnología en crianza de cerdos produce a los 5 meses de edad ejemplares de peso promedio mayor a 85 kg.

Se toma una muestra aleatoria de 8 cerdos de 5 meses producidos según la nueva tecnología , cuyos pesos resultan ser: 88 ; 89 ; 83 ; 86 ; 91 ; 82 ; 92 ; 89

¿Es posible concluir con los datos de la muestra , al nivel del 5 %, que con la nueva tecnología se obtienen cerdos de 5 meses con peso promedio mayor a 85 kg ?

19. Se desea evaluar un programa de capacitación en raleo de ciruelos a temporeros de la VI región. Para tal efecto se seleccionaron aleatoriamente 12 temporeros, a los cuales se les registró el tiempo empleado en el raleo antes y después de la capacitación. Los tiempos obtenidos se indican en la siguiente tabla:

Temporero	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Antes	**6,2**	**7,0**	**7,5**	**8,0**	**6,3**	**7,4**	**6,5**	**6,8**	**6,9**	**7,6**	**7,2**	**6,4**
Después	**6,0**	**7,2**	**7,0**	**7,6**	**5,9**	**6,9**	**6,5**	**6,4**	**6,7**	**7,1**	**7,2**	**6,2**

¿Qué conclusión se obtiene en relación a la efectividad del programa de capacitación?

20. Para evaluar el efecto de un nuevo método de procesamiento para arreglo de racimo durante el embalaje, se somete a la labor a un grupo de 10 mujeres y posteriormente se las entrena en el nuevo método. Al final se las evalúa nuevamente en la labor de arreglo de racimo. Los resultados obtenidos (en escala de 0 - 100) antes y después del entrenamiento son:

Operaria	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Nota antes	**40**	**65**	**30**	**57**	**60**	**70**	**25**	**45**	**38**	**65**
Nota después	**50**	**70**	**45**	**65**	**64**	**67**	**40**	**50**	**60**	**66**

a) ¿Puede afirmarse que el nuevo método fue efectivo en mejorar la labor de arreglo de racimo?
b) ¿Es posible afirmar que con el nuevo método se incrementa el resultado en más de 5 puntos en promedio?

Asuma que los puntajes obtenidos distribuyen Normal, y concluya con un nivel de significación del 5%.

21. Un jefe de producción desea comparar los porcentajes de descarte en uva de mesa en dos turnos A y B. Para tal efecto selecciona una muestra de descarte, en diez oportunidades al azar, en ambos turnos. Los datos obtenidos son los siguientes:

Turno A: 5,1 1,4 1,6 5,7 9,7 9,1 11,2 8,2 8,9 5,8
Turno B: 2,8 7,3 9,8 7,0 9,5 5,5 5,6 4,7 10,8 6,5

a) ¿Cuál es la conclusión basado en la muestra obtenida?
b) Construya un intervalo de confianza, del 95%, para la diferencia de medias de descarte entre el turno A y B

22. Se prueba un nuevo tipo de fertilizante S en frejol, con el fin de probar si S mejora el rendimiento respecto al fertilizante tradicional T.
a) Indique las poblaciones en estudio, interprete claramente el parámetro a probar y establezca hipótesis , nivel de significación, variable pivotal a utilizar con su **distribución** y región crítica con su **gráfico.**
b) Se siembran y fertlizan 12 parcelas con S y 10 parcelas con T. Realizada a la cosecha se obtuvo la siguiente información de los rendimientos en kg:

Fertilizante S : $\sum^{12} X_i = 398$; $\sum^{12} X_i^2 = 13322$

Fertlizante T : $\sum^{10} X_i = 300$; $\sum^{10} X_i^2 =$ 9188

¿qué puede concluirse del fertilizante S respecto al T , al nivel del 5% ?
c) ¿cuáles son las condiciones (supuestos) necesarias para la validez del desarrollo realizado en la pregunta b)?
d) Construya un intervalo del 95% de confianza para la diferencia de rendimiento entre ambos fertilizantes
e) Realice una Prueba de Hipótesis para verificar el supuesto sobre las varianzas.

23. Para probar si una dieta B produce mayor ganancia de peso en terneros, en kg, respecto a otro tipo de dieta A se alimentan 15 terneros con la dieta A y otros 15 terneros con la dieta B, seleccionados al azar,.Durante el tiempo del ensayo se enfermaron 5 terneros de la dieta A, los que tuvieron que eliminarse del ensayo.
a) Explique claramente cuales son las poblaciones en estudio, las variables y los parámetros a probar.

b) establezca hipótesis, especificándolas con precisión, nivel de significación, variable pivotal a utilizar con su **distribución** y región crítica con su **gráfico.**
c) Del procesamiento de los datos resultó la siguiente información semi procesada:

Dieta A: $\sum_{i=1}^{10} X_i = 700$; $\sum_{i=1}^{10} X_i^2 = 49227$

X_i: ganancia de peso ternero i

Dieta B: $\sum_{i=1}^{15} X_i = 1110$; $\sum_{i=1}^{15} X_i^2 = 82803$

¿Cuál es la conclusión respecto al efecto comparativo de ambas dietas?
d) ¿Qué error es posible haber cometido en la decisión tomada en c) ? **Explique.**
e) ¿Qué supuestos son necesarios para el desarrollo de la pregunta c) ? **Explíquelos**.
f) ¿Cuál será una estimación de la verdadera ganancia de peso obtenida con la dieta B, en un rango del 95% ?

24. Para probar si una hormona CP induce mayor crecimiento de bayas en uva sultanina que la hormona AG , se aplica cada hormona a 15 parras cada una. Los resultados del largo de bayas por parra son los siguientes:

Hormona AG :	$\overline{X}_1$ = 22,0	S_1^2 = 20
Hormona CP :	$\overline{X}_2$ = 23,9	S_2^2 = 32

a) ¿Puede concluirse , al nivel del 5 % , que con la hormona CP se logra mayor largo de bayas en uva sultanina que con AG ?
b) ¿Puede establecerse estadísticamente que las varianzas σ_1^2 y σ_2^2 son distintas ?
c) ¿Qué error α ó β es susceptible de estarse cometiendo en la conclusión obtenida en a) y en la obtenida en b) ?

25. Para comparar el efecto de dos dietas en la cantidad de materia grasa en la leche de vacas lecheras , se alimentan 15 vacas con la dieta A y 18 vacas con la dieta B. Los siguientes son los resultados obtenidos:

Dieta A :	$\overline{X}_A$ = 22,0	S_A = 2,8
Dieta B :	$\overline{X}_B$ = 23,9	S_B = 3,2

a) ¿Es posible establecer , al nivel del 5 % , que con la dieta A se obtiene menor contenido de grasa en la leche que con la dieta B?
b) ¿ Son homogéneas las varianzas σ_A^2 y σ_B^2 ? Plantee hipótesis y docímelas.
c) Construya un intervalo de confianza del 95 % para el contenido de materia grasa en la leche obtenido con la dieta B. Interprete conceptualmente el intervalo obtenido.

26. Para determinar si un nuevo suero es eficaz para prolongar la sobrevivencia por leucemia en ratas, se seleccionan 20 ratas que han contraído la enfermedad y están en una etapa avanzada de ella , de las cuales 12 reciben el suero. Los tiempos de supervivencia, en meses, desde que comenzó el tratamiento dio los siguientes resultados:

Con tratamiento:	$\sum X_i$ = 42	$\sum X_i^2$ = 157,78
Sin tratamiento :	$\sum X_i$ = 19,2	$\sum X_i^2$ = 52,10

i) ¿Puede concluirse, al nivel α, que el suero es eficaz para aumentar la sobrevivencia en ratas:
a) para α = 0,05 ? b) para α = 0,01 ?
ii) ¿Con cuál de los dos niveles de significación concluiría Ud. y por qué?

Resp. 8. 0,1949 **9. a)** aceptar H_0 **c)** n =20, β = 0,2977 **11. b)** n $\geq$ 29

VIII. INTERVALO DE CONFIANZA Y PRUEBAS DE HIPOTESIS PARA PROPORCIONES

1. Para estimar la proporción de pequeños agricultores que cuentan con riego tecnificado se toma una muestra aleatoria de 150 pequeños agricultores verificándose que 38 tienen este tipo de riego.

Construya un intervalo del 95 % de confianza para la proporción de pequeños agricultores con riego tecnificado.

2. Se desea tener una estimación, mediante un intervalo del 95 % de confianza, de la proporción de enraizamiento de rosas multiplicadas mediante estacas tratadas con una hormona W , para lo cual se tratan 120 estacas con la hormona W y se plantan. Al cabo de 6 meses se verifica que 36 estacas **no** echaron raíces.

a) ¿Entre qué valores se encuentra el % de estacas enraizadas? ¿Cuál es el valor del error de muestreo con este tamaño de muestra?

b) ¿cuál deberá ser el tamaño de muestra para disminuir el error de muestreo a 6%?

3. Un municipio determina iniciar una drástica campaña antirrábica si comprueba que la problación de perros vagos que presentan la enfermedad supera el 5 %. Una m.a.s. de 190 perros mostró que 13 presentaban la enfermedad. ¿Qué decisión respecto a la campaña debe tomar la municipalidad con base en la muestra obtenida, a un nivel del 5 %?

4. Se piensa que a lo más el 8 % de los cerdos de un criadero tiene triquina. En una muestra de 60 cerdos se detectan 4 que tienen triquina.

a) ¿El tamaño de la muestra es suficiente para utilizar la aproximación normal?

b) ¿Cuál es la conclusión obtenida, al nivel del 5 % , basado en la información muestral?

c) ¿Qué tipo de error es susceptible de haberse cometido en la conclusión anterior?

d) ¿Entre qué valores está el verdadero porcentaje de cerdos del criadero que tienen triquina, a un nivel de confianza del 5%?

e) Si se desea estimar la proporción de cerdos del criadero que tienen triquina, con un nivel de confianza del 95% y un error no superior a un 4 %, ¿Cuántos cerdos habría que examinar?.

5. Un laboratorio afirma que una hormona T, producida por ellos, aplicada a estacas de rosa induce un enraizamiento de éstas superior al 75%.

a) Especifique con precisión la **población** a investigar, la variable asociada al problema y el parámetro de interés?

b) Especifique las hipótesis y justifique su hipótesis H_1.

c) Para verificar tal aseveración se aplica la hormona T a 120 estacas de rosa de las que posteriormente se determina que enraizan 95 ¿qué puede concluirse respecto a la afirmación del laboratorio, a un nivel del 5% ?

d) En un rango del 95% establezca la verdadera proporción de enraizamiento lograda con la hormona T.

6. Un productor de semillas certificada asegura que al menos el 90% de sus semillas germinan. Para probar tal afirmación se siembran 120 semillas, de las cuales al cabo de unos días 98 germinan.

¿Con este resultado que conclusión debe obtenerse, al nivel del 5 %, respecto a la afirmación del productor?

7. Una empresa agroindustrial está interesada en lanzar un nuevo producto al mercado si al menos un 45% de las personas que concurren a supermercados del sector socio-económico

ABC1 aprueban el producto. Se consulta a 50 personas en cada uno de cuatro supermercados que cumplen con la condición, resultando que 102 personas en total aprueban el producto. ¿Cuál es la decisión que deberá tomar la industria respecto al producto?

8. Para probar si el fungicida A es mejor que el fungicida B en el control de Botritis en peras Winter Nellis , se aplica cada fungicida independientemente a 150 peras previamente inoculadas con el hongo. De las 150 peras tratadas con A presentaron posteriormente pudrición 21, mientras que de las tratadas con B presentaron pudrición 33.
a) ¿Puede concluirse , al nivel del 5 %, que el fungicida A controla mejor Botritis en peras que el fungicida B ?
b) ¿ Entre qué valores está la proporción de peras **sanas** tratadas con el fungicida A ? Dé un rango del 95 % de confianza

9. Un laboratorio afirma tener un nuevo producto WY menos tóxico y más efectivo que el producto BM en el control del tizón del peral. Para confirmar o rechazar tal afirmación se aplicó el producto WY y BM a 120 y 80 árboles respectivamente. Al cabo de un tiempo se detectaron 7 árboles enfermos de los tratados con BM y 6 de los tratados con WY.
¿Puede concluirse, al nivel del 5 %, que WY es mejor que BM en el control de la enfermedad?

10. Para estimar la proporción de plantas enfermas en un vivero se toma una muestra de 180 plantas elegidas aleatoriamente entre las cuales se encontraron 32 plantas **enfermas.**
Posteriormente a todas las plantas del vivero se les efectúa un tratamiento con el objeto de sanarlas. Después de algunas semanas, para determinar si hubo mejoría, se toma **otra** muestra de 120 plantas, encontrándose sólo 12 plantas enfermas.
a) explique el(los) parámetro(s) a contrastar y su interpretación
b) plantee la hipótesis alternativa y justifique en palabras su elección
c) ¿al nivel del 5% el tratamiento resultó efectivo para reducir la enfermedad?
d) ¿cuál es la proporción de **plantas sanas** en el vivero **antes** del tratamiento, en un rango del 95% de confianza?

11. Se necesita probar si un producto natural D tiene efecto para curar plantas enfermas en un vivero. Se toma una muestra aleatoria de plantas **antes** de aplicar el producto detectándose en la muestra 30 plantas enfermas y 90 plantas sanas.
a) en un rango del 95% ¿cuál es el porcentaje de plantas **sanas** del vivero?
b) días después de aplicado el producto se toma **otra muestra aleatoria** en el vivero y en el examen de las plantas seleccionadas se determina que hay 36 plantas enfermas y 114 plantas sanas.
¿Qué conclusión se obtiene respecto al efecto del producto para curar las plantas enfermas, al nivel del 5%?

12. Al alimento de gallinas ponedoras se le agrega vitamina C con el fin de probar si ella contribuye a **disminuir** la cantidad de huevos trizados. Para tal efecto a un conjunto de gallinas se les suministra la vitamina con el alimento. Después de varios días de aplicación de la vitamina se seleccionan al azar 150 huevos de gallinas alimentadas **con la vitamina**, encontrándose 6 trizados y otros 150 huevos de gallinas alimentadas **sin la vitamina** , entre los cuales se cuentan 12 huevos trizados.
¿Al nivel del 5 %, es posible concluir que conviene agregar vitamina C al alimento para disminuir la proporción de huevos trizados ?

13. Se desea probar si el acaricida B es mejor que otro acaricida A en el control de la arañita roja. Para este efecto a un árbol se le aplica el producto A , determinándose que en un conjunto de hojas hay 110 arañitas muertas y 40 vivas , mientras que en las hojas de otro árbol donde se aplicó el producto B se encontraron 100 arañitas muertas y 20 vivas.

¿Puede establecerse, al nivel del 5 %, que el producto B controla mejor que el A la arañita roja?

Respuestas.

2. b) n = 225 **4. d)** entre 0,4% y 13,0% **e)** $n \geq 177$

IX. PRUEBAS DE CONCORDANCIA Y DE ASOCIACION

1. Para probar si la proporción de plantas con virus en un vivero corresponde al 10 % , se examinan 75 plantas determinándose que 66 están libres de virus. Plantee hipótesis y obtenga conclusiones mediante la prueba de concordancia , a un nivel del 5 % , y compare esta prueba con la prueba para una proporción vista en la guía anterior.

2. Según la ley de Mendel la segregación fenotípica de dos pares de caracteres debe estar en la proporción 9:3:3:1. Para comprobar experimentalmente el cumplimiento de esta ley se analizaron 800 individuos provenientes de la cruza , encontrándose la siguiente segregación:

Segregación	AB	Ab	aB	ab
n°individuos	445	155	152	48

¿Los resultados experimentales anteriores son concordante con lo establecido por la Ley de Mendel , al nivel del 5 % ?

3. Se piensa que las tres causas A , B y C de muerte al nacer de cerdos están en la proporción 1:3:4. Para verificar la hipótesis anterior se analiza la causa de muerte de 80 cerditos , encontrándose que 14 corresponden a la causa A , 28 a la causa B y el resto a la causa C.

¿Puede establecerse , al nivel del 5 % , que estos resultados contradicen la proporción indicada ?

4. Se desea determinar si existen diferencias entre las preferencias de productores lecheros respecto de 5 marcas de insumos. Una encuesta da las siguientes preferencias para cada una de las marcas:

Marca	M_1	M_2	M_3	M_4	M_5
N°preferencias	28	25	35	39	28

Plantee hipótesis y docímelas , al nivel del 5 %.

5. En el procesamiento agroindustrial de tomates en conserva , el análisis de una muestra de 450 tarros rechazados por defectos da como resultado que fueron rechazados por abolladuras (A) 162 , por mal etiquetado (E) 145 , por oxidación (O) 103 y por sellado (S) 40.

¿Los resultados de esta muestra son concordante con la hipótesis de que las fallas por (A) son 6 veces más frecuentes que por (S) , las fallas por (E) 5 veces más frecuentes que (S) y las fallas por (O) 3 veces más frecuentes que (S) ?

6. Se asevera que en una variedad de frejol el 10 % de las semillas **no germina** , el 30 % produce plantas **anormales** y el resto son **normales**. Se siembran 180 semillas de esta variedad, **germinando** 155 de las cuales 105 resultan ser plantas **normales.**

¿ Qué puede concluirse , al nivel del 5 % , de la aseveración para esta variedad de frejol ?

7. Se vacunan contra cierta enfermedad 120 animales sanos. Después de un tiempo se encuentra que 12 adquirieron la enfermedad. De un examen de 140 animales no vacunados se encuentran que 50 adquirieron la enfermedad.

Plantee hipótesis que permitan establecer si existe asociación entre la vacunación y la incidencia de la enfermedad y docímelas.

8. Se desea establecer si tres mezclas químicas P , Q y R aplicadas a semilla de tomate producen diferencias en la germinación de éstas. Se tratan tres grupos de 200 semillas con cada una de las tres mezclas , determinándose que germinan 190 , 165 y 180 con P , Q y R respectivamente.

¿Qué puede concluirse respecto a la diferencia en la germinación de las semillas de las tres mezclas , al nivel del 5 % ?

9. De una encuesta , 600 productores lecheros fueron clasificados de acuerdo al tamaño de su plantel y a su nivel tecnológico para determinar si hay asociación entre ambas variables categóricas. La clasificación con sus frecuencias la muestra la siguiente tabla de doble entrada:

Tamaño \Nivel tecnol.	bajo	mediano	alto
pequeño	182	85	33
mediano	68	60	72
grande	20	41	39

¿Puede establecerse que la proporción de productores en los niveles tecnológicos es independiente de su tamaño ?

10.. Se desea probar si existe diferencia entre 4 pequeños productores, A,B,C,D de uva sultanina en relación a la calidad de exportación. Para tal efecto se seleccionó una muestra al azar de racimos de cada productor, contabilizándose el número de racimos aceptados para exportación.

La información se muestra en la siguiente tabla:

Condic.\Productor	A	B	C	D
Aceptados	86	230	285	132
Rechazados	14	20	15	18

¿Basado en la información anterior, es posible establecer que la calidad de exportación es diferente entre los productores?

11. Una empresa de marketing desea establecer si la preferencia por tres marcas de cereales (X, Y, Z) está asociado al nivel socioeconómico (A, B, C1 y C2). En una encuesta realizada en supermercados entregó la siguiente información:

Marca\Nivel	A	B	C1	C2
X	25	10	10	5
Y	80	65	45	10
Z	95	90	45	20

¿Cuál es la conclusión obtenida en base a la muestra anterior?

12. Con el fin de determinar el hábito de consumo de palta por grupos de edad se realizó una encuesta que dio los siguientes resultados:

Consumo \ Edad	< 20	20-29	30-60	> 60
bajo	65	66	40	34
medio	42	30	33	42
alto	93	54	27	24

¿Puede establecerse , al nivel del 5 % , que el nivel de consumo de palta está asociada a la edad de las personas?

BIBLIOGRAFIA

1 Berenson, M.L y Levine, D.M. 1996. Estadística básica en Administración: conceptos y aplicaciones. Prentice-Hall. 6ª ed. México.

2. Canavos, G. 1992. Probabilidad y Estadística: Aplicaciones y Métodos. McGraw-Hill.México

3. Chao, L.L. 1993. Estadística para las ciencias administrativas. McGraw-Hill. 3ª ed. México.

4. D'Ottone, H. 1991. Estadística Elemental. Coopecultura Ltda. Santiago, Chile.

5. Levin, R. 2006. Estadística para administradores. Prentice-Hall. México.

6. Levin, R. y Rubin, D. 1996. Estadística para Administración. Prentice-Hall. 6ª ed. México.

7. Meyer. P.L. 1992. Probabilidad y Aplicaciones Estadísticas. Addison- Wesley Iberoamericana. Wilmington, Delaware, E.U.A

8. Ostle, B. 1983. Estadística Aplicada. Limusa Wiley. México.

9. Ross, Sh. 2002. Probabilidad y Estadística para Ingenieros. McGraw-Hill, Interamericana Editores. 2ª ed. México.

10. Royo, A. 1985. Curso de Estadística. Facultad de Ciencias Agrarias, Veterinarias y Forestales. Universidad de Chile.

11. Rustom, A. 1990. Elementos de Probabilidad y su aplicación a la Agronomía. Publicación Docente Nº 1. Dirección Escuela de Agronomía, Facultad de Ciencias Agrarias y Forestales, Universidad de Chile.

12. Snedecor, G.W y Cochran,W. 1977. Métodos Estadísticos. C.E.C.S.A. México.

13. Spiegel, M.R. Teoría y Problemas de Estadística. Libros McGraw-Hill. Serie de Compendios Schaum.

14. Walpole, R.E. y Myers, R.H. 1992. Probabilidad y estadística. McGraw-Hill, 4ª ed. España.

15. Walpole, R.E., Myers, R.H.y Myers, S.L, 1999. Probabilidad y estadística para ingenieros. Prentice-Hall Hispanoamericana.

16. Zuwaylif, F.H. 1971. Estadística General aplicada. Fondo Educativo Interamericano. México.

Anexo 1

Area bajo la curva normal estándar: $\varphi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} Exp(-z^2/2) * dz$

z	Decimales de z									
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,40	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,30	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,20	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,10	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,00	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,90	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,80	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,70	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,60	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,50	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,40	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,30	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,20	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,10	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,00	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,90	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,80	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,70	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,60	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,50	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,40	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,30	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,20	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,10	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,00	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,90	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,80	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,70	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,60	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,50	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,40	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,30	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,20	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,10	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,00	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,00	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,10	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,20	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,30	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,40	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Anexo 2

Función de Distribución Acumulativa Binomial (n , p)		
x	F1(x)	F2(x)
0	0,00000	0,01687
1	0,00001	0,08716
2	0,00009	0,23214
3	0,00050	0,42948
4	0,00206	0,62886
5	0,00673	0,78837
6	0,01822	0,89361
7	0,04202	0,95249
8	0,08441	0,98101
9	0,15036	0,99316
10	0,24104	0,99776
11	0,35232	0,99933
12	0,47520	0,99982
13	0,59808	0,99995
14	0,70998	0,99999
15	0,80324	1,00000
16	0,87464	1,00000
17	0,92503	1,00000
18	0,95793	1,00000
19	0,97785	1,00000
20	0,98905	1,00000
21	0,99491	1,00000
22	0,99778	1,00000
23	0,99909	1,00000
24	0,99965	1,00000
25	0,99987	1,00000
26	0,99996	1,00000
27	0,99999	1,00000
28	1,00000	1,00000

Anexo 3

Función de Distribución Acumulativa de Poisson					
λ =	**1**	**2**	**3**	**4**	**5**
x	**F1(x)**	**F2(x)**	**F3(x)**	**F4(x)**	**F5(x)**
0	0,36788	0,13534	0,04979	0,01832	0,00674
1	0,73576	0,40601	0,19915	0,09158	0,04043
2	0,91970	0,67668	0,42319	0,23810	0,12465
3	0,98101	0,85712	0,64723	0,43347	0,26503
4	0,99634	0,94735	0,81526	0,62884	0,44049
5	0,99941	0,98344	0,91608	0,78513	0,61596
6	0,99992	0,99547	0,96649	0,88933	0,76218
7	0,99999	0,99890	0,98810	0,94887	0,86663
8	1,00000	0,99976	0,99620	0,97864	0,93191
9	1,00000	0,99995	0,99890	0,99187	0,96817
10	1,00000	0,99999	0,99971	0,99716	0,98630
11	1,00000	1,00000	0,99993	0,99908	0,99455
12	1,00000	1,00000	0,99998	0,99973	0,99798
13	1,00000	1,00000	1,00000	0,99992	0,99930
14	1,00000	1,00000	1,00000	0,99998	0,99977
15	1,00000	1,00000	1,00000	1,00000	0,99993
16	1,00000	1,00000	1,00000	1,00000	0,99998
17	1,00000	1,00000	1,00000	1,00000	0,99999
18	1,00000	1,00000	1,00000	1,00000	1,00000

Anexo 4

Percentiles de la distribución ji-cuadrado de Pearson con n grados de libertad

n \ p	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776	9.21034	10.5966
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840	11.3449	12.8382
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.1433	13.2767	14.8603
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.0705	12.8325	15.0863	16.7496
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.6446	12.5916	14.4494	16.8119	18.5476
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.2189	13.3616	15.5073	17.5345	20.0902	21.9550
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.3888	14.6837	16.9190	19.0228	21.6660	23.5894
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.5489	15.9872	18.3070	20.4832	23.2093	25.1882
11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.3410	13.7007	17.2750	19.6751	21.9200	24.7250	26.7568
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.3403	14.8454	18.5493	21.0261	23.3367	26.2170	28.2995
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.3398	15.9839	19.8119	22.3620	24.7356	27.6882	29.8195
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.1653	13.3393	17.1169	21.0641	23.6848	26.1189	29.1412	31.3193
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.0365	14.3389	18.2451	22.3071	24.9958	27.4884	30.5779	32.8013
16	5.14221	5.81221	6.90766	7.96165	9.31224	11.9122	15.3385	19.3689	23.5418	26.2962	28.8454	31.9999	34.2672
17	5.69722	6.40776	7.56419	8.67176	10.0852	12.7919	16.3382	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185
18	6.26480	7.01491	8.23075	9.39046	10.8649	13.6753	17.3379	21.6049	25.9894	28.8693	31.5264	34.8053	37.1565
19	6.84397	7.63273	8.90652	10.1170	11.6509	14.5620	18.3377	22.7178	27.2036	30.1435	32.8523	36.1909	38.5823
20	7.43384	8.26040	9.59078	10.8508	12.4426	15.4518	19.3374	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968
21	8.03365	8.89720	10.2829	11.5913	13.2396	16.3444	20.3372	24.9348	29.6151	32.6706	35.4789	38.9322	41.4011
22	8.64272	9.54249	10.9823	12.3380	14.0415	17.2396	21.3370	26.0393	30.8133	33.9244	36.7807	40.2894	42.7957
23	9.26042	10.1957	11.6886	13.0905	14.8480	18.1373	22.3369	27.1413	32.0069	35.1725	38.0756	41.6384	44.1813
24	9.88623	10.8564	12.4012	13.8484	15.6587	19.0373	23.3367	28.2412	33.1962	36.4150	39.3641	42.9798	45.5585
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366	29.3389	34.3816	37.6525	40.6465	44.3141	46.9279
26	11.1602	12.1981	13.8439	15.3792	17.2919	20.8434	25.3365	30.4346	35.5632	38.8851	41.9232	45.6417	48.2899
27	11.8076	12.8785	14.5734	16.1514	18.1139	21.7494	26.3363	31.5284	36.7412	40.1133	43.1945	46.9629	49.6449
28	12.4613	13.5647	15.3079	16.9279	18.9392	22.6572	27.3362	32.6205	37.9159	41.3371	44.4608	48.2782	50.9934
29	13.1211	14.2565	16.0471	17.7084	19.7677	23.5666	28.3361	33.7109	39.0875	42.5570	45.7223	49.5879	52.3356
30	13.7867	14.9535	16.7908	18.4927	20.5992	24.4776	29.3360	34.7997	40.2560	43.7730	46.9792	50.8922	53.6720
31	14.4578	15.6555	17.5387	19.2806	21.4336	25.3901	30.3359	35.8871	41.4217	44.9853	48.2319	52.1914	55.0027
32	15.1340	16.3622	18.2908	20.0719	22.2706	26.3041	31.3359	36.9730	42.5847	46.1943	49.4804	53.4858	56.3281
33	15.8153	17.0735	19.0467	20.8665	23.1102	27.2194	32.3358	38.0575	43.7452	47.3999	50.7251	54.7755	57.6484
34	16.5013	17.7891	19.8063	21.6643	23.9523	28.1361	33.3357	39.1408	44.9032	48.6024	51.9660	56.0609	58.9639
35	17.1918	18.5089	20.5694	22.4650	24.7967	29.0540	34.3356	40.2228	46.0588	49.8018	53.2033	57.3421	60.2748
36	17.8867	19.2327	21.3359	23.2686	25.6433	29.9730	35.3356	41.3036	47.2122	50.9985	54.4373	58.6192	61.5812
37	18.5858	19.9602	22.1056	24.0749	26.4921	30.8933	36.3355	42.3833	48.3634	52.1923	55.6680	59.8925	62.8833
38	19.2889	20.6914	22.8785	24.8839	27.3430	31.8146	37.3355	43.4619	49.5126	53.3835	56.8955	61.1621	64.1814
39	19.9959	21.4262	23.6543	25.6954	28.1958	32.7369	38.3354	44.5395	50.6598	54.5722	58.1201	62.4281	65.4756
40	20.7065	22.1643	24.4330	26.5093	29.0505	33.6603	39.3353	45.6160	51.8051	55.7585	59.3417	63.6907	66.7660
41	21.4208	22.9056	25.2145	27.3256	29.9071	34.5846	40.3353	46.6916	52.9485	56.9424	60.5606	64.9501	68.0527
42	22.1385	23.6501	25.9987	28.1440	30.7654	35.5099	41.3352	47.7663	54.0902	58.1240	61.7768	66.2062	69.3360
43	22.8595	24.3976	26.7854	28.9647	31.6255	36.4361	42.3352	48.8400	55.2302	59.3035	62.9904	67.4593	70.6159
44	23.5837	25.1480	27.5746	29.7875	32.4871	37.3631	43.3352	49.9129	56.3685	60.4809	64.2015	68.7095	71.8926
45	24.3110	25.9013	28.3662	30.6123	33.3504	38.2910	44.3351	50.9849	57.5053	61.6562	65.4102	69.9568	73.1661
46	25.0413	26.6572	29.1601	31.4390	34.2152	39.2197	45.3351	52.0562	58.6405	62.8296	66.6165	71.2014	74.4365
47	25.7746	27.4158	29.9562	32.2676	35.0814	40.1492	46.3350	53.1267	59.7743	64.0011	67.8206	72.4433	75.7041
48	26.5106	28.1770	30.7545	33.0981	35.9491	41.0794	47.3350	54.1964	60.9066	65.1708	69.0226	73.6826	76.9688
49	27.2493	28.9406	31.5549	33.9303	36.8182	42.0104	48.3350	55.2653	62.0375	66.3386	70.2224	74.9195	78.2307
50	27.9907	29.7067	32.3574	34.7643	37.6886	42.9421	49.3349	56.3336	63.1671	67.5048	71.4202	76.1539	79.4900
60	35.5345	37.4849	40.4817	43.1880	46.4589	52.2938	59.3347	66.9815	74.3970	79.0819	83.2977	88.3794	91.9517
70	43.2752	45.4417	48.7576	51.7393	55.3289	61.6983	69.3345	77.5767	85.5270	90.5312	95.0232	100.425	104.215
80	51.1719	53.5401	57.1532	60.3915	64.2778	71.1445	79.3343	88.1303	96.5782	101.879	106.629	112.329	116.321
90	59.1963	61.7541	65.6466	69.1260	73.2911	80.6247	89.3342	98.6499	107.565	113.145	118.136	124.116	128.299

Anexo 5

Percentiles de la distribución t de Student con n grados de libertad.
En la última fila aparecen los percentiles de la Normal (0, 1) para $n \to \infty$

n \ p	0.75	0.90	0.95	0.975	0.99	0.995
1	1.000000	3.07768	6.31375	12.7062	31.8205	63.6567
2	0.816497	1.88562	2.91999	4.30265	6.96456	9.92484
3	0.764892	1.63774	2.35336	3.18245	4.54070	5.84091
4	0.740697	1.53321	2.13185	2.77645	3.74695	4.60409
5	0.726687	1.47588	2.01505	2.57058	3.36493	4.03214
6	0.717558	1.43976	1.94318	2.44691	3.14267	3.70743
7	0.711142	1.41492	1.89458	2.36462	2.99795	3.49948
8	0.706387	1.39682	1.85955	2.30600	2.89646	3.35539
9	0.702722	1.38303	1.83311	2.26216	2.82144	3.24984
10	0.699812	1.37218	1.81246	2.22814	2.76377	3.16927
11	0.697445	1.36343	1.79588	2.20099	2.71808	3.10581
12	0.695483	1.35622	1.78229	2.17881	2.68100	3.05454
13	0.693829	1.35017	1.77093	2.16037	2.65031	3.01228
14	0.692417	1.34503	1.76131	2.14479	2.62449	2.97684
15	0.691197	1.34061	1.75305	2.13145	2.60248	2.94671
16	0.690132	1.33676	1.74588	2.11991	2.58349	2.92078
17	0.689195	1.33338	1.73961	2.10982	2.56693	2.89823
18	0.688364	1.33039	1.73406	2.10092	2.55238	2.87844
19	0.687621	1.32773	1.72913	2.09302	2.53948	2.86093
20	0.686954	1.32534	1.72472	2.08596	2.52798	2.84534
21	0.686352	1.32319	1.72074	2.07961	2.51765	2.83136
22	0.685805	1.32124	1.71714	2.07387	2.50832	2.81876
23	0.685306	1.31946	1.71387	2.06866	2.49987	2.80734
24	0.684850	1.31784	1.71088	2.06390	2.49216	2.79694
25	0.684430	1.31635	1.70814	2.05954	2.48511	2.78744
26	0.684043	1.31497	1.70562	2.05553	2.47863	2.77871
27	0.683685	1.31370	1.70329	2.05183	2.47266	2.77068
28	0.683353	1.31253	1.70113	2.04841	2.46714	2.76326
29	0.683044	1.31143	1.69913	2.04523	2.46202	2.75639
30	0.682756	1.31042	1.69726	2.04227	2.45726	2.75000
31	0.682486	1.30946	1.69552	2.03951	2.45282	2.74404
32	0.682234	1.30857	1.69389	2.03693	2.44868	2.73848
33	0.681997	1.30774	1.69236	2.03452	2.44479	2.73328
34	0.681774	1.30695	1.69092	2.03224	2.44115	2.72839
35	0.681564	1.30621	1.68957	2.03011	2.43772	2.72381
36	0.681366	1.30551	1.68830	2.02809	2.43449	2.71948
37	0.681178	1.30485	1.68709	2.02619	2.43145	2.71541
38	0.681001	1.30423	1.68595	2.02439	2.42857	2.71156
39	0.680833	1.30364	1.68488	2.02269	2.42584	2.70791
40	0.680673	1.30308	1.68385	2.02108	2.42326	2.70446
41	0.680521	1.30254	1.68288	2.01954	2.42080	2.70118
42	0.680376	1.30204	1.68195	2.01808	2.41847	2.69807
43	0.680238	1.30155	1.68107	2.01669	2.41625	2.69510
44	0.680107	1.30109	1.68023	2.01537	2.41413	2.69228
45	0.679981	1.30065	1.67943	2.01410	2.41212	2.68959
46	0.679861	1.30023	1.67866	2.01290	2.41019	2.68701
47	0.679746	1.29982	1.67793	2.01174	2.40835	2.68456
48	0.679635	1.29944	1.67722	2.01063	2.40658	2.68220
49	0.679530	1.29907	1.67655	2.00958	2.40489	2.67995
50	0.679428	1.29871	1.67591	2.00856	2.40327	2.67779
60	0.678601	1.29582	1.67065	2.00030	2.39012	2.66028
70	0.678011	1.29376	1.66691	1.99444	2.38081	2.64790
80	0.677569	1.29222	1.66412	1.99006	2.37387	2.63869
90	0.677225	1.29103	1.66196	1.98667	2.36850	2.63157
100	0.676951	1.29007	1.66023	1.98397	2.36422	2.62589
200	0.675718	1.28580	1.65251	1.97190	2.34514	2.60063
300	0.675308	1.28438	1.64995	1.96790	2.33884	2.59232
400	0.675104	1.28367	1.64867	1.96591	2.33571	2.58818
500	0.674981	1.28325	1.64791	1.96472	2.33383	2.58570
1000	0.674735	1.28240	1.64638	1.96234	2.33008	2.58075
∞	0.674490	1.28155	1.64485	1.95996	2.32635	2.57583

Anexo 6

Percentiles de la distribución F de Fisher–Snedecor
con m grados de libertad en el numerador y n en el denominador

n	p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.9	39.863	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.858	60.195	60.473	60.705	60.903	61.073	61.220
	0.95	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69	245.36	245.95
	0.975	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	973.03	976.71	979.84	982.53	984.87
	0.99	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5	6055.8	6083.3	6106.3	6125.9	6142.7	6157.3
	0.995	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	24505	24572	24630
2	0.9	8.5263	9.0000	9.1618	9.2434	9.2926	9.3255	9.3491	9.3668	9.3805	9.3916	9.4006	9.4081	9.4145	9.4200	9.4247
	0.95	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.405	19.413	19.419	19.424	19.429
	0.975	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.407	39.415	39.421	39.427	39.431
	0.99	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.408	99.416	99.422	99.428	99.433
	0.995	198.50	199.00	199.17	199.25	199.30	199.33	199.36	199.37	199.39	199.40	199.41	199.42	199.42	199.43	199.43
3	0.9	5.5383	5.4624	5.3908	5.3426	5.3092	5.2847	5.2662	5.2517	5.2400	5.2304	5.2224	5.2156	5.2098	5.2047	5.2003
	0.95	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7633	8.7446	8.7287	8.7149	8.7029
	0.975	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337	14.304	14.277	14.253
	0.99	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.133	27.052	26.983	26.924	26.872
	0.995	55.552	49.799	47.467	46.195	45.392	44.838	44.434	44.126	43.882	43.686	43.524	43.387	43.271	43.172	43.085
4	0.9	4.5448	4.3246	4.1909	4.1072	4.0506	4.0097	3.9790	3.9549	3.9357	3.9199	3.9067	3.8955	3.8859	3.8776	3.8704
	0.95	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9358	5.9117	5.8911	5.8733	5.8578
	0.975	12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7935	8.7512	8.7150	8.6838	8.6565
	0.99	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452	14.374	14.307	14.249	14.198
	0.995	31.333	26.284	24.259	23.155	22.456	21.975	21.622	21.352	21.139	20.967	20.824	20.705	20.603	20.515	20.438
5	0.9	4.0604	3.7797	3.6195	3.5202	3.4530	3.4045	3.3679	3.3393	3.3163	3.2974	3.2816	3.2682	3.2567	3.2468	3.2380
	0.95	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.7040	4.6777	4.6552	4.6358	4.6188
	0.975	10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5678	6.5245	6.4876	6.4556	6.4277
	0.99	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.9626	9.8883	9.8248	9.7700	9.7222
	0.995	22.785	18.314	16.530	15.556	14.940	14.513	14.200	13.961	13.772	13.618	13.491	13.384	13.293	13.215	13.146
6	0.9	3.7759	3.4633	3.2888	3.1808	3.1075	3.0546	3.0145	2.9830	2.9577	2.9369	2.9195	2.9047	2.8920	2.8809	2.8712
	0.95	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	4.0274	3.9999	3.9764	3.9559	3.9381
	0.975	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.4098	5.3662	5.3290	5.2968	5.2687
	0.99	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761	7.8741	7.7896	7.7183	7.6575	7.6049	7.5590
	0.995	18.635	14.544	12.917	12.028	11.464	11.073	10.786	10.566	10.391	10.250	10.133	10.034	9.9501	9.8774	9.8140
7	0.9	3.5894	3.2574	3.0741	2.9605	2.8833	2.8274	2.7849	2.7516	2.7247	2.7025	2.6839	2.6681	2.6545	2.6426	2.6322
	0.95	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.6030	3.5747	3.5503	3.5292	3.5107
	0.975	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.7095	4.6658	4.6285	4.5961	4.5678
	0.99	12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188	6.6201	6.5382	6.4691	6.4100	6.3590	6.3143
	0.995	16.236	12.404	10.882	10.050	9.5221	9.1553	8.8854	8.6781	8.5138	8.3803	8.2697	8.1764	8.0967	8.0279	7.9678
8	0.9	3.4579	3.1131	2.9238	2.8064	2.7264	2.6683	2.6241	2.5893	2.5612	2.5380	2.5186	2.5020	2.4876	2.4752	2.4642
	0.95	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.3130	3.2839	3.2590	3.2374	3.2184
	0.975	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.2434	4.1997	4.1622	4.1297	4.1012
	0.99	11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106	5.8143	5.7343	5.6667	5.6089	5.5589	5.5151
	0.995	14.688	11.042	9.5965	8.8051	8.3018	7.9520	7.6941	7.4959	7.3386	7.2106	7.1045	7.0149	6.9384	6.8721	6.8143
9	0.9	3.3603	3.0065	2.8129	2.6927	2.6106	2.5509	2.5053	2.4694	2.4403	2.4163	2.3961	2.3789	2.3640	2.3510	2.3396
	0.95	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.1025	3.0729	3.0475	3.0255	3.0061
	0.975	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.9121	3.8682	3.8306	3.7980	3.7694
	0.99	10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511	5.2565	5.1779	5.1114	5.0545	5.0052	4.9621
	0.995	13.614	10.107	8.7171	7.9559	7.4712	7.1339	6.8849	6.6933	6.5411	6.4172	6.3142	6.2274	6.1530	6.0887	6.0325
10	0.9	3.2850	2.9245	2.7277	2.6053	2.5216	2.4606	2.4140	2.3772	2.3473	2.3226	2.3018	2.2841	2.2687	2.2553	2.2435
	0.95	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9430	2.9130	2.8872	2.8647	2.8450
	0.975	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6649	3.6209	3.5832	3.5504	3.5217
	0.99	10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424	4.8491	4.7715	4.7059	4.6496	4.6008	4.5581
	0.995	12.826	9.4270	8.0807	7.3428	6.8724	6.5446	6.3025	6.1159	5.9676	5.8467	5.7462	5.6613	5.5887	5.5257	5.4707
11	0.9	3.2252	2.8595	2.6602	2.5362	2.4512	2.3891	2.3416	2.3040	2.2735	2.2482	2.2269	2.2087	2.1930	2.1792	2.1671
	0.95	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.8179	2.7876	2.7614	2.7386	2.7186
	0.975	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4737	3.4296	3.3917	3.3588	3.3299
	0.99	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315	4.5393	4.4624	4.3974	4.3416	4.2932	4.2509
	0.995	12.226	8.9122	7.6004	6.8809	6.4217	6.1016	5.8648	5.6821	5.5368	5.4183	5.3197	5.2363	5.1649	5.1031	5.0489
12	0.9	3.1765	2.8068	2.6055	2.4801	2.3940	2.3310	2.2828	2.2446	2.2135	2.1878	2.1660	2.1474	2.1313	2.1173	2.1049
	0.95	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.7173	2.6866	2.6602	2.6371	2.6169
	0.975	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.3215	3.2773	3.2393	3.2062	3.1772
	0.99	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875	4.2961	4.2198	4.1553	4.0999	4.0518	4.0096
	0.995	11.754	8.5096	7.2258	6.5211	6.0711	5.7570	5.5245	5.3451	5.2021	5.0855	4.9884	4.9062	4.8358	4.7748	4.7213
13	0.9	3.1362	2.7632	2.5603	2.4337	2.3467	2.2830	2.2341	2.1953	2.1638	2.1376	2.1155	2.0966	2.0802	2.0658	2.0532
	0.95	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6347	2.6037	2.5769	2.5536	2.5331
	0.975	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1975	3.1532	3.1150	3.0819	3.0527
	0.99	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911	4.1003	4.0245	3.9603	3.9052	3.8573	3.8154
	0.995	11.374	8.1865	6.9258	6.2335	5.7910	5.4819	5.2529	5.0761	4.9351	4.8199	4.7240	4.6429	4.5733	4.5129	4.4600
14	0.9	3.1022	2.7265	2.5222	2.3947	2.3069	2.2426	2.1931	2.1539	2.1220	2.0954	2.0729	2.0537	2.0370	2.0224	2.0095
	0.95	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5655	2.5342	2.5073	2.4837	2.4630
	0.975	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0946	3.0502	3.0119	2.9786	2.9493
	0.99	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297	3.9394	3.8640	3.8001	3.7452	3.6975	3.6557
	0.995	11.060	7.9216	6.6804	5.9984	5.5623	5.2574	5.0313	4.8566	4.7173	4.6034	4.5085	4.4281	4.3591	4.2993	4.2468
15	0.9	3.0732	2.6952	2.4898	2.3614	2.2730	2.2081	2.1582	2.1185	2.0862	2.0593	2.0366	2.0171	2.0001	1.9853	1.9722
	0.95	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.5068	2.4753	2.4481	2.4244	2.4034
	0.975	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	3.0078	2.9633	2.9249	2.8915	2.8621
	0.99	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948	3.8049	3.7299	3.6662	3.6115	3.5639	3.5222
	0.995	10.798	7.7008	6.4760	5.8029	5.3721	5.0708	4.8473	4.6744	4.5364	4.4235	4.3295	4.2497	4.1813	4.1219	4.0698

Percentiles de la distribución F de Fisher–Snedecor
con m grados de libertad en el numerador y n en el denominador

n	p	16	17	18	19	20	22	24	25	26	28	30	40	48	60	120
1	0.9	61.350	61.464	61.566	61.658	61.740	61.883	62.002	62.055	62.103	62.190	62.265	62.529	62.662	62.794	63.061
	0.95	246.46	246.92	247.32	247.69	248.01	248.58	249.05	249.26	249.45	249.80	250.10	251.14	251.67	252.20	253.25
	0.975	986.92	988.73	990.35	991.80	993.10	995.36	997.25	998.08	998.85	1000.2	1001.4	1005.6	1007.7	1009.8	1014.0
	0.99	6170.1	6181.4	6191.5	6200.6	6208.7	6222.8	6234.6	6239.8	6244.6	6253.2	6260.6	6286.8	6299.9	6313.0	6339.4
	0.995	24681	24727	24767	24803	24836	24892	24940	24960	24980	25014	25044	25148	25201	25253	25359
2	0.9	9.4289	9.4325	9.4358	9.4387	9.4413	9.4458	9.4496	9.4513	9.4528	9.4556	9.4579	9.4662	9.4704	9.4746	9.4829
	0.95	19.433	19.437	19.440	19.443	19.446	19.450	19.454	19.456	19.457	19.460	19.462	19.471	19.475	19.479	19.487
	0.975	39.435	39.439	39.442	39.445	39.448	39.452	39.456	39.458	39.459	39.462	39.465	39.473	39.477	39.481	39.490
	0.99	99.437	99.440	99.444	99.447	99.449	99.454	99.458	99.459	99.461	99.463	99.466	99.474	99.478	99.482	99.491
	0.995	199.44	199.44	199.44	199.45	199.45	199.45	199.46	199.46	199.46	199.46	199.47	199.47	199.48	199.48	199.49
3	0.9	5.1964	5.1929	5.1898	5.1870	5.1845	5.1801	5.1764	5.1747	5.1732	5.1705	5.1681	5.1597	5.1555	5.1512	5.1425
	0.95	8.6923	8.6829	8.6745	8.6670	8.6602	8.6484	8.6385	8.6341	8.6301	8.6229	8.6166	8.5944	8.5832	8.5720	8.5494
	0.975	14.232	14.213	14.196	14.181	14.167	14.144	14.124	14.115	14.107	14.093	14.081	14.037	14.014	13.992	13.947
	0.99	26.827	26.787	26.751	26.719	26.690	26.640	26.598	26.579	26.562	26.531	26.505	26.411	26.364	26.316	26.221
	0.995	43.008	42.941	42.880	42.826	42.778	42.693	42.622	42.591	42.562	42.511	42.466	42.308	42.229	42.149	41.989
4	0.9	3.8639	3.8582	3.8531	3.8485	3.8443	3.8371	3.8310	3.8283	3.8258	3.8213	3.8174	3.8036	3.7966	3.7896	3.7753
	0.95	5.8441	5.8320	5.8211	5.8114	5.8025	5.7872	5.7744	5.7687	5.7635	5.7541	5.7459	5.7170	5.7024	5.6877	5.6581
	0.975	8.6326	8.6113	8.5924	8.5753	8.5599	8.5332	8.5109	8.5010	8.4919	8.4755	8.4613	8.4111	8.3859	8.3604	8.3092
	0.99	14.154	14.115	14.080	14.048	14.020	13.970	13.929	13.911	13.894	13.864	13.838	13.745	13.699	13.652	13.558
	0.995	20.371	20.311	20.258	20.210	20.167	20.093	20.030	20.002	19.977	19.931	19.892	19.752	19.681	19.611	19.468
5	0.9	3.2303	3.2234	3.2172	3.2117	3.2067	3.1979	3.1905	3.1873	3.1842	3.1788	3.1741	3.1573	3.1488	3.1402	3.1228
	0.95	4.6038	4.5904	4.5785	4.5678	4.5581	4.5413	4.5272	4.5209	4.5151	4.5047	4.4957	4.4638	4.4476	4.4314	4.3985
	0.975	6.4032	6.3814	6.3619	6.3444	6.3286	6.3011	6.2780	6.2679	6.2584	6.2416	6.2269	6.1750	6.1489	6.1225	6.0693
	0.99	9.6802	9.6429	9.6096	9.5797	9.5526	9.5058	9.4665	9.4491	9.4331	9.4043	9.3793	9.2912	9.2467	9.2020	9.1118
	0.995	13.086	13.033	12.985	12.942	12.903	12.836	12.780	12.755	12.732	12.691	12.656	12.530	12.466	12.402	12.274
6	0.9	2.8626	2.8550	2.8481	2.8419	2.8363	2.8266	2.8183	2.8147	2.8113	2.8053	2.8000	2.7812	2.7716	2.7620	2.7423
	0.95	3.9223	3.9083	3.8957	3.8844	3.8742	3.8564	3.8415	3.8348	3.8287	3.8177	3.8082	3.7743	3.7571	3.7398	3.7047
	0.975	5.2439	5.2218	5.2021	5.1844	5.1684	5.1406	5.1172	5.1069	5.0973	5.0802	5.0652	5.0125	4.9858	4.9589	4.9044
	0.99	7.5186	7.4827	7.4507	7.4219	7.3958	7.3506	7.3127	7.2960	7.2805	7.2527	7.2285	7.1432	7.1001	7.0567	6.9690
	0.995	9.7582	9.7086	9.6644	9.6247	9.5888	9.5264	9.4742	9.4511	9.4298	9.3915	9.3582	9.2408	9.1816	9.1219	9.0015
7	0.9	2.6230	2.6148	2.6074	2.6008	2.5947	2.5842	2.5753	2.5714	2.5677	2.5612	2.5555	2.5351	2.5247	2.5142	2.4928
	0.95	3.4944	3.4799	3.4669	3.4551	3.4445	3.4260	3.4105	3.4036	3.3972	3.3858	3.3758	3.3404	3.3225	3.3043	3.2674
	0.975	4.5428	4.5206	4.5008	4.4829	4.4667	4.4386	4.4150	4.4045	4.3949	4.3775	4.3624	4.3089	4.2818	4.2544	4.1989
	0.99	6.2750	6.2401	6.2089	6.1808	6.1554	6.1113	6.0743	6.0580	6.0428	6.0157	5.9920	5.9084	5.8662	5.8236	5.7373
	0.995	7.9148	7.8678	7.8258	7.7881	7.7540	7.6947	7.6450	7.6230	7.6027	7.5662	7.5345	7.4224	7.3658	7.3088	7.1933
8	0.9	2.4545	2.4458	2.4380	2.4310	2.4246	2.4135	2.4041	2.3999	2.3961	2.3891	2.3830	2.3614	2.3503	2.3391	2.3162
	0.95	3.2016	3.1867	3.1733	3.1613	3.1503	3.1313	3.1152	3.1081	3.1015	3.0897	3.0794	3.0428	3.0241	3.0053	2.9669
	0.975	4.0761	4.0538	4.0338	4.0158	3.9995	3.9711	3.9472	3.9367	3.9269	3.9093	3.8940	3.8398	3.8123	3.7844	3.7279
	0.99	5.4766	5.4423	5.4116	5.3840	5.3591	5.3157	5.2793	5.2631	5.2482	5.2214	5.1981	5.1156	5.0738	5.0316	4.9461
	0.995	6.7633	6.7180	6.6775	6.6411	6.6082	6.5510	6.5029	6.4817	6.4620	6.4268	6.3961	6.2875	6.2326	6.1772	6.0649
9	0.9	2.3295	2.3205	2.3123	2.3050	2.2983	2.2867	2.2768	2.2725	2.2684	2.2611	2.2547	2.2320	2.2203	2.2085	2.1843
	0.95	2.9890	2.9737	2.9600	2.9477	2.9365	2.9169	2.9005	2.8932	2.8864	2.8743	2.8637	2.8259	2.8067	2.7872	2.7475
	0.975	3.7441	3.7216	3.7015	3.6833	3.6669	3.6383	3.6142	3.6035	3.5936	3.5759	3.5604	3.5055	3.4775	3.4493	3.3918
	0.99	4.9240	4.8902	4.8599	4.8327	4.8080	4.7651	4.7290	4.7130	4.6982	4.6717	4.6486	4.5666	4.5251	4.4831	4.3978
	0.995	5.9829	5.9388	5.8994	5.8639	5.8318	5.7760	5.7292	5.7084	5.6892	5.6548	5.6248	5.5186	5.4647	5.4104	5.3001
10	0.9	2.2330	2.2237	2.2153	2.2077	2.2007	2.1887	2.1784	2.1739	2.1697	2.1621	2.1554	2.1317	2.1195	2.1072	2.0818
	0.95	2.8276	2.8120	2.7980	2.7854	2.7740	2.7541	2.7372	2.7298	2.7229	2.7104	2.6996	2.6609	2.6411	2.6211	2.5801
	0.975	3.4963	3.4737	3.4534	3.4351	3.4185	3.3897	3.3654	3.3546	3.3446	3.3267	3.3110	3.2554	3.2271	3.1984	3.1399
	0.99	4.5204	4.4869	4.4569	4.4299	4.4054	4.3628	4.3269	4.3111	4.2963	4.2700	4.2469	4.1653	4.1238	4.0819	3.9965
	0.995	5.4221	5.3789	5.3403	5.3055	5.2740	5.2192	5.1732	5.1528	5.1339	5.1001	5.0706	4.9659	4.9128	4.8592	4.7501
11	0.9	2.1563	2.1467	2.1380	2.1302	2.1230	2.1106	2.1000	2.0953	2.0909	2.0831	2.0762	2.0516	2.0390	2.0261	1.9997
	0.95	2.7009	2.6851	2.6709	2.6581	2.6464	2.6261	2.6090	2.6014	2.5943	2.5816	2.5705	2.5309	2.5107	2.4901	2.4480
	0.975	3.3044	3.2816	3.2612	3.2428	3.2261	3.1970	3.1725	3.1616	3.1516	3.1334	3.1176	3.0613	3.0326	3.0035	2.9441
	0.99	4.2134	4.1801	4.1503	4.1234	4.0990	4.0566	4.0209	4.0051	3.9904	3.9641	3.9411	3.8596	3.8181	3.7761	3.6904
	0.995	5.0011	4.9586	4.9205	4.8863	4.8552	4.8012	4.7557	4.7356	4.7170	4.6835	4.6543	4.5508	4.4982	4.4450	4.3367
12	0.9	2.0938	2.0839	2.0750	2.0670	2.0597	2.0469	2.0360	2.0312	2.0267	2.0186	2.0115	1.9861	1.9730	1.9597	1.9323
	0.95	2.5989	2.5828	2.5684	2.5554	2.5436	2.5229	2.5055	2.4977	2.4905	2.4776	2.4663	2.4259	2.4052	2.3842	2.3410
	0.975	3.1515	3.1286	3.1081	3.0896	3.0728	3.0434	3.0187	3.0077	2.9976	2.9793	2.9633	2.9063	2.8773	2.8478	2.7874
	0.99	3.9724	3.9392	3.9095	3.8827	3.8584	3.8161	3.7805	3.7647	3.7500	3.7237	3.7008	3.6192	3.5776	3.5355	3.4494
	0.995	4.6741	4.6321	4.5945	4.5606	4.5299	4.4765	4.4314	4.4115	4.3930	4.3599	4.3309	4.2282	4.1759	4.1229	4.0149
13	0.9	2.0419	2.0318	2.0227	2.0145	2.0070	1.9939	1.9827	1.9778	1.9732	1.9649	1.9576	1.9315	1.9180	1.9043	1.8759
	0.95	2.5149	2.4987	2.4841	2.4709	2.4589	2.4379	2.4202	2.4123	2.4050	2.3918	2.3803	2.3392	2.3181	2.2966	2.2524
	0.975	3.0269	3.0039	2.9832	2.9646	2.9477	2.9181	2.8932	2.8821	2.8719	2.8534	2.8372	2.7797	2.7503	2.7204	2.6590
	0.99	3.7783	3.7452	3.7156	3.6888	3.6646	3.6224	3.5868	3.5710	3.5563	3.5300	3.5070	3.4253	3.3836	3.3413	3.2548
	0.995	4.4132	4.3716	4.3344	4.3008	4.2703	4.2173	4.1726	4.1528	4.1344	4.1015	4.0727	3.9704	3.9183	3.8655	3.7577
14	0.9	1.9981	1.9878	1.9785	1.9701	1.9625	1.9490	1.9377	1.9326	1.9279	1.9194	1.9119	1.8852	1.8714	1.8572	1.8280
	0.95	2.4446	2.4282	2.4134	2.4000	2.3879	2.3667	2.3487	2.3407	2.3333	2.3199	2.3082	2.2664	2.2449	2.2229	2.1778
	0.975	2.9234	2.9003	2.8795	2.8607	2.8437	2.8139	2.7888	2.7777	2.7673	2.7487	2.7324	2.6742	2.6444	2.6142	2.5519
	0.99	3.6187	3.5857	3.5561	3.5294	3.5052	3.4630	3.4274	3.4116	3.3969	3.3706	3.3476	3.2656	3.2238	3.1813	3.0942
	0.995	4.2005	4.1592	4.1221	4.0888	4.0585	4.0058	3.9614	3.9417	3.9234	3.8906	3.8619	3.7600	3.7080	3.6552	3.5473
15	0.9	1.9605	1.9501	1.9407	1.9321	1.9243	1.9106	1.8990	1.8939	1.8891	1.8804	1.8728	1.8454	1.8312	1.8168	1.7867
	0.95	2.3849	2.3683	2.3533	2.3398	2.3275	2.3060	2.2878	2.2797	2.2722	2.2587	2.2468	2.2043	2.1824	2.1601	2.1141
	0.975	2.8360	2.8128	2.7919	2.7730	2.7559	2.7260	2.7006	2.6894	2.6790	2.6602	2.6437	2.5850	2.5549	2.5242	2.4611
	0.99	3.4852	3.4523	3.4228	3.3961	3.3719	3.3297	3.2940	3.2782	3.2635	3.2372	3.2141	3.1319	3.0899	3.0471	2.9595
	0.995	4.0237	3.9827	3.9459	3.9127	3.8826	3.8301	3.7859	3.7662	3.7480	3.7153	3.6867	3.5850	3.5330	3.4803	3.3722

Percentiles de la distribución F de Fisher–Snedecor
con m grados de libertad en el numerador y n en el denominador

n	p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	0.9	3.0481	2.6682	2.4618	2.3327	2.2438	2.1783	2.1280	2.0880	2.0553	2.0281	2.0051	1.9854	1.9682	1.9532	1.9399
	0.95	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4564	2.4247	2.3973	2.3733	2.3522
	0.975	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.9337	2.8890	2.8506	2.8170	2.7875
	0.99	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804	3.6909	3.6162	3.5527	3.4981	3.4506	3.4089
	0.995	10.575	7.5138	6.3034	5.6378	5.2117	4.9134	4.6920	4.5207	4.3838	4.2719	4.1785	4.0994	4.0314	3.9723	3.9205
17	0.9	3.0262	2.6446	2.4374	2.3077	2.2183	2.1524	2.1017	2.0613	2.0284	2.0009	1.9777	1.9577	1.9404	1.9252	1.9117
	0.95	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.4126	2.3807	2.3531	2.3290	2.3077
	0.975	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8696	2.8249	2.7863	2.7526	2.7230
	0.99	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822	3.5931	3.5185	3.4552	3.4007	3.3533	3.3117
	0.995	10.384	7.3536	6.1556	5.4967	5.0746	4.7789	4.5594	4.3894	4.2535	4.1424	4.0496	3.9709	3.9033	3.8445	3.7929
18	0.9	3.0070	2.6239	2.4160	2.2858	2.1958	2.1296	2.0785	2.0379	2.0047	1.9770	1.9535	1.9333	1.9158	1.9004	1.8868
	0.95	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3742	2.3421	2.3143	2.2900	2.2686
	0.975	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.8137	2.7689	2.7302	2.6964	2.6667
	0.99	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971	3.5082	3.4338	3.3706	3.3162	3.2689	3.2273
	0.995	10.218	7.2148	6.0278	5.3746	4.9560	4.6627	4.4448	4.2759	4.1410	4.0305	3.9382	3.8599	3.7926	3.7341	3.6827
19	0.9	2.9899	2.6056	2.3970	2.2663	2.1760	2.1094	2.0580	2.0171	1.9836	1.9557	1.9321	1.9117	1.8940	1.8785	1.8647
	0.95	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3402	2.3080	2.2800	2.2556	2.2341
	0.975	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7645	2.7196	2.6808	2.6469	2.6171
	0.99	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225	3.4338	3.3596	3.2965	3.2422	3.1949	3.1533
	0.995	10.073	7.0935	5.9161	5.2681	4.8526	4.5614	4.3448	4.1770	4.0428	3.9329	3.8410	3.7631	3.6961	3.6378	3.5866
20	0.9	2.9747	2.5893	2.3801	2.2489	2.1582	2.0913	2.0397	1.9985	1.9649	1.9367	1.9129	1.8924	1.8745	1.8588	1.8449
	0.95	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.3100	2.2776	2.2495	2.2250	2.2033
	0.975	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.7209	2.6758	2.6369	2.6030	2.5731
	0.99	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567	3.3682	3.2941	3.2311	3.1769	3.1296	3.0880
	0.995	9.9439	6.9865	5.8177	5.1743	4.7616	4.4721	4.2569	4.0900	3.9564	3.8470	3.7555	3.6779	3.6111	3.5530	3.5020
22	0.9	2.9486	2.5613	2.3512	2.2193	2.1279	2.0605	2.0084	1.9668	1.9327	1.9043	1.8801	1.8593	1.8411	1.8252	1.8111
	0.95	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2585	2.2258	2.1975	2.1727	2.1508
	0.975	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6469	2.6017	2.5626	2.5285	2.4984
	0.99	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458	3.2576	3.1837	3.1209	3.0667	3.0195	2.9779
	0.995	9.7271	6.8064	5.6524	5.0168	4.6088	4.3225	4.1094	3.9440	3.8116	3.7030	3.6122	3.5350	3.4686	3.4108	3.3600
24	0.9	2.9271	2.5383	2.3274	2.1949	2.1030	2.0351	1.9826	1.9407	1.9063	1.8775	1.8530	1.8319	1.8136	1.7974	1.7831
	0.95	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.2163	2.1834	2.1548	2.1298	2.1077
	0.975	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5865	2.5411	2.5019	2.4677	2.4374
	0.99	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560	3.1681	3.0944	3.0316	2.9775	2.9303	2.8887
	0.995	9.5513	6.6609	5.5190	4.8898	4.4857	4.2019	3.9905	3.8264	3.6949	3.5870	3.4967	3.4199	3.3538	3.2962	3.2456
25	0.9	2.9177	2.5283	2.3170	2.1842	2.0922	2.0241	1.9714	1.9292	1.8947	1.8658	1.8412	1.8200	1.8015	1.7853	1.7708
	0.95	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1979	2.1649	2.1362	2.1111	2.0889
	0.975	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5603	2.5149	2.4756	2.4413	2.4110
	0.99	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172	3.1294	3.0558	2.9931	2.9389	2.8917	2.8502
	0.995	9.4753	6.5982	5.4615	4.8351	4.4327	4.1500	3.9394	3.7758	3.6447	3.5370	3.4470	3.3704	3.3044	3.2469	3.1963
26	0.9	2.9091	2.5191	2.3075	2.1745	2.0822	2.0139	1.9610	1.9188	1.8841	1.8550	1.8303	1.8090	1.7904	1.7741	1.7596
	0.95	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1811	2.1479	2.1192	2.0939	2.0716
	0.975	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.5363	2.4908	2.4515	2.4171	2.3867
	0.99	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818	3.0941	3.0205	2.9578	2.9038	2.8566	2.8150
	0.995	9.4059	6.5409	5.4091	4.7852	4.3844	4.1027	3.8928	3.7297	3.5989	3.4916	3.4017	3.3252	3.2594	3.2020	3.1515
28	0.9	2.8938	2.5028	2.2906	2.1571	2.0645	1.9959	1.9427	1.9001	1.8652	1.8359	1.8110	1.7895	1.7708	1.7542	1.7395
	0.95	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1512	2.1179	2.0889	2.0635	2.0411
	0.975	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4940	2.4484	2.4089	2.3743	2.3438
	0.99	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195	3.0320	2.9585	2.8959	2.8418	2.7946	2.7530
	0.995	9.2838	6.4403	5.3170	4.6977	4.2996	4.0197	3.8110	3.6487	3.5186	3.4117	3.3222	3.2460	3.1803	3.1231	3.0727
30	0.9	2.8807	2.4887	2.2761	2.1422	2.0492	1.9803	1.9269	1.8841	1.8490	1.8195	1.7944	1.7727	1.7538	1.7371	1.7223
	0.95	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.1256	2.0921	2.0630	2.0374	2.0148
	0.975	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4577	2.4120	2.3724	2.3378	2.3072
	0.99	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665	2.9791	2.9057	2.8431	2.7890	2.7418	2.7002
	0.995	9.1797	6.3547	5.2388	4.6234	4.2276	3.9492	3.7416	3.5801	3.4505	3.3440	3.2547	3.1787	3.1132	3.0560	3.0057
40	0.9	2.8354	2.4404	2.2261	2.0909	1.9968	1.9269	1.8725	1.8289	1.7929	1.7627	1.7369	1.7146	1.6950	1.6778	1.6624
	0.95	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0376	2.0035	1.9738	1.9476	1.9245
	0.975	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.3343	2.2882	2.2481	2.2130	2.1819
	0.99	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876	2.8005	2.7274	2.6648	2.6107	2.5634	2.5216
	0.995	8.8279	6.0664	4.9758	4.3738	3.9860	3.7129	3.5088	3.3498	3.2220	3.1167	3.0284	2.9531	2.8880	2.8312	2.7811
48	0.9	2.8131	2.4167	2.2016	2.0658	1.9711	1.9006	1.8458	1.8017	1.7653	1.7347	1.7085	1.6859	1.6660	1.6485	1.6328
	0.95	4.0427	3.1907	2.7981	2.5652	2.4085	2.2946	2.2074	2.1382	2.0817	2.0346	1.9946	1.9601	1.9301	1.9037	1.8802
	0.975	5.3541	3.9875	3.4022	3.0662	2.8444	2.6852	2.5646	2.4696	2.3925	2.3286	2.2745	2.2281	2.1878	2.1524	2.1210
	0.99	7.1942	5.0767	4.2180	3.7374	3.4251	3.2036	3.0372	2.9069	2.8018	2.7150	2.6418	2.5793	2.5252	2.4777	2.4358
	0.995	8.6590	5.9287	4.8505	4.2549	3.8711	3.6005	3.3981	3.2403	3.1133	3.0087	2.9208	2.8458	2.7809	2.7242	2.6741
60	0.9	2.7911	2.3933	2.1774	2.0410	1.9457	1.8747	1.8194	1.7748	1.7380	1.7070	1.6805	1.6574	1.6372	1.6193	1.6034
	0.95	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9522	1.9174	1.8870	1.8602	1.8364
	0.975	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.2159	2.1692	2.1286	2.0929	2.0613
	0.99	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185	2.6318	2.5587	2.4961	2.4419	2.3943	2.3523
	0.995	8.4946	5.7950	4.7290	4.1399	3.7599	3.4918	3.2911	3.1344	3.0083	2.9042	2.8166	2.7419	2.6771	2.6205	2.5705
120	0.9	2.7478	2.3473	2.1300	1.9923	1.8959	1.8238	1.7675	1.7220	1.6842	1.6524	1.6250	1.6012	1.5803	1.5617	1.5450
	0.95	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8693	1.8337	1.8026	1.7750	1.7505
	0.975	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.1021	2.0548	2.0136	1.9773	1.9450
	0.99	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586	2.4721	2.3990	2.3363	2.2818	2.2339	2.1915
	0.995	8.1788	5.5393	4.4972	3.9207	3.5482	3.2849	3.0874	2.9330	2.8083	2.7052	2.6183	2.5439	2.4794	2.4228	2.3727

Percentiles de la distribución F de Fisher–Snedecor
con m grados de libertad en el numerador y n en el denominador

n	p	16	17	18	19	20	22	24	25	26	28	30	40	48	60	120
16	0.9	1.9281	1.9175	1.9079	1.8992	1.8913	1.8774	1.8656	1.8603	1.8554	1.8466	1.8388	1.8108	1.7964	1.7816	1.7507
	0.95	2.3335	2.3167	2.3016	2.2880	2.2756	2.2538	2.2354	2.2272	2.2196	2.2059	2.1938	2.1507	2.1285	2.1058	2.0589
	0.975	2.7614	2.7380	2.7170	2.6980	2.6808	2.6507	2.6252	2.6138	2.6033	2.5844	2.5678	2.5085	2.4781	2.4471	2.3831
	0.99	3.3720	3.3391	3.3096	3.2829	3.2587	3.2165	3.1808	3.1650	3.1503	3.1238	3.1007	3.0182	2.9760	2.9330	2.8447
	0.995	3.8747	3.8338	3.7972	3.7641	3.7342	3.6819	3.6378	3.6182	3.6000	3.5674	3.5389	3.4372	3.3852	3.3324	3.2240
17	0.9	1.8997	1.8889	1.8792	1.8704	1.8624	1.8482	1.8362	1.8309	1.8259	1.8169	1.8090	1.7805	1.7658	1.7506	1.7191
	0.95	2.2888	2.2719	2.2567	2.2429	2.2304	2.2084	2.1898	2.1815	2.1738	2.1599	2.1477	2.1040	2.0815	2.0584	2.0107
	0.975	2.6968	2.6733	2.6522	2.6331	2.6158	2.5855	2.5598	2.5484	2.5378	2.5187	2.5020	2.4422	2.4115	2.3801	2.3153
	0.99	3.2748	3.2419	3.2124	3.1857	3.1615	3.1192	3.0835	3.0676	3.0529	3.0264	3.0032	2.9205	2.8780	2.8348	2.7459
	0.995	3.7473	3.7066	3.6701	3.6372	3.6073	3.5552	3.5112	3.4916	3.4735	3.4409	3.4124	3.3108	3.2587	3.2058	3.0971
18	0.9	1.8747	1.8638	1.8539	1.8450	1.8368	1.8225	1.8103	1.8049	1.7999	1.7907	1.7827	1.7537	1.7387	1.7232	1.6910
	0.95	2.2496	2.2325	2.2172	2.2033	2.1906	2.1685	2.1497	2.1413	2.1335	2.1195	2.1071	2.0629	2.0400	2.0166	1.9681
	0.975	2.6404	2.6168	2.5956	2.5764	2.5590	2.5285	2.5027	2.4912	2.4806	2.4613	2.4445	2.3842	2.3531	2.3214	2.2558
	0.99	3.1904	3.1575	3.1280	3.1013	3.0771	3.0348	2.9990	2.9831	2.9683	2.9418	2.9185	2.8354	2.7928	2.7493	2.6597
	0.995	3.6373	3.5967	3.5603	3.5275	3.4977	3.4456	3.4017	3.3822	3.3641	3.3315	3.3030	3.2014	3.1493	3.0962	2.9871
19	0.9	1.8524	1.8414	1.8314	1.8224	1.8142	1.7997	1.7873	1.7818	1.7767	1.7674	1.7592	1.7298	1.7145	1.6988	1.6659
	0.95	2.2149	2.1977	2.1823	2.1683	2.1555	2.1331	2.1141	2.1057	2.0978	2.0836	2.0712	2.0264	2.0033	1.9795	1.9302
	0.975	2.5907	2.5670	2.5457	2.5265	2.5089	2.4783	2.4523	2.4408	2.4300	2.4107	2.3937	2.3329	2.3016	2.2696	2.2032
	0.99	3.1165	3.0836	3.0541	3.0274	3.0031	2.9607	2.9249	2.9089	2.8941	2.8675	2.8442	2.7608	2.7179	2.6742	2.5839
	0.995	3.5412	3.5008	3.4645	3.4318	3.4020	3.3500	3.3062	3.2867	3.2686	3.2360	3.2075	3.1058	3.0536	3.0004	2.8908
20	0.9	1.8325	1.8214	1.8113	1.8022	1.7938	1.7792	1.7667	1.7611	1.7559	1.7465	1.7382	1.7083	1.6928	1.6768	1.6433
	0.95	2.1840	2.1667	2.1511	2.1370	2.1242	2.1016	2.0825	2.0739	2.0660	2.0517	2.0391	1.9938	1.9704	1.9464	1.8963
	0.975	2.5465	2.5228	2.5014	2.4821	2.4645	2.4337	2.4076	2.3959	2.3851	2.3657	2.3486	2.2873	2.2557	2.2234	2.1562
	0.99	3.0512	3.0183	2.9887	2.9620	2.9377	2.8953	2.8594	2.8434	2.8286	2.8019	2.7785	2.6947	2.6517	2.6077	2.5168
	0.995	3.4568	3.4164	3.3802	3.3475	3.3178	3.2659	3.2220	3.2025	3.1845	3.1519	3.1234	3.0215	2.9692	2.9159	2.8058
22	0.9	1.7984	1.7871	1.7768	1.7675	1.7590	1.7440	1.7312	1.7255	1.7202	1.7106	1.7021	1.6714	1.6554	1.6389	1.6041
	0.95	2.1313	2.1138	2.0980	2.0837	2.0707	2.0478	2.0283	2.0196	2.0116	1.9970	1.9842	1.9380	1.9141	1.8894	1.8380
	0.975	2.4717	2.4478	2.4262	2.4067	2.3890	2.3579	2.3315	2.3198	2.3088	2.2891	2.2718	2.2097	2.1775	2.1446	2.0760
	0.99	2.9411	2.9082	2.8786	2.8518	2.8274	2.7849	2.7488	2.7328	2.7179	2.6910	2.6675	2.5831	2.5396	2.4951	2.4029
	0.995	3.3150	3.2748	3.2387	3.2060	3.1764	3.1246	3.0807	3.0613	3.0432	3.0106	2.9821	2.8799	2.8273	2.7736	2.6625
24	0.9	1.7703	1.7587	1.7483	1.7388	1.7302	1.7149	1.7019	1.6960	1.6906	1.6808	1.6721	1.6407	1.6242	1.6073	1.5715
	0.95	2.0880	2.0703	2.0543	2.0399	2.0267	2.0035	1.9838	1.9750	1.9668	1.9520	1.9390	1.8920	1.8675	1.8424	1.7896
	0.975	2.4105	2.3865	2.3648	2.3452	2.3273	2.2959	2.2693	2.2574	2.2464	2.2265	2.2090	2.1460	2.1134	2.0799	2.0099
	0.99	2.8519	2.8189	2.7892	2.7624	2.7380	2.6953	2.6591	2.6430	2.6280	2.6010	2.5773	2.4923	2.4484	2.4035	2.3100
	0.995	3.2007	3.1606	3.1246	3.0920	3.0624	3.0106	2.9667	2.9472	2.9291	2.8965	2.8679	2.7654	2.7125	2.6585	2.5463
25	0.9	1.7579	1.7463	1.7358	1.7263	1.7175	1.7021	1.6890	1.6831	1.6776	1.6677	1.6589	1.6272	1.6105	1.5934	1.5570
	0.95	2.0691	2.0513	2.0353	2.0207	2.0075	1.9842	1.9643	1.9554	1.9472	1.9323	1.9192	1.8718	1.8471	1.8217	1.7684
	0.975	2.3840	2.3599	2.3381	2.3184	2.3005	2.2690	2.2422	2.2303	2.2192	2.1992	2.1816	2.1183	2.0854	2.0516	1.9811
	0.99	2.8133	2.7803	2.7506	2.7238	2.6993	2.6565	2.6203	2.6041	2.5891	2.5620	2.5383	2.4530	2.4089	2.3637	2.2696
	0.995	3.1515	3.1114	3.0754	3.0429	3.0133	2.9615	2.9176	2.8981	2.8800	2.8473	2.8187	2.7160	2.6630	2.6088	2.4961
26	0.9	1.7466	1.7349	1.7243	1.7147	1.7059	1.6904	1.6771	1.6712	1.6657	1.6556	1.6468	1.6147	1.5979	1.5805	1.5437
	0.95	2.0518	2.0339	2.0178	2.0032	1.9898	1.9664	1.9464	1.9375	1.9292	1.9142	1.9010	1.8533	1.8284	1.8027	1.7488
	0.975	2.3597	2.3355	2.3137	2.2939	2.2759	2.2443	2.2174	2.2054	2.1943	2.1742	2.1565	2.0928	2.0597	2.0257	1.9545
	0.99	2.7781	2.7451	2.7153	2.6885	2.6640	2.6211	2.5848	2.5686	2.5536	2.5264	2.5026	2.4170	2.3727	2.3273	2.2325
	0.995	3.1067	3.0666	3.0306	2.9981	2.9685	2.9167	2.8728	2.8533	2.8352	2.8025	2.7738	2.6709	2.6178	2.5633	2.4501
28	0.9	1.7264	1.7146	1.7039	1.6941	1.6852	1.6695	1.6560	1.6500	1.6444	1.6342	1.6252	1.5925	1.5753	1.5575	1.5198
	0.95	2.0210	2.0030	1.9868	1.9720	1.9586	1.9349	1.9147	1.9057	1.8973	1.8821	1.8687	1.8203	1.7950	1.7689	1.7138
	0.975	2.3167	2.2924	2.2704	2.2505	2.2324	2.2006	2.1735	2.1615	2.1502	2.1299	2.1121	2.0477	2.0142	1.9797	1.9072
	0.99	2.7160	2.6830	2.6532	2.6263	2.6017	2.5587	2.5223	2.5060	2.4909	2.4636	2.4397	2.3535	2.3088	2.2629	2.1670
	0.995	3.0279	2.9879	2.9520	2.9194	2.8899	2.8380	2.7941	2.7746	2.7564	2.7236	2.6949	2.5916	2.5381	2.4834	2.3690
30	0.9	1.7090	1.6970	1.6862	1.6763	1.6673	1.6514	1.6377	1.6316	1.6259	1.6156	1.6065	1.5732	1.5557	1.5376	1.4989
	0.95	1.9946	1.9765	1.9601	1.9452	1.9317	1.9077	1.8874	1.8782	1.8698	1.8544	1.8409	1.7918	1.7661	1.7396	1.6835
	0.975	2.2799	2.2554	2.2334	2.2134	2.1952	2.1631	2.1359	2.1237	2.1124	2.0919	2.0739	2.0089	1.9750	1.9400	1.8664
	0.99	2.6632	2.6301	2.6003	2.5732	2.5487	2.5055	2.4689	2.4526	2.4374	2.4100	2.3860	2.2992	2.2542	2.2079	2.1108
	0.995	2.9611	2.9211	2.8852	2.8526	2.8230	2.7712	2.7272	2.7076	2.6894	2.6566	2.6278	2.5241	2.4703	2.4151	2.2998
40	0.9	1.6486	1.6362	1.6249	1.6146	1.6052	1.5884	1.5741	1.5677	1.5617	1.5507	1.5411	1.5056	1.4868	1.4672	1.4248
	0.95	1.9037	1.8851	1.8682	1.8529	1.8389	1.8141	1.7929	1.7835	1.7746	1.7586	1.7444	1.6928	1.6656	1.6373	1.5766
	0.975	2.1542	2.1293	2.1068	2.0864	2.0677	2.0349	2.0069	1.9943	1.9827	1.9615	1.9429	1.8752	1.8396	1.8028	1.7242
	0.99	2.4844	2.4511	2.4210	2.3937	2.3689	2.3252	2.2880	2.2714	2.2559	2.2280	2.2034	2.1142	2.0676	2.0194	1.9172
	0.995	2.7365	2.6966	2.6607	2.6281	2.5984	2.5463	2.5020	2.4823	2.4639	2.4307	2.4015	2.2958	2.2407	2.1838	2.0636
48	0.9	1.6187	1.6060	1.5945	1.5839	1.5743	1.5571	1.5424	1.5358	1.5296	1.5183	1.5084	1.4716	1.4520	1.4314	1.3867
	0.95	1.8592	1.8402	1.8231	1.8075	1.7932	1.7680	1.7464	1.7367	1.7276	1.7112	1.6967	1.6435	1.6154	1.5859	1.5224
	0.975	2.0931	2.0679	2.0452	2.0245	2.0056	1.9723	1.9438	1.9311	1.9192	1.8977	1.8787	1.8094	1.7728	1.7347	1.6529
	0.99	2.3985	2.3650	2.3348	2.3073	2.2823	2.2383	2.2007	2.1839	2.1683	2.1400	2.1150	2.0244	1.9768	1.9273	1.8217
	0.995	2.6295	2.5896	2.5536	2.5210	2.4912	2.4389	2.3944	2.3745	2.3560	2.3225	2.2930	2.1861	2.1300	2.0720	1.9483
60	0.9	1.5890	1.5760	1.5642	1.5534	1.5435	1.5259	1.5107	1.5039	1.4975	1.4859	1.4755	1.4373	1.4168	1.3952	1.3476
	0.95	1.8151	1.7959	1.7784	1.7625	1.7480	1.7222	1.7001	1.6902	1.6809	1.6641	1.6491	1.5943	1.5650	1.5343	1.4673
	0.975	2.0330	2.0076	1.9846	1.9636	1.9445	1.9106	1.8817	1.8687	1.8566	1.8346	1.8152	1.7440	1.7063	1.6668	1.5810
	0.99	2.3148	2.2811	2.2507	2.2230	2.1978	2.1533	2.1154	2.0984	2.0825	2.0538	2.0285	1.9360	1.8872	1.8363	1.7263
	0.995	2.5259	2.4859	2.4498	2.4171	2.3872	2.3346	2.2898	2.2697	2.2511	2.2172	2.1874	2.0789	2.0217	1.9622	1.8341
120	0.9	1.5300	1.5164	1.5039	1.4926	1.4821	1.4634	1.4472	1.4399	1.4331	1.4205	1.4094	1.3676	1.3448	1.3203	1.2646
	0.95	1.7285	1.7085	1.6904	1.6739	1.6587	1.6317	1.6084	1.5980	1.5881	1.5703	1.5543	1.4952	1.4632	1.4290	1.3519
	0.975	1.9161	1.8900	1.8663	1.8447	1.8249	1.7899	1.7597	1.7462	1.7335	1.7104	1.6899	1.6141	1.5733	1.5299	1.4327
	0.99	2.1536	2.1194	2.0885	2.0604	2.0346	1.9891	1.9500	1.9325	1.9161	1.8864	1.8600	1.7628	1.7108	1.6557	1.5330
	0.995	2.3280	2.2878	2.2514	2.2183	2.1881	2.1347	2.0890	2.0686	2.0494	2.0147	1.9840	1.8709	1.8106	1.7469	1.6055

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS AGRONOMICAS
Departamento de Economía Agraria