ÍNDICE

APUNTES DE ESTADÍSTICA DESCRIPTIVA

(José Santic A.)

1 CONCEPTOS BÁSICOS 4
1.1 Estadística descriptiva: 4
1.2 Universo y población 4
1.2.1 Universo 4
1.2.2 Población 4
1.2.3 Ejemplos de poblaciones: 4
1.2.4 Clasificación de las poblaciones: 4
1.3 Muestra 5
1.3.1 Formas de elegir una muestra 5
1.3.2 Ejemplos de muestras: 5
1.3.3 Representatividad de la muestra 6
1.3.4 Ventajas de las muestras: 6
1.4 Variables estadísticas: 7
1.4.1 Variables cuantitativas: 7
1.4.2 Variables cualitativas o atributos: 8
1.5 Encuesta: 8
2 RECOPILACIÓN DE LOS DATOS 10
3 ORGANIZACIÓN DE LOS DATOS 10
3.1 Proceso de organización de los datos: 10
3.2 Escalas de medición de variables 11
3.2.1 Escala nominal: 11
3.2.2 Escala ordinal: 11
3.2.3 Escala de intervalos: 12
3.2.4 Escala de razón: 12
3.3 Distribución de frecuencias 13
3.3.1 Tablas de frecuencia de tipo I: 13
3.3.2 Tablas de frecuencia de tipo II: 13
3.3.3 Tablas de frecuencia de tipo III (tabla de intervalos de clase): 14
3.4 Construcción de una tabla de distribución de frecuencias de tipo III (tabla de intervalos de clase) 14
3.5 Representación gráfica de una tabla de distribución de frecuencias de variable continua (datos agrupados) 20

3.5.1 Histograma: ... 20
3.5.2 Polígono de frecuencias: ... 21
3.5.3 Polígono de frecuencias acumuladas u Ojiva ... 22
3.5.4 Gráfico de líneas: ... 23
3.5.5 Gráficos circulares o de torta: ... 23
3.5.6 Diagrama de tallo y hojas: ... 24
3.6 Representación gráfica de una tabla de distribución de frecuencias de variable discreta (datos no agrupados) ... 25
3.7 Ejercicios propuestos: ... 25
4 ESTADÍGRAFOS ... 29
4.1 ESTADÍGRAFOS DE POSICIÓN O MEDIDAS DE TENDENCIA CENTRAL 29
4.1.1 Media aritmética ... 30
4.1.2 Mediana ... 34
4.1.3 Moda o valor modal ... 39
4.1.4 Relación entre la media aritmética, mediana y moda ... 44
4.1.5 Media geométrica ... 44
4.1.6 Ejercicios resueltos: ... 45
4.1.7 Ejercicios propuestos: ... 48
4.2 ESTADÍGRAFOS DE LOCALIZACIÓN ... 49
4.2.1 Cuartiles: ... 49
4.2.2 Deciles: ... 62
4.2.3 Percentiles: ... 67
4.3 ESTADÍGRAFOS DE DISPERSIÓN O DE VARIABILIDAD ... 70
4.3.1 Rango o recorrido: ... 70
4.3.2 Coeficiente de apertura: ... 71
4.3.3 Desviación media: ... 71
4.3.4 Varianza: ... 75
4.3.5 Desviación estándar o típica: ... 77
4.3.6 Coeficiente de variación: ... 88
4.3.7 Coeficiente de variabilidad o Coeficiente de variación ... 88
5 PUNTUACIONES TÍPICAS ... 89
5.1 Puntuación sigma z: (letra z en minúscula) ... 89
5.2 Puntuación típica Z (letra Z en mayúscula) ... 91
6 DISTRIBUCIÓN NORMAL ... 92

6.1 Conceptos básicos: ... 92

6.1.1 Variable aleatoria continua: ... 92

6.1.2 Distribución normal, curva normal o campana de Gauss: ... 93

6.2 Cálculo de probabilidades en distribuciones normales estándar o tipificada ... 95

6.2.1 Tabla de distribución normal tipificada $N(0,1)$, que proporciona, para cada valor de z, el área que queda a su izquierda (Tabla I). ... 96

6.2.2 Tabla de distribución normal tipificada $N(0,1)$, que proporciona el área comprendida entre 0 y z (Tabla II). ... 96

6.3 Ejercicios sobre la curva normal ... 111

7 REFERENCIAS BIBLIOGRÁFICAS ... 120

1 CONCEPTOS BÁSICOS

1.1 Estadística descriptiva:

Es una rama de la Estadística que se ocupa de la recopilación, organización, análisis y representación de los datos obtenidos de un estudio referido a una población o a una muestra, es decir, a toda o a solo una parte de ella.

1.2 Universo y población

1.2.1 Universo

Es el conjunto de elementos (individuos u objetos) que interesa estudiar.

1.2.2 Población

Es un conjunto de datos de una determinada característica que se observa o mide en cada elemento. Según esta definición, en un mismo universo se puede disponer de varias poblaciones dependiendo de las características que se quiere estudiar. Esas características, que se denominan variables, pueden ser: sexo, peso, nacionalidad, estado civil, ingreso mensual, nivel de estudios, orientación política, etc.

1.2.3 Ejemplos de poblaciones:

- Las personas afiliadas a las Administradoras de Fondos de Pensiones (AFP).
- Los habitantes de la comuna de Maipú.
- Los países de la Unión Europea.
- Los libros de una biblioteca.
- Los clubes de fútbol de la Asociación Nacional de Futbol Profesional de Chile (ANFP).
- Los colegios municipalizados en Chile.
- Los militantes de los partidos políticos.

1.2.4 Clasificación de las poblaciones:

- Poblaciones infinitas: Cuando la cantidad de elementos es infinita o muy grande. Ejemplos: los números enteros positivos, los puntos de una línea, la cantidad de granos en un granero.

- Poblaciones finitas: Si la cantidad de elementos es finito o limitado. Por ejemplo:
    - Los estudiantes del Colegio San Gabriel.
    - Los profesores normalistas del país.
    - Las educadoras de párvulos de la Región Metropolitana.
    - Los habitantes de la comuna de Recoleta en Santiago.
    - Los libros de una biblioteca en un momento determinado.
    - La cantidad de teléfonos móviles por hogar en la ciudad de Curicó.
    - Los artículos fabricados por una empresa en un mes.

1.3 Muestra

Es un subconjunto de elementos representativo de la población y que son objeto de estudio. En Estadística se suele trabajar con muestras y no con poblaciones, por ser éstas generalmente muy grandes o sus miembros estar muy dispersos, por lo que los estudios demandan más recursos económicos y de tiempo. Por ejemplo, no es posible conocer la masa corporal de todos los habitantes del país que son adultos, por lo que se elige una muestra que, si es representativa de todos esos habitantes, el resultado del estudio se puede hacer extensivo a toda la población adulta. El número de elementos incluidos en una muestra se llama *tamaño de la muestra*, y puede variar entre un elemento y el número total de la población. La estadística descriptiva solo busca describir y analizar los elementos de la muestra sin sacar conclusiones o inferencias a la población correspondiente. De esto último se ocupa la estadística inferencial.

1.3.1 Formas de elegir una muestra

- Muestra dirigida: La selección de los individuos de la población se efectúa según la conveniencia del investigador.

- Muestra aleatoria: Son aquellas donde los individuos de la población son seleccionados al azar, por ejemplo, a través de números aleatorios.

1.3.2 Ejemplos de muestras:

- Si la población son los estudiantes de la Universidad Andrés Bello, una muestra podría estar integrada por los matriculados en la Casa Central ubicada en la calle República de Santiago o por aquellos que están cursando una determinada carrera, como, por ejemplo, ingeniería comercial.

- Si la población son los estudiantes de la escuela mixta Juan Pablo Duarte de Providencia, una muestra podría corresponder a sus alumnos de sexo femenino.

- Si la población son los habitantes de Punta Arenas, una muestra podría ser los habitantes que residen en el Barrio Sur.

- Medir el tiempo de duración de un lote de ampolletas.

  El caso particular de una muestra que incluye a todos los elementos de la población se le llama *censo*.

1.3.3 Representatividad de la muestra

Para que una muestra sea representativa de la población, debe cumplir con dos requisitos:

- Sus elementos deben estar seleccionados al azar.

- Los elementos de la población deben estar bien representados en la muestra. Por ejemplo, si se quiere determinar la estatura promedio del joven chileno cuando entra a la adultez, es decir, a los 18 años, la muestra debe incluir a jóvenes de todas las regiones, tanto urbanas como rurales, e incluir a personas tanto de sexo masculino como femenino.

1.3.4 Ventajas de las muestras:

Según Grassau (1964), las principales ventajas de las muestras son las siguientes:

- Reducción del costo: Por la menor cantidad de material y personal que se utiliza en la investigación y porque la muestra es menos dispersa que la población, lo que implica una economía de gastos en pasajes y viáticos.

- Menor tiempo para realizar la investigación: Por el menor número de personas u objetos que se van a estudiar.

- Menores exigencias al público: el público debe responder a una menor cantidad de cuestionarios.

- Disminución de personal y espacios de trabajo: El trabajo de campo y el de codificación y tabulación de datos requiere de menos horas. Además, el personal disponible para realizar estas dos últimas funciones suele ser relativamente escaso.

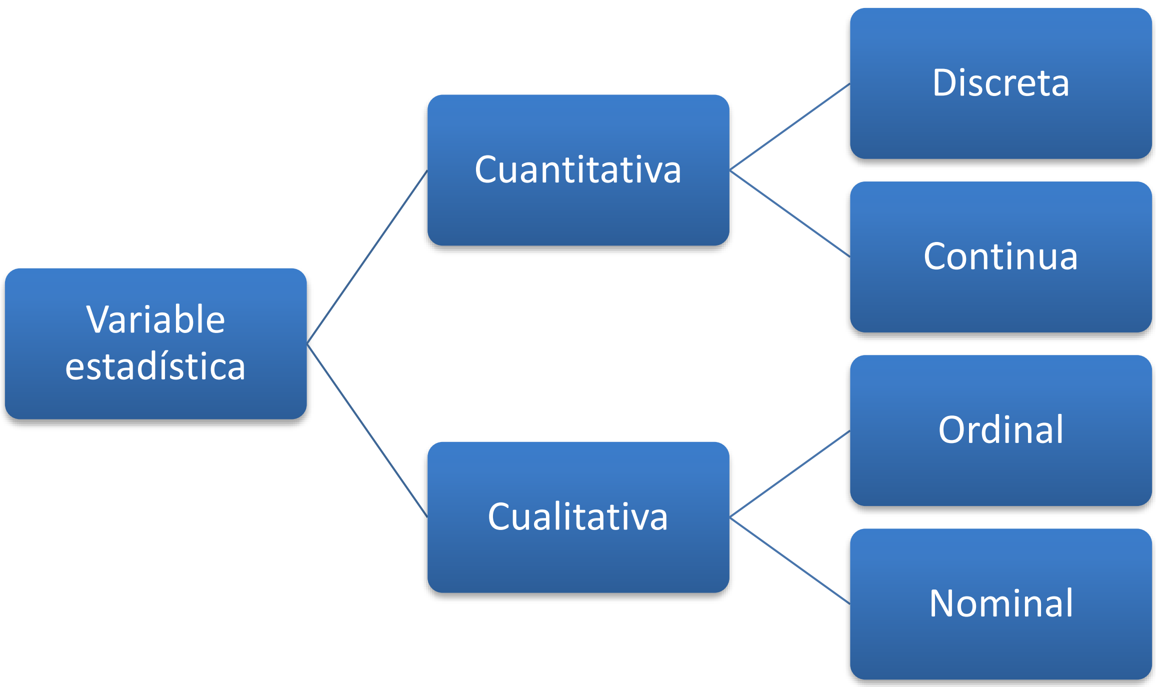
Desventajas de las muestras:

- El procedimiento de muestreo es inaplicable en aquellos casos en que se requieren datos de todos y cada uno de los integrantes de una población, como es el caso de las cuentas bancarias o las Fichas de Protección Social.

- Necesidad de personas expertas en los procesos de muestreo.

1.4 Variables estadísticas:

Una variable, cuyo símbolo suele ser *x* o *y*, es una característica medible (puede tomar un conjunto de valores) y observable asociada a elementos de una población o muestra. Los diferentes valores que asumen las variables se denominan *datos estadísticos*. El conjunto de donde la variable toma sus valores se denomina *dominio de la variable*. La variables se clasifican en dos tipos: Cuantitativas y cualitativas.



1.4.1 Variables cuantitativas:

Son las que toman valores numéricos. Son de dos tipos:

1.4.1.1 **Variables cuantitativas discretas**: Cuando solo se miden por medio de números enteros, es decir, sin decimales. En otras palabras, son variables que solo pueden tomar algunos valores en un determinado intervalo. Por ejemplo, el número de hermanos o el tamaño de la familia, la cantidad de establecimientos educacionales en una comuna, la cantidad de sucursales de una entidad bancaria, el número de accidentes de tránsito por día, la cantidad de viajes de un autobús en un mes, la cantidad de libros en una biblioteca, el número de electrones en un átomo, etc.

1.4.1.2 **Variables cuantitativas continuas**: Asumen valores intermedios entre dos números enteros, es decir, admiten decimales. En otras palabras, son variables que pueden tomar cualquier valor en un cierto intervalo. Por ejemplo, el peso de las personas, el tiempo empleado por un corredor de los 100 metros planos, la temperatura, los logros en una prueba, la altura de los individuos, la edad de las personas, la cantidad de lluvia caída en Santiago, volumen de agua en un embalse, la velocidad de un automóvil, etc.

1.4.2 Variables cualitativas o atributos:

Son las que señalan una cualidad del individuo, no es medible con números y su valor es una característica del individuo. Por ejemplo, un color (verde, rojo, blanco, etc.); la profesión del individuo (profesor, abogado, ingeniero, etc.); el sexo (masculino, femenino); el estado civil (casado, soltero, etc.). Son de dos tipos:

1.4.2.1 Variables cualitativas ordinales:

Son aquellas susceptibles de ordenación siguiendo una escala establecida y permiten establecer comparaciones entre categorías. Por ejemplo, nivel de estudios (básicos, medios, técnico-profesional, superior), nivel de desempeño profesional (insuficiente, aceptable, sobresaliente), desempeño de un estudiante en una tarea (deficiente, insuficiente, aceptable, bueno, excelente), nivel socio-económico (alto, medio, bajo), grado de cultura de una persona (muy culta, más o menos culta, poco culta, inculta), intensidad del dolor en una enfermedad (leve, moderado, fuerte), puesto ocupado en una actividad deportiva (1º, 2º, 3º), etc.

1.4.2.2 Variables cualitativas nominales:

Son aquellas que no son susceptible de ordenación, salvo el orden alfabético. Por ejemplo, el estado civil (casado, soltero, divorciado, viudo, separado), el sexo (masculino, femenino), nacionalidad (chilena, extranjera), los colores de las flores (rojo, blanco, violeta), lugar de nacimiento, etc.

1.5 Encuesta:

Es un conjunto de preguntas tipificadas dirigidas a una población o muestra con el propósito de conocer estados de opinión, determinadas características o hechos específicos.

Ejercicio: (del libro PSU matemática, Proyecto Clave)

Variable	Tipo de variable
Equipo de fútbol favorito	Cualitativa nominal
Número de trabajadores de una empresa	Cuantitativa discreta
Número de peces en un acuario	Cuantitativa discreta
Estatura de los integrantes de un equipo de fútbol	Cuantitativa continua
Color preferido	Cualitativa nominal

Nacionalidad de una persona	Cualitativa nominal
Tipo de animal en una tienda de mascotas	Cualitativa nominal
Número de estudiantes en una sala de clases	Cuantitativa discreta
Ingresos anuales de una empresa	Cuantitativa continua
Curso al que pertenece un estudiante	Cualitativa ordinal
Presión arterial	Cuantitativa continua
Cantidad de personas que se encuentran en una fila	Cuantitativa discreta
Metros cuadrados de un terreno	Cuantitativa continua
Estado civil de una persona	Cualitativa nominal
Número de habitantes de una ciudad	Cuantitativa discreta
Color de pelo	Cualitativa nominal
Promedio de notas del primer semestre	Cuantitativa continua
Masa corporal de una persona	Cuantitativa continua
IPC en los últimos cinco años	Cuantitativa continua
Profesión de una persona	Cualitativa nominal
Artículos más vendidos en un supermercado	Cualitativa nominal
Número de escritorios en una oficina	Cuantitativa discreta
Grupo de sangre de una persona	Cualitativa nominal
Estatura de los estudiantes de un curso	Cuantitativa continua
Deporte preferido por los estudiantes de un curso	Cualitativa nominal
Tiempo en recorrer dos kilómetros	Cuantitativa continua
Puntaje obtenido en la PSU	Cuantitativa discreta
Capacidad de una piscina	Cuantitativa continua
Nivel de inglés de un estudiante (básico, medio, avanzado)	Cualitativa ordinal
Diámetro de las ruedas de un vehículo	Cuantitativa continua
Asignatura aprobada durante el semestre	Cualitativa nominal
Valor del dólar	Cuantitativa continua
Lugar de llegada en una competencia	Cualitativa ordinal
Área de un rectángulo	Cuantitativa continua
Sueldo mensual de los trabajadores de una empresa	Cuantitativa continua
Marca de automóvil preferido	Cualitativa nominal
Edad de una persona	Cuantitativa continua
Día del mes en que cada persona está de cumpleaños	Cuantitativa discreta

Ejemplos de diferenciación entre población y muestra y entre tipos de variables

Ejemplo 1: Se quiere determinar el peso de los estudiantes universitarios de la ciudad de Punta Arenas:

La población: Todos los estudiantes universitarios de la ciudad de Punta Arenas.
La muestra: Los estudiantes universitarios que cursan el primer año de universidad.
La variable: El peso de los estudiantes en kilogramos. Es una variable cuantitativa, porque se puede medir, y es continua, porque la medición del peso admite decimales.

Ejemplo 2: Se quiere determinar los gustos musicales de los jóvenes de la ciudad de Curicó:

La población: Todos los jóvenes de la ciudad de Curicó.
La muestra: Los jóvenes entre 17 y 20 años.
La variable: El género musical (Rock, pop, clásica, folclórica, etc.).

La preferencia musical es una variable cualitativa, porque no es medible ni contable. Es cualitativa nominal, es decir, no ordenable.

2 RECOPILACIÓN DE LOS DATOS

Se entiende por *medida* el proceso de asignar números a objetos, fenómenos o características mediante reglas de cuantificación. Recopilar datos de una población o muestra es, entonces, hacer las mediciones necesarias para conocer los valores que asumen las variables consideradas en un estudio.

Los datos pueden obtenerse de las siguientes fuentes:

- Datos originales: Son los que se recopilan específicamente para un determinado estudio. Entre las técnicas más utilizadas para estos efectos están las encuestas, los cuestionarios y las entrevistas.

- Fuentes primarias: Son datos publicados por la misma entidad que los recopiló. Por ejemplo, el Instituto Nacional de Estadísticas (INE) es fuente primaria de los datos del censo poblacional. Organismos internacionales como la Organización de las Naciones Unidas (ONU), o la UNESCO, también son fuente de datos.

- Fuentes secundarias: Datos publicados por un organismo distinto a quien los recopiló originalmente. Por ejemplo, el Diario El Mercurio es fuente secundaria de datos publicados sobre diversas materias en su cuerpo Economía y Negocios.

3 ORGANIZACIÓN DE LOS DATOS

3.1 Proceso de organización de los datos:

Una vez recopilados los datos, éstos deben organizarse, proceso que comprende:

- Corrección de los datos: Consiste en la revisión de los datos para evitar omisiones, inconsistencias y errores.

- Clasificación de los datos: Consiste en la elección de categorías de acuerdo con criterios de tiempo, lugar, cantidad y cualidad. Los criterios más utilizados son los siguientes:

- Intervalos de tiempo: Siglos, decenios, años, meses, etc.
- De lugar: Continentes, países, regiones, ciudades, etc.
- De cantidad: Por ejemplo, de 0 a 20 Kg; de 21 a 30 Kg; de 31 a 40 Kg; de 41 a 51 Kg, etc.
- De cualidad: Colegios municipales, particulares subvencionados, particulares pagados, etc.

- Tabulación de los datos: Consiste en numerar los datos semejantes y registrarlos de acuerdo con las clasificaciones que se hayan definido.

3.2 Escalas de medición de variables

Toda medida de una variable está referida a una escala, que puede ser: nominal, ordinal, intervalar y de razón.

3.2.1 Escala nominal:

Las observaciones del atributo de la variable son clasificadas en categorías o clases y no son ordenables. Nombran, pero no miden la variable.

Por ejemplo:

- Estado civil (casado, soltero, divorciado, viudo, separado)
- Sexo (masculino, femenino)
- Nacionalidad (chilena, extranjera)
- Categorías de profesores universitarios (catedráticos, profesores asociados, profesores ayudantes)
- Estado de salud (sanos, enfermos)
- Color de los ojos
- Marca de automóvil
- Afiliación religiosa (cristianismo, judaísmo, islamismo, etc.)

3.2.2 Escala ordinal:

Se pueden establecer relaciones de orden entre los datos de la variable: mayor que, menor que, igual a.

Por ejemplo:

- Nivel de estudios (básicos, medios, técnico-profesional, superior)
- Nivel de desempeño profesional (insuficiente, aceptable, sobresaliente).
- Calificación de la tarea de un estudiante (deficiente, insuficiente, aceptable, bueno, excelente).
- Nivel socioeconómico (alto, medio, bajo),
- Estructura jerárquica en una organización (presidente, vicepresidente, gerente general, subgerente, etc.)

- Grado militar (general, coronel, teniente coronel, mayor, etc.)
- Etapas de desarrollo de un ser vivo
- Altura de los estudiantes de un curso

Ejemplo:

Alumnos	Altura en cm	Diferencia entre alumnos en cm	Rango (medida ordinal)
A	180		1º
B	178	2	2º
C	173	5	3º
D	168	5	4º
E	166	2	5º

3.2.3 Escala de intervalos:

Ordenan las medidas y permiten realizar comparaciones entre dos medidas, es decir, en cuanto más o en cuánto menos está presente una determinada característica o propiedad. Usan un cero relativo o arbitrario.

Por ejemplo:

- La escala de termómetros (usa un cero arbitrario, ya que el cero no significa ausencia de temperatura)
- Sobrepeso de las personas
- Coeficiente intelectual
- Altura de las montañas

3.2.4 Escala de razón:

Posee las características de una escala de intervalos y, además, posee un cero absoluto o verdadero; este último es un indicador de ausencia de la variable. Se pueden establecer razones, es decir, comparar mediciones a través de un cociente entre los datos.

Por ejemplo:

- El peso, la altura o el ingreso de las personas.
- Cantidad de litros de agua consumidos diariamente por una persona.
- Calificaciones en un examen.
- Cantidad de alumnos por establecimiento educacional.
- Cantidad de goles convertidos por un jugador en un partido.
- Utilidades de las empresas en un cierto período.

3.3 Distribución de frecuencias

Una tabla de distribución de frecuencias es un resumen de los datos originales recopilados.

Un concepto previo: Se denomina *recorrido o rango* de una variable a la diferencia entre el mayor y el menor valor que toma esa variable. Dependiendo del número de observaciones y del recorrido de la variable, las tablas de distribución de frecuencias se clasifican en:

3.3.1 Tablas de frecuencia de tipo I:

Cuando el tamaño de la población o de la muestra es pequeño. Por ejemplo, la altura en centímetros de 5 personas: 160, 180, 175, 170, 182. La tabla se ordena de forma creciente o decreciente.

Persona	Altura en cm
A	160
B	170
C	175
D	180
F	182

3.3.2 Tablas de frecuencia de tipo II:

Cuando el tamaño de la población o de la muestra es grande y el recorrido de la variable es pequeño. Por ejemplo, el valor de las notas de un estudiante durante un año. Sea un total de 30 notas en el año, con un recorrido de la variable de 1 a 7. Las notas podrían ser:

4	5	5	4	6	6
3	5	4	6	4	5
6	6	5	5	4	7
7	5	2	3	4	6
1	5	5	4	5	5

Para construir la tabla de frecuencia:

Primero: se ordenan los datos (las notas) de forma creciente o decreciente, y se realiza el conteo correspondiente, como sigue:

Notas	Conteo
1	I
2	I
3	II
4	IIIII II
5	IIIII IIIII I
6	IIIII I
7	II

Segundo: Se construye la tabla de frecuencia reemplazando el conteo por un número:

Notas	Frecuencia
1	1
2	1
3	2
4	7
5	11
6	6
7	2
Total	30

3.3.3 Tablas de frecuencia de tipo III (tabla de intervalos de clase):

Cuando tanto el tamaño de la población o muestra, como el recorrido de la variable, es grande.

3.4 Construcción de una tabla de distribución de frecuencias de tipo III (tabla de intervalos de clase)

Conceptos previos:

Frecuencia absoluta (f_i): Es el número de veces que se repite un dato en un intervalo o clase.

Frecuencia absoluta total (N): Es la suma de las frecuencias absolutas de cada uno de los valores de la variable.

Frecuencia absoluta acumulada (F_i): Es la suma de las frecuencias absolutas de los valores menores o iguales al valor de la variable en cuestión.

Por ejemplo, si las notas de un alumno en Ciencias Sociales son: 7, 5, 6, 5, 5, 6, 4, 7, la frecuencia absoluta de la nota 7 es 2, ya que dicha nota se repite dos veces, la del 5 es 3, ya que se repite tres veces, de la nota 6, es 2 y de la nota 4 es 1. Esta información se muestra en una tabla de frecuencia como sigue:

Notas del alumno	Frecuencia absoluta ($f_{i.}$)	Frecuencia absoluta acumulada ($F_{i.}$)
4	1	1
5	3	4
6	2	6
7	2	8
Frecuencia absoluta total (N)	8	

Frecuencia relativa (h_i): Es el cociente entre la frecuencia absoluta y la frecuencia absoluta total, es decir, entre la frecuencia absoluta y el tamaño de la población o muestra.

Frecuencia relativa acumulada (H_i): Es la suma de las frecuencias relativas de los valores menores o iguales al valor de la variable en cuestión.

Notas del alumno	Frecuencia absoluta ($f_{i\cdot}$)	Frecuencia relativa (h_i)	Frecuencia relativa acumulada ($H_{i\cdot}$)
4	1	$1/8 = 0{,}125$	$0{,}125$
5	3	$3/8 = 0{,}375$	$0{,}125 + 0{,}375 = 0{,}500$
6	2	$2/8 = 0{,}250$	$0{,}500 + 0{,}250 = 0{,}750$
7	2	$2/8 = 0{,}250$	$0{,}750 + 0{,}250 = 1{,}000$
Total	8	$1{,}000$	

Nótese que la frecuencias están expresadas en tanto por uno, por lo que la suma de todas ellas es igual a 1. Si se hubieran expresado en porcentajes, la suma sería 100%.

<u>Ejercicio 1</u>:

Construir una tabla de distribución de frecuencias en intervalos de clase con las estaturas de 80 alumnos de un colegio.

Valores de la variable estatura de los 80 alumnos (cm):

150	152	153	153	154	155	156	157
150	152	153	154	154	155	156	157
150	152	153	154	154	155	156	157
152	153	153	154	155	155	156	157
152	153	153	154	155	155	157	158
158	161	161	162	165	166	168	170
158	161	162	162	165	166	168	170
159	161	162	162	165	166	168	172
159	161	162	163	165	167	169	172
160	161	162	163	166	167	169	174

<u>Primer paso</u>: Calcular el rango o recorrido de los valores de la variable estatura:

Rango o recorrido (R) = Mayor valor – Menor valor

R = 174 – 150 = 24

<u>Segundo paso</u>: Establecer el número de intervalos de clase. Se acostumbra a utilizar un número cercano a 10, por ejemplo, 8 intervalos. Alternativamente, se puede determinar la cantidad de intervalos extrayendo la raíz cuadrada del total de las observaciones, que en este

ejemplo es 80, y si el valor no es exacto se aproxima al entero más próximo. Por lo tanto, $\sqrt{80} = 8{,}9$ que se aproxima a 9 intervalos.

<u>Tercer paso</u>: Calcular el tamaño o anchura del intervalo considerando 8 intervalos.

$$Tamaño\ de\ intervalo = \frac{Recorrido}{número\ de\ intervalos}$$

$$Tamaño\ de\ intervalo = \frac{24}{8} = 3$$

Si el tamaño calculado no es un número entero, se debe aproximarlo al entero más próximo. Es conveniente que todos los intervalos tengan el mismo tamaño, de manera que cualquier diferencia que se produzca al hacer la distribución debe agregarse al último intervalo, el que sí tendrá un tamaño algo diferente.

<u>Cuarto paso</u>: Determinar el límite inferior y superior de cada intervalo de clase:

El límite inferior del primer intervalo: Es el menor valor de la variable, es decir, 150.

El límite inferior del segundo intervalo: Es el límite inferior del primer intervalo más el tamaño del intervalo, es decir, 150 + 3 = 153.

El límite inferior del tercer intervalo: Es límite inferior del segundo intervalo más el tamaño del intervalo, es decir, 153 + 3 = 156. Y así sucesivamente, hasta completar 8 valores.

Estatura (x)
150
153
156
159
162
165
168
171
174

El límite superior de cada intervalo se obtiene sumando al límite inferior el tamaño del intervalo, que en este ejemplo es 3:

Estatura (x)
150 - 153
153 - 156
156 -159
159 -162
162 - 165
165 - 168

168 - 171
171 - 174

Los intervalos de clase así determinados se denominan cerrados por la izquierda – abiertos por la derecha. Esto quiere decir que un dato que es igual al valor del límite superior de un intervalo, corresponde considerarlo en el intervalo siguiente. Así, por ejemplo, los alumnos que midan 153 cm no deben incluirse en el intervalo 150 – 153, sino en el intervalo 153 – 156. De igual manera, los que midan 156 cm se deben clasificar en el intervalo 156 – 159 y no en el intervalo 153 - 156, etc.

En ocasiones, los intervalos se presentan con corchetes como sigue:

[150 – 153[o bien [150 – 153) intervalo cerrado por la izquierda y abierto por la derecha
[153 – 156[o bien [153 – 156)

<u>Quinto paso</u>: Determinar las marcas de clase (x_i), que corresponden al valor medio de cada intervalo de clase, es decir, es el valor representativo de cada intervalo.

Se calcula así:

$$Marca\ de\ clase = \frac{límite\ inferior + límite\ superior}{2}$$

De esta manera, la marca de clase del primer intervalo será:

$$Marca\ de\ clase = \frac{150 + 153}{2} = 151{,}5$$

La marca de clase del segundo intervalo será:

$$Marca\ de\ clase = \frac{153 + 156}{2} = 154{,}5$$

Estatura (x)	Marca de clase (x_i)
150 - 153	151,5
153 - 156	154,5
156 -159	157,5
159 -162	160,5
162 - 165	163,5
165 - 168	166,5
168 - 171	169,5
171 - 174	172,5

<u>Sexto paso</u>: Calcular las frecuencias absolutas y relativas y sus correspondientes valores acumulados, de acuerdo con el procedimiento señalado anteriormente. La frecuencia absoluta de cada intervalo es el número de valores que contiene ese intervalo y la frecuencia relativa es el cociente de ese valor absoluto y el total de valores ($n = 80$ en el ejemplo).

De esta manera, la tabla de distribución solicitada para este ejercicio queda como sigue:

Estatura en cm (x)	Marca de clase (xi)	fi	Fi	hi	Hi
150 - 153	151,5	8	8	0,1000	0,1000
153 - 156	154,5	22	30	0,2750	0,3750
156 -159	157,5	12	42	0,1500	0,5250
159 -162	160,5	9	51	0,1125	0,6375
162 - 165	163,5	9	60	0,1125	0,7500
165 - 168	166,5	10	70	0,1250	0,8750
168 - 171	169,5	7	77	0,0875	0,9625
171 - 174	172,5	3	80	0,0375	1,0000
Total		80		1,000	

Como se aprecia en la tabla anterior, la suma de las frecuencias relativa es igual a 1. Los valores de las frecuencias relativas (h_i) y frecuencias relativas acumuladas (H_i) se pueden expresar en porcentaje multiplicando los respectivos valores por 100. Por ejemplo, $0{,}1000\ x\ 100\ =\ 10\%;\ 0{,}2750\ x\ 100 = 27{,}5\%;\ 1\ x\ 100 = 100\%$.

Interpretación de la tabla de frecuencias

Marcas de clase x_i: 151,5 es un valor representativo del intervalo 150-153, es decir, la estatura media de estudiantes ubicados en ese intervalo es de 151,5 cm. De igual manera, 154,5 es un valor representativo del intervalo 153-156, es decir, la estatura media de estudiantes ubicados en ese intervalo es de 154,5 cm, etc.

Frecuencia absoluta f_i: Hay 8 alumnos que miden entre 150 y 153 cm., sin considerar a los que miden exactamente 153 cm que se ubican en el intervalo siguiente (153-156); asimismo, hay 22 alumnos que miden entre 153 y 156 cm, sin considerar a los que miden exactamente 156 cm que se ubican en el intervalo siguiente (156-159), etc.

Frecuencia absoluta acumulada H_i: Por ejemplo, hay 30 estudiantes cuyas alturas son menores que 156 cm (límite superior del intervalo 153-156), o bien, que hay 30 estudiantes cuya estatura es mayor o igual que 150 cm, pero menor a 156 cm; Otro ejemplo: hay 70 estudiantes que miden menos de 168 cm. (límite superior del intervalo 165-168), o bien, que hay 70 estudiantes cuya estatura es mayor o igual que 150 cm, pero menor a 168 cm., etc.

Frecuencia relativa h_i: Por ejemplo, el 10% (0,1000 x 100) de los alumnos miden entre 150 y 153 cm., sin considerar a los que miden exactamente 153 cm que se ubican en el intervalo siguiente (153-156); En otras palabras, los 8 alumnos que miden entre 150 y 153 cm en relación con el total de alumnos, que es 80, representan el 10% de ese total; Otro ejemplo: el 27,5% (0,2750 x 100) de los alumnos miden entre 153 y 156 cm., sin considerar a los que miden exactamente 156 cm que se ubican en el intervalo siguiente (156-159); En otras palabras, los 22 alumnos que miden entre 153 y 156 cm en relación con el total de alumnos, que es 80, representan el 27,5% de ese total, etc.

Frecuencia relativa acumulada H_i: El 37,5% (0,375 x 100) de los estudiantes miden menos que 156 cm (límite superior del intervalo 153-156). En otras palabras, los 30 alumnos que miden menos de 156 cm. representan el 37,5% del total de alumnos. O bien, el 37,5% de los estudiantes tienen una estatura que es mayor o igual que 150, pero menor que 156cm.; de igual manera, el 87,5% (0,875 x 100) de los estudiantes miden menos que 168 cm (límite superior del intervalo 165-168). En otras palabras, los 70 alumnos que miden menos de 168 cm. representan el 87,5% del total de alumnos. O bien, 87,5% de los estudiantes tienen una estatura que es mayor o igual que 150 cm, pero menor que 168 cm.

Ejercicio 2:

Confeccionar una tabla de distribución de frecuencias tipo II considerando las siguientes temperaturas registradas en un mes de diciembre en Santiago de Chile.

Día	$T°$
1	32
2	31
3	28
4	29
5	33
6	32
7	31
8	30
9	31
10	31
11	27
12	28
13	29
14	30
15	32

Día	$T°$	Día	$T°$
16	31	31	29
17	31		
18	30		
19	30		
20	29		
21	29		
22	30		
23	30		
24	31		
25	30		
26	31		
27	34		
28	33		
29	33		
30	29		

Grados Celsius (x_i)	Recuento	fi	Fi	h_i	Hi
27	I	1	1	3,2%	3,2%
28	II	2	3	6,5%	9,7%
29	IIIII I	6	9	19,4%	29,0%
30	IIIII II	7	16	22,6%	51,6%
31	IIIII III	8	24	25,8%	77,4%
32	III	3	27	9,7%	87,1%
33	III	3	30	9,7%	96,8%

34	I	1	31	3,2%	100,0%
	n=	31		100%	

Ejercicio 3: Para ejercitación:

- Si en una distribución F_5 representa la frecuencia acumulada correspondiente al quinto intervalo, ¿qué significa $F_5 = 60$?

- Si H_5 representa la frecuencia relativa acumulada correspondiente al quinto intervalo, ¿qué significa $H_5 = 0{,}750$?

- Si h_4 representa la frecuencia relativa correspondiente al cuarto intervalo, ¿qué significa h_4=15,25%?

- Si H_4 representa la frecuencia relativa acumulada correspondiente al cuarto intervalo, ¿qué significa H_4=0,6375?

3.5 Representación gráfica de una tabla de distribución de frecuencias de variable continua (datos agrupados)

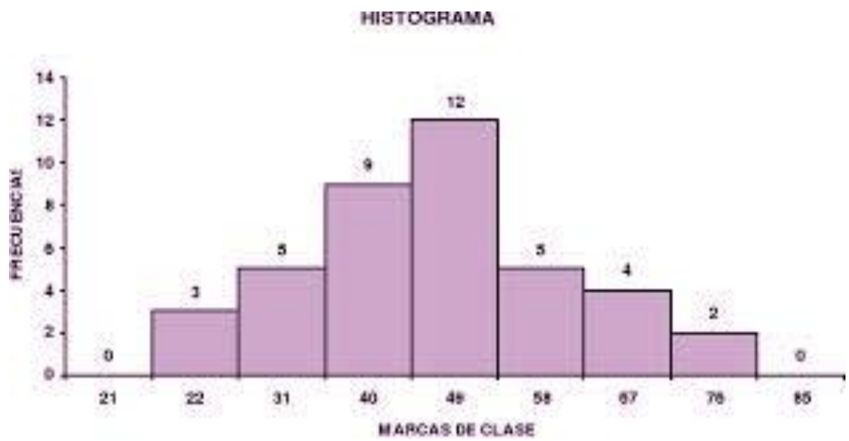
Los gráficos hacen posible tener una visión clara del comportamiento de la variable estudiada, por ejemplo, de la estatura en cm de un grupo de estudiantes. Existen varios tipos de gráficos, entre ellos: histogramas, polígonos de frecuencias, circulares o de torta y pictogramas.

3.5.1 Histograma:

Es un gráfico de barras que consiste en un conjunto de rectángulos verticales. Se usa frecuentemente con valores de variables cuantitativas continuas, agrupados en intervalos. Posee las siguientes características:

a. La longitud de la base de cada rectángulo en el eje X del plano cartesiano, es igual a la amplitud del intervalo. Para dibujar dicha longitud, se parte marcando en dicho eje las distintas marcas de clase; luego, se restan del valor de cada marca las unidades necesarias para señalar en el eje el límite inferior del intervalo y se adicionan las que correspondan para marcar el límite superior correspondiente. Uniendo ambos puntos se obtiene y dibuja la base de cada rectángulo.

b. Si los intervalos de clase son todos de igual tamaño, las alturas de cada rectángulo corresponden a las frecuencias absolutas de cada intervalo (en el eje y del plano cartesiano). Si los intervalos de clase no son de igual tamaño, hay que calcular dichas alturas.

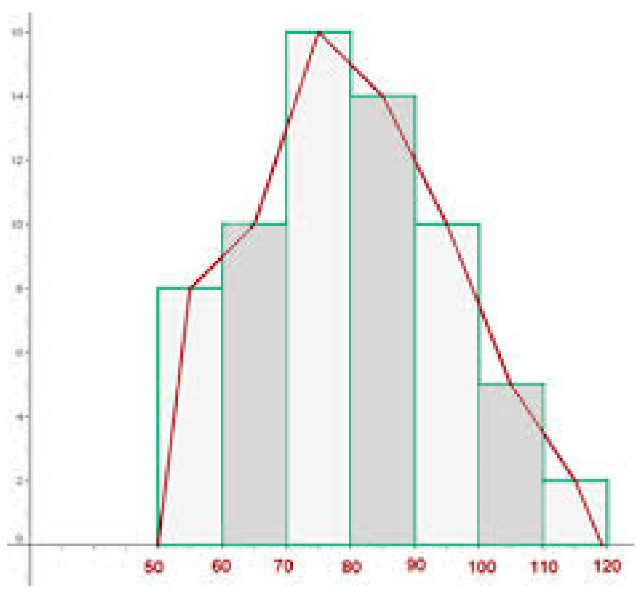


3.5.2 <u>Polígono de frecuencias</u>:

Es una representación gráfica que forma un polígono compuesto por la línea poligonal que se forma al unir los puntos referidos a las marcas de clase de cada intervalo y el eje de las abscisas. Se usa frecuentemente con valores de variables cuantitativas continuas. Son dos las posibilidades de graficarlo:

a. Uniendo los puntos medios de los techos de los rectángulos de un histograma.

b. Uniendo las marcas de clase de los intervalos con las frecuencias absolutas respectivas formando pares ordenados de puntos. Esta opción exige marcar en el eje x la marca de clase de un intervalo hipotético anterior al primero de ellos y la de un intervalo posterior al último. De esta forma, la poligonal de frecuencia se forma uniendo los pares ordenados y se cierra uniendo los extremos con las marcas de clase de los intervalos hipotéticos señalados.



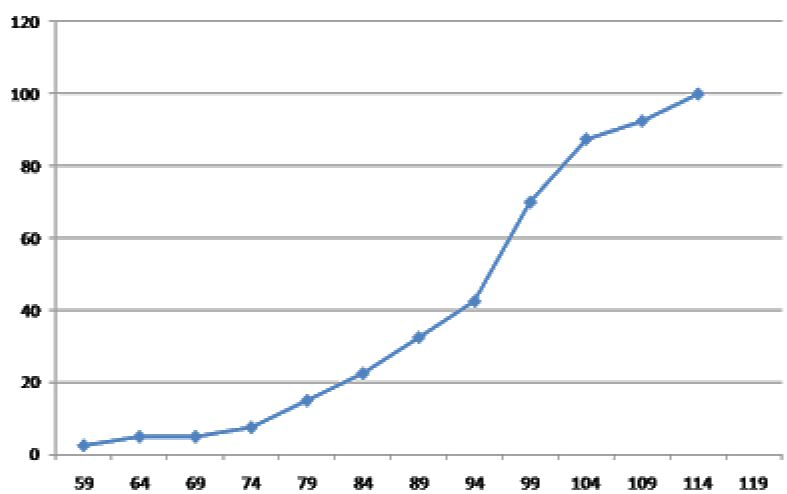
3.5.3 Polígono de frecuencias acumuladas u Ojiva

Es un gráfico que muestra las frecuencias que se encuentran por encima o por debajo de ciertos valores, en vez de exhibir los números asignados a cada intervalo.

Supóngase que las alturas en centímetros de 40 estudiantes se agruparon en una tabla de frecuencias como la siguiente:

Intervalos	fi	Fi	h_i	Hi
60-64	1	1	2,5	2,5
65-69	1	2	2,5	5,0
70-74	0	2	0,0	5,0
75-79	1	3	2,5	7,5
80-84	3	6	7,5	15,0
85-89	3	9	7,5	22,5
90-94	4	13	10,0	32,5
95-99	4	17	10,0	42,5
100-104	11	28	27,5	70,0
105-109	7	35	17,5	87,5
110-114	2	37	5,0	92,5
115-119	3	40	7,5	100,0
Total	40		100,0	

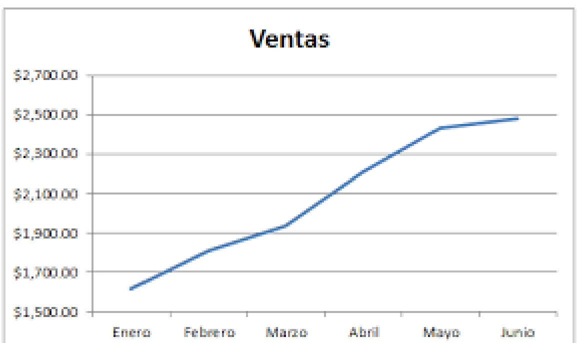
El polígono de frecuencias acumuladas, grafica la columna $Hi,$ que muestra, en este ejemplo, una ojiva denominada "menor que", ya que grafica las frecuencias acumuladas menores que cualquier límite real superior de los intervalos de clase, tiene la siguiente representación:

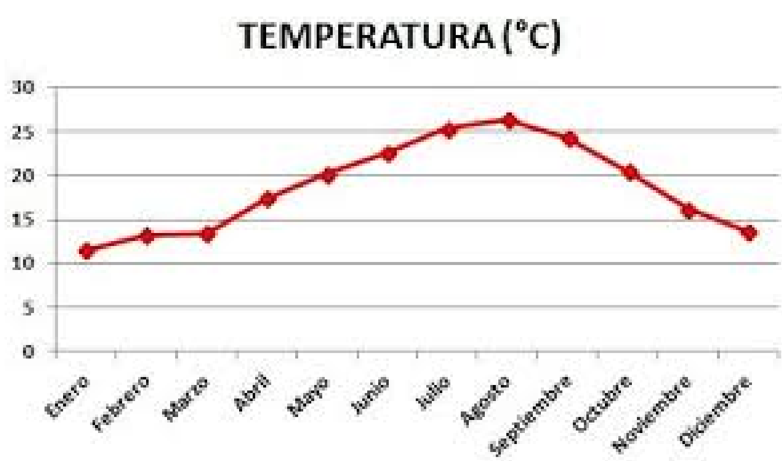


La ojiva puede ser del tipo "mayor que" "o más" y su representación es la siguiente:

3.5.4 <u>Gráfico de líneas</u>:

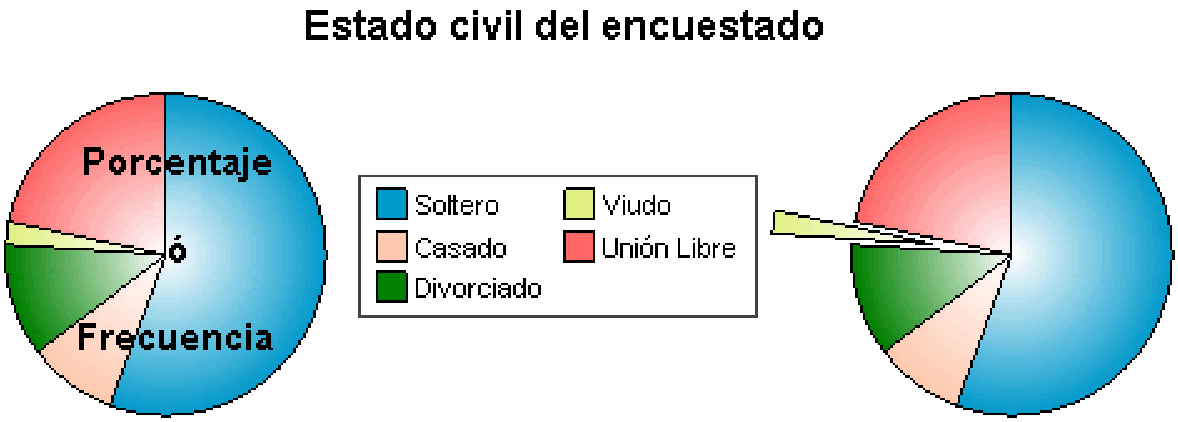
Es un gráfico formado por un conjunto de puntos unidos con líneas rectas. Por ejemplo, para reflejar las ventas mensuales de una empresa o la temperatura en grados Celsius de una determinada localidad, mes a mes durante el año.





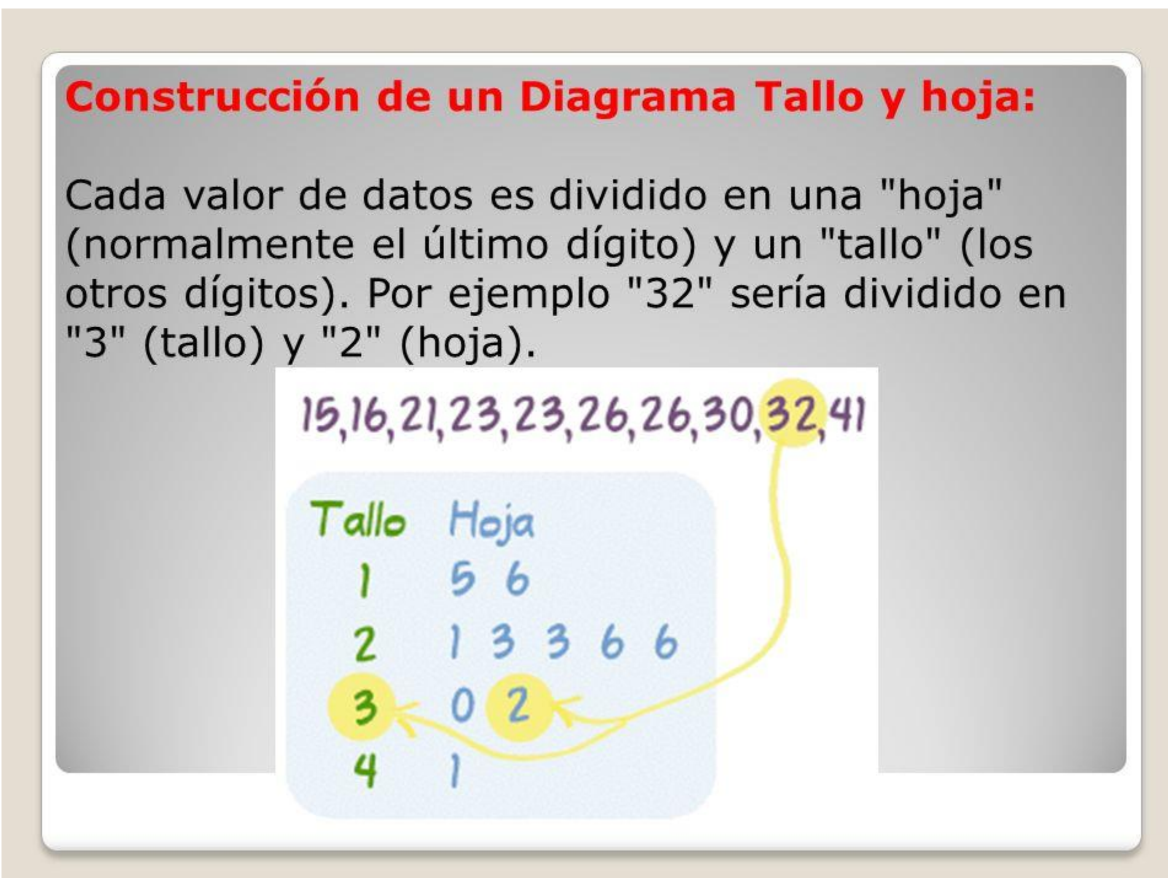
3.5.5 <u>Gráficos circulares o de torta</u>:

Es un gráfico que busca destacar la fracción que cada categoría de una variable representa del total. El ángulo del centro de la circunferencia correspondiente a cada sector circular es proporcional a la frecuencia relativa de cada valor de la variable. Se utilizan para representar tanto variables cualitativas como cuantitativas, como podría ser el estado civil de los informantes de un estudio.



3.5.6 Diagrama de tallo y hojas:

Cada número se divide en uno o más dígitos principales, que forman el tallo, y uno o más dígitos secundarios, que constituyen las hojas



3.6 Representación gráfica de una tabla de distribución de frecuencias de variable discreta (datos no agrupados)

La frecuencia correspondiente a cada valor de la variable se representa por una barra vertical.

3.7 Ejercicios propuestos:

1) Los 50 alumnos de un colegio que rindieron la Prueba de Selección Universitaria (PSU) en el 2013 obtuvieron los siguientes puntajes:

567	732	567	534	697	635	752	801	789	456
603	645	801	509	496	745	659	599	523	801
686	743	456	489	596	643	489	499	801	785
487	725	667	654	578	598	496	726	647	658
501	652	501	427	647	543	589	599	473	499

Se pide:

a) Construir una tabla de distribución de frecuencias tipo III (intervalar), de amplitud 50.

b) ¿Qué significa f_4?, ¿Qué significa F_6?

2) La siguiente tabla muestra la distribución de los puntajes obtenidos por los alumnos de un curso, en una prueba. ¿Qué porcentaje de los alumnos del curso obtuvo menos de 12 puntos en la prueba?

Puntaje	Cantidad de alumnos
0-5	3
6-11	3
12-17	5
18-23	15
24-29	4

Alternativas: A) 80 %, B) 20 %; C) 30,5 %; D) 35 %; E) ninguna de las anteriores.

3) Se aplica una prueba a un grupo de 12 alumnos de 4° medio. Los resultados contenidos están representados en la tabla de frecuencia que se muestra a continuación. ¿Qué porcentaje de los alumnos obtuvo nota mayor o igual a 4?
Nota: [1,2) es la notación para un intervalo cerrado por la izquierda y abierto por

la derecha. Ídem para el resto de los intervalos, excepto el último, que es cerrado tanto por la izquierda como por la derecha.

Nota	Frecuencia absoluta
[1,2)	0
[2,3)	1
[3,4)	2
[4,5)	3
[5,6)	4
[6,7]	2
Total	12

Alternativas: A) 25%; B) 45%; C) 75%; D) 55%; E) Ninguna de las anteriores.

4) En la siguiente tabla se muestra la distribución de los puntajes obtenidos en una prueba de Estadística:

Nota	Frecuencia absoluta
10-20	5
20-30	7
30-40	15
40-50	5
50-60	8

a) ¿Cuántos alumnos rindieron la prueba?

b) ¿Cuántos alumnos obtuvieron un puntaje menor que 40 puntos?

c) ¿Qué porcentaje de alumnos obtuvo un puntaje mayor o igual que 20 pero menor que 40 puntos?

5) Los siguientes valores corresponden a una encuesta a 48 profesores de un establecimiento educacional para conocer el número de familiares que están o estuvieron matriculados en él.

3	1	5	2	4	2	4	6	5	1
3	2	3	2	4	3	3	3	1	4
2	1	1	2	1	4	4	3	3	2
1	0	1	2	3	3	3	2	2	2
3	3	3	1	1	2	4	1		

Se solicita construir una tabla de frecuencias y también:

a) Identificar el tipo de variable y de escala.
b) ¿Cuántas personas tienen 5 familiares?
c) ¿Cuántas personas tienen más de 2 familiares?
d) ¿Qué porcentaje de profesores tiene 4 familiares en el establecimiento?
e) ¿Qué porcentaje de profesores tiene 3 familiares en el establecimiento?
f) ¿Qué porcentaje de profesores tiene menos de 2 familiares en el establecimiento?
g) ¿Qué porcentaje de profesores tiene más de 3 familiares?
h) ¿Qué significa la frecuencia absoluta correspondiente al cuarto valor de la variable?

6) En la tabla siguiente se muestran los puntajes obtenidos en un curso de Química integrado por 33 alumnos:

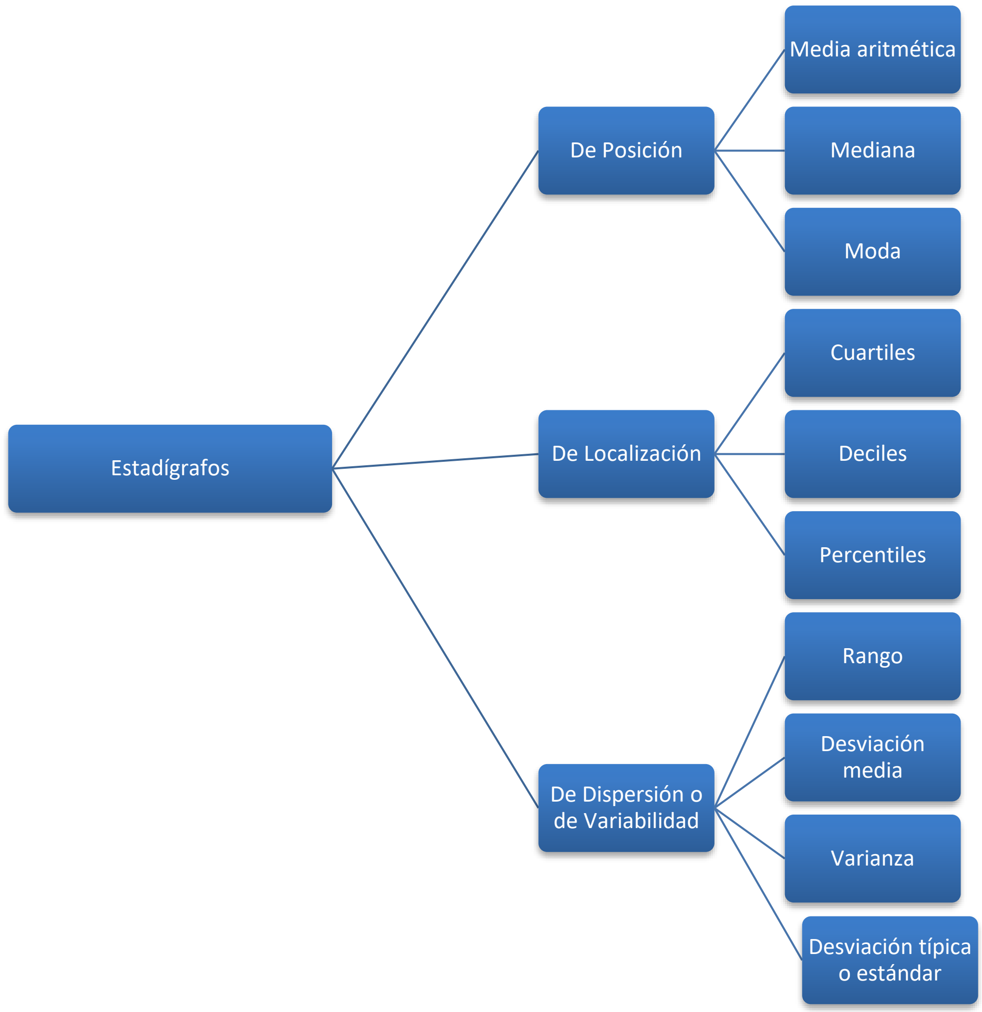
Puntaje (x)	Marca de clase (x_i)	f_i	F_i	h_i	H_i
13,5 - 21,5	17,5	1	1	3%	3%
21,5 - 29,5	25,5	2	3	6%	9%
29,5 - 37,5	33,5	2	5	6%	15%
37,5 - 45,5	41,5	5	10	15%	30%
45,5 - 53,5	49,5	6	16	19%	49%
53,5 - 61,5	57,5	4	20	12%	61%
61,5 - 69,5	65,5	0	20	0%	61%
69,5 - 77,5	73,5	4	24	12%	73%
77,5 - 85,5	81,5	7	31	21%	94%
85,5 - 93,5	89,5	2	33	6%	100%
Total		33		100%	

Responder:

a) ¿Cuál es la amplitud de cada intervalo?
b) ¿Cuál es el límite inferior del intervalo 6?
c) ¿Cuál es la marca de clase del 3º intervalo?
d) ¿Cuál es el límite superior del 5º intervalo?
e) ¿Cuál es el intervalo de mayor frecuencia absoluta?
f) ¿Cuál es el porcentaje de alumnos con puntaje inferior a 69,5?
g) ¿Cuál es el número de alumnos con puntaje igual o inferior a 61,5?
h) ¿Qué significan las distintas frecuencias correspondientes al 6º intervalo?

4 ESTADÍGRAFOS

Un estadígrafo es una expresión numérica que permite cuantificar un conjunto de datos de una variable provenientes de una población o de una muestra, previamente ordenados en tablas de frecuencias Se clasifican en estadígrafos de posición, de localización y de dispersión.



4.1 ESTADÍGRAFOS DE POSICIÓN O MEDIDAS DE TENDENCIA CENTRAL

Una medida de tendencia central es un número que se considera representativo de todos los números en un conjunto de datos. Representan valores centrales en torno de los cuales se agrupan las observaciones o puntuaciones. Los estadígrafos de posición son:

- La media aritmética
- La mediana
- La moda.

4.1.1 Media aritmética

Se escribe $\bar{x}$ o M_a. Es el cociente entre la suma de los valores de la variable y el total de datos (la frecuencia total). Se acostumbra a denominar media o promedio.

4.1.1.1 Media aritmética de datos no agrupados (tabla de frecuencia tipo I)

$$\boldsymbol{\bar{x}} = \frac{(\boldsymbol{x_1} + \boldsymbol{x_2} + \boldsymbol{x_3} + \cdots + \boldsymbol{x_n})}{\mathbf{n}}$$

o bien:

$$\boldsymbol{\bar{x}} = \frac{\sum_{i=1}^{n} \boldsymbol{x_i}}{\mathbf{n}}$$

donde el símbolo $\sum$ es la letra griega mayúscula sigma que significa suma o sumatoria.

Ejemplo: Calcular la estatura promedio de las cinco personas siguientes:

Persona	Altura en cm
A	160
B	170
C	175
D	180
F	182

$$\bar{x} = \frac{160 + 170 + 175 + 180 + 182}{5} = 173{,}4\ cm$$

4.1.1.2 Media aritmética de datos agrupados en tablas de frecuencia tipo II)

Es la suma de los valores de la variable multiplicados por sus respectivas frecuencia, dividida entre la frecuencia total. Nota: se asume que el asterisco significa multiplicado por.

$$\boldsymbol{\bar{x}} = \frac{\boldsymbol{f_1} \cdot \boldsymbol{x_1} + \boldsymbol{f_2} \cdot \boldsymbol{x_2} + \boldsymbol{f_3} \cdot \boldsymbol{x_3} \ldots + \boldsymbol{f_n} \cdot \boldsymbol{x_n}}{\boldsymbol{n}}$$

o bien:

$$\boldsymbol{\bar{x}} = \frac{\sum_{i=1}^{n} (\boldsymbol{x_i} \cdot \boldsymbol{f_i})}{\mathbf{n}}$$

donde $\boldsymbol{n} = \sum_{i=1}^{n} \boldsymbol{f_i}$

Ejemplo: Un estudiante obtuvo las siguientes 30 notas en un año. ¿Cuál es la nota media obtenida?

Notas	Frecuencia
1	1
2	1
3	2
4	7
5	11
6	6
7	2
Total n	30

$$\bar{x} = \frac{1 \cdot 1 + 2 \cdot 1 + 3 \cdot 2 + 4 \cdot 7 + 5 \cdot 11 + 6 \cdot 6 + 7 \cdot 2}{30} = \frac{142}{30} = 4{,}7$$

Alternativamente, los datos se pueden presentar de la siguiente manera:

Notas (x_i)	f_i	$x_i \cdot f_i$
1	1	$1 \cdot 1 = 1$
2	1	$2 \cdot 1 = 2$
3	2	$3 \cdot 2 = 6$
4	7	$4 \cdot 7 = 28$
5	11	$11 \cdot 5 = 55$
6	6	$6 \cdot 6 = 36$
7	2	$7 \cdot 2 = 14$
Total	30	142

$$\bar{x} = \frac{142}{30} = 4{,}7$$

Las frecuencias f_i $(f_1, f_2, f_3, \dots, f_k)$ se pueden asociar con factores o pesos w_i $(w_1, w_2, w_3, \dots, w_k)$.

En este caso se habla de una media aritmética ponderada:

$$\bar{x} = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 + \cdots + x_k \cdot w_k}{w_1 + w_2 + w_3 + \cdots + w_k}$$

Ejemplo: Si un estudiante tiene dos notas parciales de 70 y 90 puntos y una nota de prueba global cuyo coeficiente es 3, es decir, pondera tres veces el puntaje de las pruebas parciales, ¿Cuál es la nota promedio?

w_1= 1
w_2= 1
w_3= 3

$\bar{x} = \frac{[(70\cdot1+90\cdot1+85\cdot3)]}{1+1+3} = \frac{415}{5} = 83$

Ejemplo:
Andrea ha obtenido en matemática las siguientes notas: 5,6 – 7,0 – 6,2 – 6,3 – 6,3 – 4,8. ¿Cuál debe ser su última nota si quiere tener un promedio igual o superior a 6?

Sea x esa nota:

$$\frac{5,6\ +\ 7,0\ +\ 6,2\ +\ 6,3\ +\ 6,3\ +\ 4,8 + x}{7} = 6$$

$$\frac{36,2 + x}{7} = 6$$

$$36,2 + x = 43$$

$$x = 43 - 36,2 = 5,8$$

4.1.1.3 Media aritmética de datos agrupados en tablas de frecuencia tipo III (intervalar)

Es la suma de las marcas de clase de cada intervalo multiplicadas por sus respectivas frecuencia, dividida entre la frecuencia total.

$$\bar{\boldsymbol{x}} = \frac{\sum_{i=1}^{n}(\boldsymbol{x}_{mci} \cdot \boldsymbol{f}_i)}{\mathbf{n}}$$

donde $\boldsymbol{x}_{mci}$ es marca de clase (punto medio del intervalo)

Ejemplo: Calcular la estatura promedio de 80 estudiantes según la información presentada en la siguiente tabla de frecuencias:

Estatura en cm (x)	Marca de clase $\boldsymbol{x}_{mci}$	$\boldsymbol{f}_i$
150 - 153	151,5	8
153 - 156	154,5	22
156 -159	157,5	12
159 -162	160,5	9
162 - 165	163,5	9
165 - 168	166,5	10
168 - 171	169,5	7
171 - 174	172,5	3
Total		80

$$\bar{x} = \frac{8 \cdot 151{,}5 + 22 \cdot 154{,}5 + 12 \cdot 157{,}5 + 9 \cdot 160{,}5 + 9 \cdot 163{,}5 + 10 \cdot 166{,}5 + 7 \cdot 169{,}5 + 3 \cdot 172{,}5}{80}$$
$$= 159{,}8$$

Alternativamente, los datos se pueden presentar como sigue:

Estatura en cm (x)	Marca de clase (x_i)	f_i	$f_i \cdot x_i$
150 - 153	151,5	8	1.212,0
153 - 156	154,5	22	3.399,0
156 - 159	157,5	12	1.890,0
159 - 162	160,5	9	1.444,5
162 - 165	163,5	9	1.471,5
165 - 168	166,5	10	1.665,0
168 - 171	169,5	7	1.186,5
171 - 174	172,5	3	517,5
Total		80	12.786,0

$$\bar{x} = \frac{12.786}{80} = 159{,}8$$

<u>Ejemplo</u>: De 500 alumnos de un colegio, cuya estatura promedio es 1,67 m, 150 son mujeres. Si la estatura promedio de todas las mujeres es 1,60 m, ¿Cuál es la estatura promedio de los varones del colegio?
Fuente: Tapia, O. y Ormazábal M. (2012). Matemática PSU. Cuaderno de ejercicios. Ediciones UC., p. 402

Solución:

Los varones son $500 - 150 = 350$
Sea x la estatura de los estudiantes varones

$$\frac{[1{,}60(150) + x(350)]}{500} = 1{,}67$$

$$1{,}60(150) + x(350) = 1{,}67(500)$$
$$240 + 350x = 835$$
$$350x = 595$$

$$x = 1{,}70\ m$$

<u>Ejemplo</u>: El promedio de notas de una asignatura de un curso A de 20 alumnos es 6 y el del curso B, que también tiene 20 alumnos, es 5, ¿Cuál es el promedio de notas de la asignatura de los alumnos de ambos cursos?

Fuente: Tapia, O. y Ormazábal M. (2012). Matemática PSU. Cuaderno de ejercicios. Ediciones UC., p. 402

Solución:
Totalidad de alumnos de ambos cursos: 40

$$\frac{[6(20) + 5(20)]}{40} = \frac{120 + 100}{40} = \frac{220}{40} = 5{,}5$$

4.1.1.4 Inconveniente de la media aritmética:

Es un estadígrafo muy sensible a los valores extremos de la variable, es decir, a aquellos que distan mucho de los valores centrales de la distribución. Como no son representativos de ella, pueden hacer variar significativamente el promedio calculado.

Ejemplo: Supóngase que los sueldos de los operarios en una pequeña empresa son los siguientes:

Operarios	Sueldo en $
A	400.000
B	450.000
C	800.000
D	420.000
E	450.000

$$\bar{x} = \frac{400000 + 450000 + 800000 + 420000 + 450000}{5} = \frac{2520000}{5} = 504.000$$

Obviamente el sueldo promedio de $504.000 no es representativo de esta empresa, porque está distorsionado por el sueldo del operario C, que es de $800.000.

Cuando en la distribución se tienen valores extremos, se suele utilizar otro estadígrafo, la Mediana, que se pasa a describir a continuación.

4.1.2 Mediana

Se escribe $\boldsymbol{M_e}$. Es el valor central de una distribución una vez ordenados los datos de manera decreciente o creciente. Por encima de la mediana se encuentra el 50% de los valores de la distribución y por debajo de ella el otro 50%. Es una medida de tendencia central menos sensible que la media aritmética a los valores extremos de la variable.

4.1.2.1 Mediana ($\boldsymbol{M_e}$) de datos no agrupados (tabla de frecuencia de tipo I):

Se ordenan los datos en forma creciente o decreciente. Si la cantidad de datos es un número impar, la mediana será el valor central; si es un número par, la mediana será la media aritmética entre los dos valores centrales.

Ejemplo 1: Calcular la mediana de la estatura de las 5 personas que se muestran en la tabla siguiente: $B = 170cm$; $A = 160\ cm$; $D = 180\ cm$; $E = 182$ cm y $C = 175\ cm$.

Se ordenan de forma creciente como se ve en la tabla. La mediana es el valor central, es decir, 175 cm.

Persona	Altura en cm
A	160
B	170
C	175
D	180
E	182

Ejemplo 2: Calcular la mediana del conjunto de notas siguientes: 4-5-4-6-6-7-7-4.

Se ordenan las notas de manera creciente: $4 - 4 - 4 - 5 - 6 - 6 - 7 - 7$. Los valores centrales son 5 y 6, por lo tanto, la mediana es la media entre esos valores, es decir, $(5 + 6)/2 = 5{,}5$.

Ejemplo 3: El gasto en telefonía de ocho familias es: $40000 - 47000 - 60000 - 70000 - 78000 - 80000 - 90000 - 180000$. Calcular la mediana.

Los datos ya están ordenados de menor a mayor. Como la cantidad de observaciones es un número par, la mediana será la media aritmética entre 70000 y 78000, es decir, $(70000 + 78000)/2 = \$74.000$.

Nótese la poca sensibilidad de la mediana a los valores extremos, en este caso a $40.000 y $180.000. Si este último dato hubiese sido, $250.000, la mediana seguiría siendo $74.000.

Ejemplo 4: Supóngase que los sueldos en una pequeña empresa son los que se muestran en la tabla siguiente. Calcular la mediana.

Operarios	Sueldo en $
A	400.000
B	450.000
C	800.000
D	420.000
E	450.000

Ordenando los datos, se tiene:

Operarios	Sueldo en $
A	400.000
D	420.000
B	450.000
E	450.000
C	800.000

Como la cantidad de datos es impar, la mediana es $450.000. Este valor es, sin duda, más representativo de la media aritmética, que el valor de $504.000 calculado en un ejemplo anterior.

4.1.2.2 Mediana ($\boldsymbol{M_e}$) de datos no agrupados (tabla de frecuencia de tipo II)

La mediana corresponde al primer valor de la variable cuya frecuencia acumulada es igual o mayor a la mitad de la frecuencia total.

Ejemplo 5: Determinar la mediana en la siguiente distribución:

Notas del alumno	Frecuencia absoluta ($f_{i\cdot}$)
3	1
4	5
5	3
6	2
7	1
Frecuencia total (n)	12

Paso 1: Se calcula la mitad de la frecuencia total, es decir: $\frac{n}{2} = \frac{12}{2} = 6$

Paso 2: Se determina la frecuencia acumulada hasta el primer valor cuya frecuencia acumulada sea igual o mayor que $\boldsymbol{\frac{n}{2} = \frac{12}{2} = 6}$. En este ejemplo, es 6, porque 6 es igual que 6. Por lo tanto, la nota mediana del alumno es 4.

Notas del alumno	Frecuencia absoluta ($f_{i\cdot}$)	Frecuencia absoluta acumulada ($F_{i\cdot}$)
3	1	1
4	5	6
5	3	9
6	2	
7	1	
Frecuencia total (n)	12	

4.1.2.3 Mediana (M_e) de datos agrupados en intervalos de clase (tabla de frecuencia de tipo III)

Se calcula usando la siguiente fórmula:

$$\boldsymbol{Me = L.I.M + [\frac{\frac{n}{2} - f(ac.ant)}{fm}] \cdot c}$$

Donde:

$L.I.M =$ Límite inferior del intervalo mediano. Se denomina intervalo mediano al que le corresponde una frecuencia acumulada igual o mayor que la mitad de la frecuencia total.

n = Frecuencia total (número total de datos).

$f(ac.ant)$ = Frecuencia acumulada del intervalo anterior al intervalo mediano.

fm = Frecuencia absoluta del intervalo mediano.

c = Tamaño o amplitud del intervalo.

<u>Ejemplo 1</u>: Calcular la mediana de la siguiente distribución:

Se calculan la frecuencia absoluta acumulada hasta que sea igual o mayor que la mitad de la frecuencia total, es decir, $\boldsymbol{n/2 = 47/2 = 23,5}$:

x_i	f_i	$F_i(fac.ant)$
$0-2$	5	5
$2-4$	8	13
$4-6$	9	22
$6-8$	12	34
$8-10$	8	
$10-12$	5	
	$n = 47$	

La frecuencia acumulada mayor que 23,5 es 34 y a esta frecuencia le corresponde el intervalo mediano $6-8$.

El L.I.M del intervalo mediano es 6.
$n = 47$

$f(ac.ant)$: la frecuencia acumulada anterior a la frecuencia acumulada del intervalo mediano es 22.

fm: La frecuencia absoluta del intervalo mediano es 12.

$\boldsymbol{c} =$ La amplitud del intervalo, es decir la diferencia entre el límite superior y el inferior es 2. Se observa, por ejemplo, en el intervalo mediano, donde $8 - 6 = 2$ y en todos los demás.

Aplicando la fórmula, se tiene:

$$Me = 6 + \frac{\frac{47}{2} - 22}{12} \cdot 2 = 6 + (23{,}5 - 22)/12 \cdot 2 = 6 + \frac{1{,}5}{12} \cdot 2 = 6 + \frac{3{,}0}{12} = 6 + 0{,}25 = 6{,}25$$

<u>Ejemplo 2</u>: Calcular la mediana de los salarios semanales en miles de pesos de 65 operarios de una firma constructora:

Salarios en miles de pesos x_i		Número de operarios
$50{,}00 - 59{,}99$		8
$60{,}00 - 69{,}99$		10
$70{,}00 - 79{,}99$		16
$80{,}00 - 89{,}99$		14
$90{,}00 - 99{,}99$		10
$100{,}00 - 109{,}99$		5
$110{,}00 - 199{,}99$		2
		$n = 65$

Se calcula la frecuencia absoluta acumulada hasta aquella que sea igual o mayor que la mitad de la frecuencia total, es decir, $n/2 = 65/2 = 32{,}5$:

Salarios en miles de pesos x_i	Número de operarios f_i	$F_i(fac.ant)$
$50{,}00 - 59{,}99$	8	8
$60{,}00 - 69{,}99$	10	18
$70{,}00 - 79{,}99$	16	34
$80{,}00 - 89{,}99$	14	
$90{,}00 - 99{,}99$	10	
$100{,}00 - 109{,}99$	5	
$110{,}00 - 199{,}99$	2	
	$n = 65$	

La frecuencia acumulada mayor que 32,5 es 34 y a esta frecuencia le corresponde el intervalo mediano $70{,}00 - 79{,}99$.

El $L.I.M$ del intervalo mediano es 70.

$n = 65$

$f(ac.ant)$: la frecuencia acumulada anterior a la frecuencia acumulada del intervalo mediano es 18.

fm: La frecuencia absoluta del intervalo mediano es 16.

$c =$ La amplitud del intervalo, es decir la diferencia entre el límite superior y el inferior es 10. Se observa, por ejemplo, en el intervalo mediano, donde $79{,}99 - 70 = 9{,}99$, que se aproxima a 10.

Aplicando la fórmula, se tiene que la mediana de los salarios es 79,06 miles de pesos semanales:

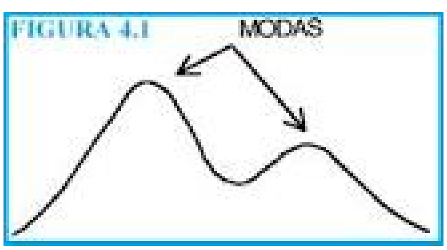
$$Me = 70 + \frac{\frac{65}{2} - 18}{16} \cdot 10 = 70 + (32{,}5 - 18)/16 \cdot 10 = 70 + \frac{14{,}5}{16} \cdot 10 = 70 + \frac{145}{16}$$
$$= 70 + 9{,}06 = 79{,}06$$

4.1.3 Moda o valor modal

En una distribución no tabulada, la moda es el valor de la variable (cualitativa o cuantitativa) que tiene la mayor frecuencia absoluta, es decir, el valor que más se repite en un conjunto de datos. Es un valor de la variable representado en el eje de las abscisas al que corresponde la frecuencia máxima, pero no es una frecuencia.

Si hay una sola moda, la distribución es unimodal; si hay dos, bimodal y si hay más de dos, polimodal o multimodal. Puede ocurrir que la distribución no tenga una moda.

La gráfica siguiente muestra una distribución bimodal:



En una distribución tabulada (tabla de distribución de frecuencias tipo III), la moda puede determinarse de dos maneras:

- La moda es la marca de clase del intervalo con mayor frecuencia absoluta.

- La objeción al método de cálculo anterior es que en una distribución tabulada, los valores de la variable en un intervalo son crecientes gradualmente entre su límite inferior y superior, por lo que no hay valores que se repitan. Por ello, una definición más correcta es decir que la moda es el valor en torno al cual se concentra la mayor cantidad de casos de la distribución.

Por lo mismo, se han propuesto distintas maneras de calcular la moda, una de las cuales es la siguiente:

$$\boldsymbol{Mo = L.I.M + \left(\frac{d_1}{d_1 + d_2}\right) \cdot c}$$

Donde:

$L.I.M$ = Límite inferior del intervalo modal.

d_1 = Diferencia entre la frecuencia absoluta correspondiente al intervalo modal menos la frecuencia contigua de la clase anterior.

d_2 = Diferencia entre la frecuencia absoluta correspondiente al intervalo modal menos la frecuencia contigua de la clase siguiente.

c $=$ Tamaño del intervalo.

La moda es el estadígrafo más representativo en el caso de distribuciones de variables en escala nominal, ya que sus datos no son ordenables y no se pueden hacer operaciones elementales con sus observaciones. Se utiliza, igualmente, cuando la variable tiende a concentrarse hacia un valor determinado.

Ejemplo1: Determinar la moda del siguiente conjunto de notas de un estudiante: 3, 4, 7, 5, 6, 5, 5. La nota que más se repite es el 5, por lo tanto, la moda Mo es 5.

Ejemplo 2: Se encuestó a 180 personas y se le preguntó la marca de automóvil de su preferencia. Los resultados se muestran en la tabla siguiente. ¿Cuál es la moda?

Marca de automóvil	f_i
Mitsubishi	18
Toyota	22
Nissan	40
Chevrolet	25
Ford	75
Total	$n = 180$

De acuerdo con estos resultados, la marca Nissan es la moda, ya que alcanzó la máxima frecuencia.

Ejemplo 3: Las notas obtenidas por 28 estudiantes de un curso en una prueba de Física son:

Notas (x_i)	f_i
1	0
2	0
3	3
4	10
5	9
6	5
7	1
	$n = 28$

La moda $M_o = 4$ ya que es la que tiene la mayor frecuencia.

Ejemplo 4: En la tabla siguiente se muestra el número de familias que tienen entre 0 y 6 o más cargas familiares. Encontrar la moda.

Número de cargas familiares (x_i)	Número de familias (f_i)
0	70
1	110
2	200
3	370
4	170
5	50
6 o más	30
Total	1.000

La frecuencia máxima es 370 que corresponde al valor de la variable 3, por lo que la moda es 3 cargas familiares por familia.

Ejemplo 5: Los sueldos de 120 trabajadores de una empresa se muestran en la siguiente tabla de frecuencias:

Sueldos ($) x	Marca de clase x_i	Frecuencia absoluta f_i
80000-110000	95.000	23
110000-140000	125.000	35
140000-170000	155.000	26
170000-200000	185.000	13
200000-230000	215.000	7
230000-270000	250.000	4
		$n = 120$

La moda, definida en términos de la marca de clase es $125.000, porque corresponde a la marca de clase del intervalo con mayor frecuencia.

Por otra parte, utilizando la fórmula anterior, la moda es:

$$\boldsymbol{M_o = L.I.M + \left(\frac{d1}{d1 + d2}\right) \cdot c}$$

Intervalo modal: $110.000 - 140.000$

L.I.M = 110.000

$d_1 = 35 - 23 = 12$

$d_2 = 35 - 26 = 9$

$c = 140000 - 110000 = 30000$

$$M_o = 110000 + \left(\frac{12}{12+9}\right) \cdot 30000 = 110000 + \frac{12}{21} \cdot 30000 = 127.143$$

<u>Ejemplo 6</u>: La edad de un grupo de personas se presenta en la tabla siguiente. Calcular la moda de la distribución.

Edades	f_i
40-45	7
45-50	15
50-55	20
55-60	30
60-65	17
65-70	12
70-75	9

Utilizando la fórmula anterior, la moda es:

$$M_o = L.I.M + \left(\frac{d1}{d1 + d2}\right) \cdot c$$

Intervalo modal: $60 - 55$
$L.I.M = 55$
$d_1 = 30 - 20 = 10$
$d_2 = 30 - 17 = 13$
$c = 60 - 55 = 5$

$$M_o = 55 + \left(\frac{10}{10 + 13}\right) \cdot 5 = 55 + \frac{10}{23} \cdot 5 = 55 + \frac{50}{23} = \frac{1315}{23} = 57{,}17$$

Algunos autores, Molina (1998), calculan la moda simplemente como el punto medio del intervalo modal:

$$M_o = \frac{55 + 60}{2} = 57{,}5$$

que es un valor muy parecido al determinado con la fórmula anterior.

Ejercicios propuestos sobre la Moda:

1. La tabla siguiente muestra la distribución de los diámetros de las cabezas de los remaches fabricados por una compañía. Calcular la moda de los diámetros.

Diámetro (en pulgadas)	Frecuencia (f_i)
0,7247 - 0,7249	2
0,7250 - 0,7252	6
0,7253 - 0,7255	8
0,7256 - 0,7258	15
0,7259 - 0,7261	42
0,7262 - 0,7264	68
0,7265 - 0,7267	49
0,7268 - 0,7270	25
0,7271 - 0,7273	18
0,7274 - 0,7276	12
0,7277 - 0,7279	4
0,7280 - 0,7282	1
Total	250

Respuesta: $Mo = 0{,}72632$.

2. La moda en la siguiente distribución estadística es 67,8. Explicar cómo se llega a ese resultado.

Diámetro (en pulgadas)	Frecuencia (f_i)
[60 – 63)	5
[63 – 66)	18
[66 – 69)	42
[69 – 72)	27
[72 – 75)	8
Total	100

4.1.4 Relación entre la media aritmética, mediana y moda

4.1.5 Media geométrica

Se define como la raíz enésima del producto de los n valores de la variable.

4.1.5.1 Para datos no agrupados, la fórmula de cálculo es la siguiente:

$$\boldsymbol{Mg = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n}}$$

En la práctica, la M_g se calcula por logaritmos:

$$\mathbf{log}\, \boldsymbol{Mg = \frac{1}{n} log\, \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \dots x_n}}$$

Ejemplo: calcular la media geométrica de los números: 2. 4, 8, 4.

$Mg = \sqrt[4]{2 \cdot 4 \cdot 8 \cdot 4} = \sqrt[4]{256} = 4$

Ejemplo: Calcular la media geométrica de los números: 3, 5, 6, 6, 7, 10, 12 mediante logaritmos.

$\sqrt[7]{3 \cdot 5 \cdot\ 6 \cdot\ 6 \cdot\ 7 \cdot\ 10 \cdot\ 12} = \sqrt[7]{453600}$

$\log Mg = \frac{1}{7} log\ 453600$

$\log Mg = \frac{1}{7} \cdot 5{,}6567 = 0{,}8081$

$M_g = antilogaritmo\ de\ 0{,}8081 = 10^{0.8081} \approx 6{,}43$

4.1.5.2 Para datos agrupados, la fórmula de cálculo es la siguiente:

$$\boldsymbol{Log\ M_g = \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i}$$

$$\boldsymbol{M_g = antilogaritmo\ \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i}$$

Ejemplo: Las calificaciones de 25 estudiantes se distribuyen como sigue:

Calificaciones	f_i
$70 - 75$	2
$75 - 80$	3
$80 - 85$	4
$85 - 90$	5
$90 - 95$	6
$95 - 100$	5
Total	25

Se necesitan las marcas de clase:

Calificaciones	x_i	f_i	$\log x_i$	$f_i \cdot \log x_i$
$70 - 75$	72,5	2	1.8603	3,7206
$75 - 80$	77,5	3	1,8893	5,6679
$80 - 85$	82,5	4	1,9164	7,6656
$85 - 90$	87,5	5	1,9420	9,7100
$90 - 95$	92,5	6	1,9661	11,7966
$95 - 100$	97,5	5	1,9890	9,9450
Total	25	25		48,5057

$$M_g = antilogaritmo\ \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i$$

$$M_g = antilogaritmo\ \frac{1}{25} \cdot 48{,}5057$$

$$M_g = antilogaritmo\ 1.9402$$

$$M_g = 10^{1.9402} = 87{,}14$$

Por tanto, la media geométrica de las calificaciones de los 25 estudiantes es 87,14.

4.1.6 Ejercicios resueltos:

Ejercicio:

Fuente: Tapia, O. y Ormazábal, M. (2012). Matemática PSU. Cuaderno de ejercicios, p. 403. Ediciones UC

Considere la siguiente tabla:

x_i	1	2	3	4	5	6	7
f_i	2	5	4	8	9	12	10

Y los estadígrafos:

I. Moda = 6
II. Mediana = 9,5
III. Media aritmética = 9

¿Cuál (es) corresponde(n) a la información que entrega la tabla?

A. Solo I
B. Solo I y III
C. Solo II y III
D. Ninguna
E. Todas

Solución:

La moda es correcta, ya que es el valor que más se repite (el 6 se repite 12 veces).
Para calcular la mediana, hay que ordenar los valores de menor a mayor, como sigue:

2-4-5-8-9-10-12
La mediana es 8, no 9, 5.

La media aritmética requiere ponderar los valores de la variable x_i por las correspondientes frecuencia y su suma dividirla entre la frecuencia acumulada que es: $2 + 5 + 4 + 8 + 9 + 12 + 10 = 50$

$$\bar{x} = \frac{2 + 10 + 12 + 32 + 45 + 72 + 70}{50} = \frac{243}{50} = 4{,}9$$

La media es 4,9, no 9.

Por tanto, la alternativa correcta es A).

<u>Ejercicio</u>:
Fuente: Tapia, O. y Ormazábal, M. (2012). Matemática PSU. Cuaderno de ejercicios, p. 404. Ediciones UC

En la siguiente tabla faltan algunos valores:

y_i	n_i	f_i	N_i	H_i
3	3	0,06	3	v
4	4	y	7	0,14
5	x	0,14	14	0,28
6	20	0,4	z	0,68
7	16	0,32	50	1

Los valores de x, y, z, v son, respectivamente:

A. 7-0,08-34-0,06
B. 4-0,06-32-0,28
C. 6-0,2-42-0,05
D. 3-0,05-24-0,03
E. Ninguna de las anteriores

Solución:
Se requiere conocer la frecuencia total acumulada (suma de n_i). Para ello se puede utilizar alguno de las frecuencias relativas f_i, por ejemplo 0,06. Sea F esa frecuencia acumulada:

$$\frac{3}{F} = 0{,}06$$

$$F = \frac{3}{0{,}06} = 50.$$
$$x = 50 - 3 - 4 - 20 - 16 = 7$$

$$y = \frac{4}{50} = 0{,}08$$

$$z = 14 + 20 = 34$$
$$v = f_1 = 0{,}06$$

Por consiguiente, la alternativa correcta es la A.

<u>Ejercicio</u>:
Fuente: Tapia, O. y Ormazábal, M. (2012). Matemática PSU. Cuaderno de ejercicios, p. 406. Ediciones UC

Para calcular la nota final de una asignatura, las tres pruebas del semestre se ponderan con un 30%, 30% y 40%, respectivamente. Isabel tiene un 5 y un 4 en las dos primeras. Si su nota final fue 5,1, entonces en la tercera prueba obtuvo un:

A. 5,0
B. 5,1
C. 5,2
D. 6,0
E. 6,3

Solución:

Sea z la tercera nota:

$$\bar{x} = \frac{x_1 \cdot w_1 \ + x_2 \cdot w_2 + x_3 \cdot w_3 + \cdots + x_k \cdot w_k}{w_1 + w_2 + w_3 \ + \cdots + w_k}$$

Sea:

$x_1 = 5$ y $w_1 = 0{,}30$
$x_2 = 4$ y $w_2 = 0{,}30$
$x_2 = 4$ y $w_2 = 0{,}30$
$x_3 =?$ y $w_3 = 0{,}40$

$$\frac{[5{,}0(0{,}30) + \ 4{,}0(0{,}30) + z(0{,}40)]}{0{,}30 + 0{,}30 + 0{,}40} = 5{,}1$$
$$\frac{[5{,}0(0{,}30) + \ 4{,}0(0{,}30) + z(0{,}40)]}{1} = 5{,}1$$

$$0{,}4z = 5{,}1 - 1{,}5 - 1{,}2 = 2{,}4$$

$$x = \frac{2{,}4}{0{,}4} = 6$$

Por tanto, la alternativa correcta es la D.

4.1.7 Ejercicios propuestos:

1. Las calificaciones de un estudiante en cinco asignaturas fueron 85, 76, 93, 82 y 96. Hallar la mediana de dichas calificaciones.

2. Los tiempos de reacción de un individuo a determinados estímulos fueron 0,53, 0,46, 0,50, 0,49, 0,52, 0,53, 0,44 y 0,55 segundos, respectivamente. Determinar la mediana del tiempo de reacción del individuo a los estímulos.

3. Según una encuesta, el número de libros leídos por un grupo de estudiantes de 8º básico durante el año 2001 fue el siguiente:

3	14	9	17	1
22	12	8	9	15
16	8	13	20	12
4	15	24	18	10
6	11	18	15	25

Se pide a) confeccionar una tabla de distribución de frecuencias intervalar siendo el primer intervalo: 1-5, y b) calcular el valor de la mediana.

4. La mediana del conjunto de datos cuya distribución está dada por la siguiente tabla es: A) 1,0; B) 1,5, C) 2,0, D)2,5, E) 3,0.

x_i	f_i
1	7
2	14
3	9
4	8
5	2

5. Se preguntó la edad a un grupo de personas, obteniendo la siguiente distribución:

Edades (x_i)	f_i	F_i
40-45	7	7
45-50	15	22
50-55	20	42
55-60	30	72
60-65	17	89
65-70	12	101
70-75	9	110

Se solicita calcular la mediana de las edades.

4.2 ESTADÍGRAFOS DE LOCALIZACIÓN

Son los deciles y percentiles que, en términos genéricos, se denominan cuantiles. Permiten separar la distribución en grupos con su porcentaje representativo.

4.2.1 Cuartiles:

Son los tres valores de una variable que dividen al conjunto de datos en cuatro partes iguales, de modo que en cada una de ellas está incluido el 25% de los valores de la distribución. Se distingue el primer cuartil ($\boldsymbol{Q_1}$), el segundo cuartil ($\boldsymbol{Q_2}$) y el tercer cuartil ($\boldsymbol{Q_3}$).

Rango intercuartil: $\boldsymbol{r = Q_3 - Q_1}$

Primer cuartil: Es el valor de la variable que deja a la izquierda el 25% de los datos, es decir, un 25% del conjunto de datos queda por debajo del primer cuartil y el 75% sobre él.

25%	75%

Segundo cuartil: Es el valor de la variable que deja a la izquierda el 50% de los datos, por lo que resulta ser igual a la mediana. $\boldsymbol{Q_2 = M_e}$.

50%	50%

Tercer cuartil: Es el valor de la variable que deja a la izquierda el 75% de los datos.

75%	25%

4.2.1.1 Cálculo de cuartiles para datos no agrupados:

1. Ordenación de los datos: de menor a mayor.

2. Determinación de la posición: Se calcula la posición del cuartil buscado aplicando la siguiente fórmula:

$$\boldsymbol{\frac{k(n)}{4}\ si\ el\ número\ de\ datos\ es\ par}$$

$$\boldsymbol{\frac{k(n+1)}{4}\ si\ el\ número\ de\ datos\ es\ impar}$$

donde:
n = número total de observaciones (o tamaño de la muestra)
k la medida de la posición que se está calculando ($k = 1,2,3$).

Ejemplo 1: (datos no presentados en una tabla de frecuencias)

Calcular el cuartil 1, 2 y 3 en el siguiente conjunto de datos: 1,8,8,7,5,6,8,5,4,3,4.

Cálculo del $\boldsymbol{Q_1}$:

1. Ordenación de los datos: de menor a mayor: 1,3,4,4,5,5,6,7,8,8,8. Son 11 datos (impar).
2. Determinación de la posición: $\boldsymbol{\frac{k(n+1)}{4}}$.

   $k = 1$
   $n = 11$ (impar)

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}} = \frac{\boldsymbol{1(11+1)}}{\boldsymbol{4}} = \boldsymbol{3}$ (posición)
4. En la posición 3 está el cuartil 1, es decir, $\boldsymbol{Q_1 = 4}$

Cálculo del $\boldsymbol{Q_2}$:

1. Ordenación de los datos: de menor a mayor: 1,3,4,4,5,5,6,7,8,8,8
2. Determinación de la posición: $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$
3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$

   $k = 2$
   $n = 11$ (corresponde al número de datos, no a la suma de estos)

4. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}} = \frac{\boldsymbol{2(11+1)}}{\boldsymbol{4}} = \boldsymbol{6}$ (posición)
5. En la posición 6 está el cuartil 2, es decir, $Q_{\boldsymbol{2}} = \boldsymbol{5}$

Cálculo del $\boldsymbol{Q_3}$:

1. Ordenación de los datos: de menor a mayor: 1,3,4,4,5,5,6,7,8,8,8
2. Determinación de la posición: $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$
3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$

   $k = 3$
   $n = 11$ (corresponde al número de datos, no a la suma de estos).

4. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}} = \frac{\boldsymbol{3(11+1)}}{\boldsymbol{4}} = \boldsymbol{9}$ (posición)
5. En la posición 9 está el cuartil 3, es decir, $\boldsymbol{Q_3 = 8}$

Ejemplo 2: (datos no presentados en una tabla de frecuencias)

Calcular $\boldsymbol{Q_1, Q_2, Q_3}$

52,53,57,62,64,65,66,68
70,70,71,71,72,75,77,78
78,82,82,83,83,84,85,86
90,91,92,94,96,101,102,102

Cálculo del $\boldsymbol{Q_1}$:

1. Ordenación de los datos de menor a mayor: Ya están ordenados.
2. Determinación de la posición: $\frac{\boldsymbol{k(n)}}{\boldsymbol{4}}$

$k = 1$
$n = 32$ (número par y corresponde al número de datos, no a la suma de estos).

3. $\frac{k(n)}{4} = \frac{1(32)}{4} = \frac{32}{4} = 8\ es\ la\ posición.$

4. Como la posición es 8, $\boldsymbol{Q_1 = 68}$, es decir, el 25% de las observaciones es inferior a 68.

Cálculo del $\boldsymbol{Q_2}$:

1. Ordenación de los datos: de menor a mayor: Están ya ordenados

2. Determinación de la posición: $\boldsymbol{\frac{k(n)}{4}}$

$k = 2$
$n = 32$

3. $\frac{k(n)}{4} = \frac{2(32)}{4} = \frac{64}{4} = 16\ es\ la\ posición$:

4. Como la posición es 16, $\boldsymbol{Q_2 = 78}$, es decir, el 50% de las observaciones es inferior a 78 (corresponde a la mediana).

Cálculo del $\boldsymbol{Q_3}$:

1. Ordenación de los datos: de menor a mayor: Están ya ordenados

2. Determinación de la posición: $\boldsymbol{\frac{k(n)}{4}}$

$k = 3$
$n = 32$

3. $\boldsymbol{\frac{k(n)}{4}} = \boldsymbol{\frac{3(32)}{4}} = \boldsymbol{\frac{96}{4}} = 24, es\ la\ posición$:

4. Como 24 es la posición, $\boldsymbol{Q_3 = 86}$, es decir, el 75% de las observaciones es inferior a 86.

Ejemplo 3: (datos no presentados en una tabla de frecuencias)

La cantidad de alumnos que ha asistido a clases en un colegio durante la primera quincena de clases (15 días) entre lunes y viernes, es la siguiente:

30 28 27 30 25
30 29 29 27 29
28 30 30 30 29

Calcular $\boldsymbol{Q_1} \boldsymbol{y} \ \boldsymbol{Q_3}$.

Cálculo de $\boldsymbol{Q_1}$:

1. Ordenación de los datos de menor a mayor:

25 27 27 28 28
29 29 29 29 30
30 30 30 30 30

2. Determinación de la posición: $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$

$k = 1$
$n = 15$ (número impar de datos)

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}} = \frac{\boldsymbol{1(15+1)}}{\boldsymbol{4}} = \frac{16}{4} = 4$

4. Como la posición es 4, $\boldsymbol{Q_1 = 28}$, es decir, el 25% de los alumnos ha asistido 28 días o menos a clase

Cálculo de $\boldsymbol{Q_3}$:

1. Ordenación de los datos de menor a mayor:

25 27 27 28 28
29 29 29 29 30
30 30 30 30 30

2. Determinación de la posición: $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}}$

$k = 3$
$n = 15$ (número impar de datos)

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{4}} = \frac{\boldsymbol{3(15+1)}}{\boldsymbol{4}} = \frac{48}{4} = 12 \ es \ la \ posición$

4. Como 12 es la posición, $\boldsymbol{Q_3 = 30}$, es decir, el 75% de los alumnos ha asistido 30 días o menos a clase.

Ejemplo 4: Las edades de 10 alumnos es la siguiente. Determinar los tres cuartiles.

15 – 16 – 16 – 17 -17 – 18- 19 – 19 – 20 -21

Cálculo del $\boldsymbol{Q_1}$:

1. Ordenación de los datos de menor a mayor: Ya están ordenados. Los datos son 10 (par).
2. Determinación de la posición: $\frac{\boldsymbol{k(n)}}{\boldsymbol{4}}$

$k = 1$
$n = 10$ (número par y corresponde al número de datos, no a la suma de los mismos).

3. $\frac{\boldsymbol{k(n)}}{\boldsymbol{4}} = \frac{\boldsymbol{1(10)}}{\boldsymbol{4}} = \frac{\boldsymbol{10}}{\boldsymbol{4}} = 2{,}5\ es\ la\ posición.$

4. Como la posición es un número decimal, una manera es promediar el valor de esa posición (16) con la siguiente (16), es decir: $\boldsymbol{Q_1} = \frac{\boldsymbol{16+16}}{\boldsymbol{2}} = \frac{\boldsymbol{32}}{\boldsymbol{2}} = \boldsymbol{16}$, es decir, el 25% de los alumnos tienen una edad inferior a 16 años. Otra manera de promediar es: $16 + (16 - 16)(0{,}50) = 16$ (el 0,50 corresponde a los decimales de 2,5).

Cálculo del $\boldsymbol{Q_2}$:

1. Ordenación de los datos de menor a mayor: Ya están ordenados.
2. Determinación de la posición: $\frac{\boldsymbol{k(n)}}{\boldsymbol{4}}$

$k = 2$
$n = 10$ (número par y corresponde al número de datos, no a la suma de los mismos).

3. $\frac{k(n)}{4} = \frac{2(10)}{4} = \frac{20}{4} = 5\ es\ la\ posición.$

4. En este caso, el valor de la posición es 5 es 17, pero como el número de datos es par, hay que interpolar con la posición 6, que es 18, es decir, $Q_2 = \frac{17+18}{2} = 17{,}5$, o sea, el 50% de los alumnos tiene 17,5 años o menos.

Cálculo del $\boldsymbol{Q_3}$:

1. Ordenación de los datos de menor a mayor: Ya están ordenados.
2. Determinación de la posición: $\frac{\boldsymbol{k(n)}}{\boldsymbol{4}}$

$k = 3$
$n = 10$ (número par y corresponde al número de datos, no a la suma de estos).

3. $\frac{k(n)}{4} = \frac{3(10)}{4} = \frac{30}{4} = 7{,}5\ es\ la\ posición.$

4. En este caso, la posición es 7,5 es decir, un número decimal, por lo que hay que interpolar, es decir, $Q_3 = \frac{19+19}{2} = 19$, es decir, el 75% de los alumnos tiene 19 años o menos.

Ejemplo 5: (datos presentados en una tabla de frecuencias)

El número de hijos de una muestra de 20 familias es la siguiente:

1-1-1-1-1-2-2-2-2-2-2-2-3-3-3-3-4-5.

Hallar los cuartiles $\boldsymbol{Q_1, Q_2, y\ Q_3}$, mostrando la información en una tabla de frecuencias:

Número de hijos	Número de familias
1	6
2	7
3	4
4	2
5	1
Total	20

Se determinan las frecuencias acumuladas:

Número de hijos	Número de familias $\boldsymbol{f_i}$	$\boldsymbol{F_i}$
1	6	6
2	7	13
3	4	17
4	2	19
5	1	20
Total	20	

Cálculo $\boldsymbol{Q_1}$:

$$\frac{kn}{4}$$
$$\frac{1(20)}{4} = 5\ (posición)$$

Se ubica el 5 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 6. A esta frecuencia le corresponde el valor 1 de la variable "número de hijos" por lo que $\boldsymbol{Q_1 = 1}$.

Cálculo $\boldsymbol{Q_2}$:

$\frac{kn}{4} = \frac{2(20)}{4} = 10\ es\ la\ posición$

Se ubica el 10 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 13. A esta frecuencia le corresponde el valor 2 de la variable "número de hijos" por lo que $\boldsymbol{Q_1 = 2}$.

Cálculo $\boldsymbol{Q_3}$:

$\frac{kn}{4}$

$\frac{3(20)}{4} = 15\ (posición)$

Se ubica el 15 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 17. A esta frecuencia le corresponde el valor 3 de la variable "número de hijos" por lo que $\boldsymbol{Q_3 = 3}$.

Ejemplo 6: (datos presentados en una tabla de frecuencias)
Las edades de 60 estudiantes son las que se indican en la siguiente tabla.

Hallar los cuartiles $\boldsymbol{Q_1, Q_2, y\ Q_3}$:

Edades	Número de familias
13	3
14	14
15	23
16	10
17	5
18	4
19	1
Total	60

Se determinan las frecuencias acumuladas:

Edades	Número de familias	$\boldsymbol{F_i}$
13	3	3
14	14	17
15	23	40
16	10	50
17	5	55
18	4	59
19	1	60
Total	60	

Cálculo $\boldsymbol{Q_1}$:

$$\frac{kn}{4}$$

$$\frac{1(60)}{4} = 15\ (posición)$$

Se ubica el 15 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 17. A esta frecuencia le corresponde el valor 14 de la variable “edades” por lo que $\boldsymbol{Q_1 = 14}$.

Cálculo $\boldsymbol{Q_2}$:

$$\frac{kn}{4}$$

$$\frac{2(60)}{4} = 30\ (posición)$$

Se ubica el 30 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 40. A esta frecuencia le corresponde el valor 15 de la variable “número de hijos” por lo que $\boldsymbol{Q_1 = 15}$.

Cálculo $\boldsymbol{Q_3}$:

$$\frac{kn}{4}$$

$$\frac{3(60)}{4} = 45\ (posición)$$

Se ubica el 45 en la columna $\boldsymbol{F_i}$. En este caso no aparece y cuando esto ocurre se elige el valor más próximo que sea mayor, en este caso, el 60. A esta frecuencia le corresponde el valor 16 de la variable “edades” por lo que $\boldsymbol{Q_3 = 16}$.

4.2.1.2 Cálculo de cuartiles para datos agrupados:

Para datos agrupados en intervalos, los cuartiles se calculan con la siguiente fórmula:

$$\boldsymbol{Q_k = LI + \left(\frac{\frac{kn}{4} - fac.ant}{fac - fac.ant}\right) \cdot c}$$

Donde:

$\boldsymbol{LI}$ = Límite inferior del intervalo que contiene al cuartil que se busca.

$\boldsymbol{k = 1, 2, 3}$.

$\boldsymbol{n}$ =Totalidad de datos

$\boldsymbol{fac.ant}$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que corresponde al cuartil buscado.

$\boldsymbol{fac}$ = Frecuencia absoluta acumulada que contiene al cuartil buscado.
$\boldsymbol{c}$ = Amplitud de los intervalos.

Ejemplo 1:

Hallar los cuartiles $\boldsymbol{Q_1, Q_2, y\ Q_3}$ de los salarios semanales en dólares de los empleados de una compañía:

Salarios (dólares)	Número de empleados
50 – 59,99	8
60 – 69.99	10
70 – 79,99	16
80 – 89,99	14
90 – 99,99	10
100 – 109,99	5
110 – 119,99	2
Total	65

Fuente: Estadística, serie Schaum, p. 62

Se determinan las frecuencias relativas y acumuladas:

Salarios (dólares)	Número de empleados $\boldsymbol{f_i}$	$\boldsymbol{F_i}$	$\boldsymbol{h_i}$	$\boldsymbol{H_i}$
50 – 59,99	8	8	12,3 %	12,3 %
60 – 69.99	10	18	15,4%	27,7%
70 – 79,99	16	34	24,6%	52,3%
80 – 89,99	14	48	21,5%	73,8%
90 – 99,99	10	58	15,4%	89,2%
100 – 109,99	5	63	7,7%	96,9%
110 – 119,99	2	65	3,1%	100,0%
Total	65			

$$\boldsymbol{Q_k = LI + \left(\frac{\frac{kn}{4} - fac.ant}{fac - fac.ant}\right) \cdot c}$$

Cálculo $\boldsymbol{Q_1}$:

$LI = 60$, porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 25% (en este caso, 27,7%).

$$Q_1 = 60 + \left(\frac{\frac{1(65)}{4} - 8}{18 - 8}\right) \cdot 10 = 60 + 8{,}25 = 68{,}25$$

Es decir, el 25% de los empleados gana $\$\boldsymbol{68,25}$ o menos.

Cálculo $\boldsymbol{Q_2}$:

$LI = 90$, porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 50% (en este caso 52,3%).

$$Q_2 = 70 + \left(\frac{\frac{2(65)}{4} - 18}{34 - 18}\right) \cdot 10 = 70 + 9{,}06 = 79{,}06$$

Es decir, el 50% de los empleados gana $79,06 o menos.

Cálculo $\boldsymbol{Q_3}$:

$LI = 90$, porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 75% (en este caso 89,2%).

$$Q_3 = 90 + \left(\frac{\frac{3(65)}{4} - 48}{58 - 48}\right) \cdot 10 = 90 + 0{,}75 = 90{,}75$$

Es decir, el 75% de los empleados gana $90,75 o menos.

<u>Ejemplo 2</u>: Las edades de un grupo de 100 personas se muestran en la tabla de distribución de frecuencias siguiente. Se pide calcular Q1, Q2 y Q3.

Edades	f_i	F_i
40-45	7	7
45-50	15	22
50-55	20	42
55-60	30	72
60-65	17	89
65-70	2	91
70-75	9	100
	100	

Se calculas las frecuencias relativas y acumuladas:

Edades	f_i	F_i	h_i	H_i
40-45	7	7	7,0%	7,0%
45-50	15	22	15,0%	22,0%
50-55	20	42	20,0%	42,0%
55-60	30	72	30,0%	72,0%
60-65	17	89	17,0%	89,0%
65-70	2	91	2,0%	91,0%
70-75	9	100	9,0%	100,0%
	100		100,00%	

$$\boldsymbol{Q1 = LI + \left(\frac{\frac{n}{4} - fac.ant}{fac - fac.ant}\right) \cdot c}$$

Cálculo de $\boldsymbol{Q_1}$:

$LI = 50,$ porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 25% (en este caso, 42,0%).

$$Q_1 = 50 + \left(\frac{\frac{1(100)}{4} - 22}{42 - 22}\right) \cdot 5 = 50 + 0,75 = 50,75$$

Es decir, el 25% de las personas tiene 50,75 años ≈ 51 años o menos.

Cálculo $\boldsymbol{Q_2}$:

$LI = 55,$ porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 50% (en este caso, 72,0%).

$$Q_2 = 55 + \left(\frac{\frac{2(100)}{4} - 42}{72 - 42} \right) \cdot 5 = 55 + 1{,}33 = 56{,}33$$

Es decir, el 50% de las personas tiene 56,33 ≈ 56 años o menos.

Cálculo $\boldsymbol{Q_3}$:

$LI = 60,$ porque corresponde al límite inferior del intervalo que corresponde a la frecuencia acumulada igual o superior al 75% (en este caso, 89,0%).

$$Q_3 = 60 + \left(\frac{\frac{3(100)}{4} - 72}{89 - 72} \right) \cdot 5 = 60 + 0{,}88 = 60{,}88$$

Es decir, el 75% de las personas tiene 60,88 años ≈ 61 años o menos.

<u>Ejemplo 3:</u>
Fuente: PSU matemáticas, Proyecto Clave, p. 372.

La cantidad de respuestas correctas en un examen de matemática es la siguiente:

Cantidad de respuestas correctas	5	6	7	8	9	19
$\boldsymbol{f_i}$	4	8	8	12	8	4

¿Cuántas respuestas correctas debe tener un estudiante para asegurar que está sobre el primer cuartil de las calificaciones?

Se tiene que completar la siguiente tabla de frecuencias:

Cantidad de respuestas	$\boldsymbol{f_i}$	$\boldsymbol{F_i}$	$\boldsymbol{f_i}$(%)	$\boldsymbol{F_i}$ (%)
5	4	4	9,1	9,1
6	8	12	18,2	27,3
7	8	20	18,2	45,5
8	12	32	27,3	72,8
9	8	40	18,2	9,1
10	4	44	9,1	100,1

Como $\boldsymbol{Q_1}$corresponde al 25% de los estudiantes que obtuvo menor cantidad de respuestas correctas, es posible identificar en la tabla de frecuencias que $\boldsymbol{Q_1 = 6}$; por lo tanto, para asegurar estar sobre el 25% de las calificaciones, se deben responder correctamente, a lo menos, 7 preguntas.

4.2.2 Deciles:

Son los nueve valores que dividen al conjunto de datos en diez partes iguales, de modo que en cada una de ellas está incluido el 10% de los valores de la distribución. Se denominan $\boldsymbol{D_1}, \boldsymbol{D_2}, \boldsymbol{D_3}, \dots, \boldsymbol{D_k}$. El decil $\boldsymbol{D_k}$ es el valor de la variable que deja a la izquierda $k \cdot 10\%$ de los datos ($\boldsymbol{k = 1, 2, 3, \dots, 9}$); en otras palabras, indica el valor por debajo del cual se encuentra un determinado porcentaje de observaciones.

$k \cdot 10\%$	$(100 - k) \cdot 10\%$

Cálculo de deciles para datos no agrupados:

1. Se ordenan las observaciones de menor a mayor valor.
2. Se calcula la posición que ocupa el decil buscado aplicando la siguiente fórmula:

Para número par de datos:

$$\frac{\boldsymbol{k(n)}}{\boldsymbol{10}}$$

Para número impar de datos:

$$\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}}$$

siendo n el número total de observaciones (o tamaño de la muestra) analizadas y la letra "k" el decil buscado.

Ejemplo 1: Calcular los deciles $\boldsymbol{D_1}$, $\boldsymbol{D_6}$ $\boldsymbol{y}$ $\boldsymbol{D_8}$ de las siguientes observaciones:10,11,11,12, 12, 13, 13, 13, 14, 15, 17, 18, 20.

Cálculo del $\boldsymbol{D_1}$:

1. Se ordenan las observaciones de menor a mayor: 10, 11,11,12, 12, 13, 13, 13, 14, 15, 17,18, 20.

2. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}}$
   k=1
   n=13

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{1(13+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{14}}{\boldsymbol{10}} =$
   $1{,}4\ es\ la\ posición$: $es\ a\ la\ derecha\ del\ 10, ya\ que\ 10\ es\ la\ posición\ 1$.

4. Se interpola si la posición sale con decimales: $10 + (11 - 10)(0{,}40) = 10{,}4$. Se multiplicó por 0,40 porque corresponde al decimal de 1,4.

5. $D_1 = 10{,}4$, es decir, el 10% de las observaciones es inferior a 10,4.

Cálculo del $\boldsymbol{D_6}$:

1. Se ordenan las observaciones de menor a mayor: 10, 11,11,12, 12, 13, 13, 13, 14, 15, 17,18, 20.

2. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}}$
   k=6
   n=13

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{6(13+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{84}}{\boldsymbol{10}} =$
   $8{,}4\ es\ la\ posición: es\ a\ la\ derecha\ del\ 13, ya\ que\ 13\ es\ la\ posición\ 8.$

4. Se interpola si la posición sale con decimales: $13 + (14 - 13)(0{,}40) = 13{,}4$. Se multiplicó por 0,40 porque corresponde al decimal de 8,4.

5. $D_6 = 13{,}4$, es decir, el 60% de las observaciones es inferior a 13,4.

Cálculo del $\boldsymbol{D_8}$:

1. Se ordenan las observaciones de menor a mayor: 10, 11,11,12, 12, 13, 13, 13, 14, 15, 17,18, 20.

2. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}}$
   k=8
   n=13

3. $\frac{\boldsymbol{k(n+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{8(13+1)}}{\boldsymbol{10}} = \frac{\boldsymbol{112}}{\boldsymbol{10}} =$
   $11{,}2\ es\ la\ posición: es\ a\ la\ derecha\ del\ 17, ya\ que\ 17\ es\ la\ posición\ 11.$

4. Se interpola si la posición sale con decimales: $\boldsymbol{17 + (18 - 17)(0{,}20) = 17{,}2}$. Se multiplicó por 0,20 porque corresponde al decimal de11,2.

5. $D_8 = 17{,}2$, es decir, el 80% de las observaciones es inferior a 17,2.

Cálculo de deciles para datos agrupados en intervalos:

Se utiliza la siguiente fórmula:

$$\boldsymbol{Dk = LI + \frac{k \cdot \frac{n}{10} - fac.ant}{fac - fac.ant} \cdot c}$$

Donde:

LI = Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el k% de los datos.

$\boldsymbol{k} =$ 1,2,3,5,5,6,7,8,9,10, para los deciles, D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, respectivamente.

n = Total de datos.

fac = Frecuencia absoluta acumulada que supera el k% de los datos.

$fac.ant$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el k% de los datos.

c = Amplitud de los intervalos.

Ejemplo 1: Las edades de un grupo de 100 personas se muestran en la siguiente la tabla de distribución de frecuencias. Se pide calcular $\boldsymbol{D_2}$.

Edades	f_i	F_i
[40-45)	7	7
[45-50)	15	22
[50-55)	20	42
[55-60)	30	72
[60-65)	17	89
[65-70)	12	91
70-75	9	100

fac = Como se pide el segundo decil, en la columna F_i de frecuencias acumuladas se busca el valor que supere 20% del total de personas, o sea, 0,20*100=20 personas. Ese valor corresponde a la frecuencia absoluta acumulada = 22.

LI = Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el 20% de los datos. Es 45.

$fac.ant.$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el 20% de los datos. Es 7.

$n = 100$ (total de personas)

$\boldsymbol{c}$ = Amplitud de los intervalos = 5

$$\boldsymbol{D_2 = LI + \frac{k \cdot \frac{n}{10} - fac.ant}{fac - fac.ant} \cdot c}$$

$D_2 = 45 + \frac{2*\frac{100}{10}-7}{22-7} \cdot 5 = \frac{13}{15} \cdot 5 = 45 + 4{,}33 = 49{,}33$≈49 años.

Es decir, el 20% de las edades no superan los 49 años.

Ejemplo 2: Calcular el tercer decil de los salarios semanales en miles de pesos de 65 operarios de una firma constructora:

(Salarios en miles de pesos)	Número de operarios f_i	$F_i(fac.ant.)$
50,00-59,99	8	8
60,00-69,99	10	18
70,00-79,99	16	34
80,00-89,99	14	48
90,00-99,99	10	58
100,00-109,99	5	63
110,00-199,99	2	65
	n=65	

fac = Como se pide el tercer decil, en la columna F_i de frecuencias acumuladas se busca el valor que supere 30% del total de personas, o sea, 0,30*65=19,5 personas. Ese valor corresponde a la frecuencia absoluta acumulada = 34.

LI = Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el 30% de los datos. Es 70.

$fac.ant.$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el 30% de los datos. Es 18.

$n = 65$ (total de personas)

$\boldsymbol{c}$ = Amplitud de los intervalos = 10

$$\boldsymbol{Dk = LI + \frac{k \cdot \frac{n}{10} - fac.ant}{fac - fac.ant} \cdot c}$$

$D_3 = \mathbf{70} + \frac{3 \cdot \frac{65}{10} - 18}{34 - 18} \cdot 10 = 70 + \frac{19{,}5 - 18}{16} \cdot 10 = 70 + \frac{1{,}5}{16} \cdot 10 = 70 + 0{,}94 = 70{,}94$ años.

Es decir, el 30% de los operarios no superan los 70,94 ≈ 71 miles de pesos de ingresos semanales o, si se prefiere, el 30% de los operarios gana 71 miles de pesos o menos.

Ejemplo 3 de un cálculo de deciles: Calcular el decil 1 y el decil 8 de la siguiente distribución:

Intervalos	$\boldsymbol{f_i}$	$\boldsymbol{F_i}$
[5,2-6,1)	3	3
[6,1-7,0)	5	8
[7,0-7,9)	9	17
[7,9-8,8)	7	24
[8,8-9,7)	5	29
[9,7-10,6)	3	32
	32	

Cálculo del primer decil: $\boldsymbol{D_1}$

fac = Como se pide el primer decil, en la columna F_i de frecuencias acumuladas se busca el valor que supere 10% del total de personas, o sea, 0,10*32=3,2 personas. Ese valor corresponde a la frecuencia absoluta acumulada = 8.

LI = Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el 10% de los datos. Es 6,1.

$fac.ant$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el 10% de los datos. Es 3.

$n = 32$ (total de personas)

$\boldsymbol{c}$ = Amplitud de los intervalos = 0,9

$$\boldsymbol{D_1 = LI + \frac{k \cdot \frac{n}{10} - fac.ant}{fac - fac.ant} \cdot c}$$

$$D_1 = 6{,}1 + \frac{1 \cdot \frac{32}{10} - 3}{8 - 3} \cdot 0{,}9 = 6{,}1 + \left(\frac{0{,}2}{5}\right) \cdot 0{,}9 = 6{,}1 + 0{,}036 = 6{,}136$$

Cálculo del decil 8: $\boldsymbol{D_8}$

fac = Como se pide el decil 8, en la columna F_i de frecuencias acumuladas se busca el valor que supere 80% del total de personas, o sea, 0,80*32=25,6. Ese valor corresponde a la frecuencia absoluta acumulada = 29.

LI = Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el 80% de los datos. Es 8,8.

$fac.ant$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el 80% de los datos. Es 24.

$n = 32$ (total de personas)

$\boldsymbol{c}$ = Amplitud de los intervalos = 0,9

$$D_8 = 8{,}8 + \frac{8 \cdot \frac{32}{10} - 24}{29 - 24} \cdot 0{,}9 = 8{,}8 + \left(\frac{1{,}6}{5}\right) \cdot 0{,}9 = 8{,}8 + 0{,}288 = 9{,}088$$

Significa que el 80% de los datos es menor que 9,088 y el 20% restante serán superiores a 9,088.

4.2.3 Percentiles:

Son medidas de localización que dividen al conjunto de datos en cien partes iguales, de modo que en cada una de ellas está incluido el 1% de los valores de la distribución. Se denominan $\boldsymbol{P_1}, \boldsymbol{P_2}, \boldsymbol{P_3}, \ldots, \boldsymbol{P_k}$. El percentil $\boldsymbol{P_k}$ es el valor de la variable que deja a la izquierda k% de los datos ($k = 1{,}2{,}3, \ldots, 99$).

$k \cdot 1\%$	$(100 - k) \cdot 1\%$

Por ejemplo, si "k" es 30 y "n" es 50, entonces el percentil 30 (P_{30}) es un número tal que garantiza que en la muestra un 30% ($50 \cdot 0{,}30 = 15\ observaciones$) son menores que ($P_{30}$) y un 70% ($50 \cdot 0{,}70 = 35\ observaciones$) son mayores a ($P_{30}$).

Son medidas estadística muy utilizada cuando se quiere clasificar o ubicar características en las personas, por ejemplo, su peso, estatura o calificaciones.

Algunos casos particulares, son:

P_0 = Percentil 0 = Mínimo

P_{100} = Percentil 100 = Máximo

P_{50} = Percentil 50 = Mediana = Segundo cuartil (Q_2)

P_{25} = Percentil 25 = Primer Cuartil (Q_1)

P_{75} = Percentil 75 = Tercer Cuartil

P_{10} = Percentil 10 = Primer decil

P_{90} = Percentil 90 = Noveno decil.

4.2.3.1 Cálculo de percentiles para datos no agrupados:

1. Se ordenan las observaciones de menor a mayor valor.
2. Se calcula la posición que ocupa el percentil buscado aplicando la siguiente fórmula:

Para número par de datos:

$$\frac{\boldsymbol{k(n)}}{\mathbf{100}}$$

Para número impar de datos:

$$\frac{\boldsymbol{k(n+1)}}{\mathbf{100}}$$

$n = número\ total\ de\ obsrvaciones$
$k = percentil\ buscado = 1,2,3,4,5, \ldots, 99)$

Ejemplo 1: Calcular el percentil 30 de las siguientes observaciones: 3, 5, 6, 11, 14, 18, 18, 20, 24, 25, 27, 27, 28, 31, 33, 34, 36, 44, 45, 47, 48, 48, 50, 50, 52.

$n = 25 = cantidad\ impar\ de\ datos$
$k = 30 = percentil\ buscado$

Se ordenan las observaciones de menor a mayor: Ya lo están.

Como la cantidad de observaciones es 25, un número impar:

$$\frac{k(n+1)}{100} = \frac{30(25+1)}{100} = \frac{30 \cdot 26}{100} = 7{,}8$$

Al ser la posición un número decimal, se debe interpolar el valor que corresponde a la posición 7 (que es la observación 18) con la siguiente (que es la 20), como sigue: $18 + (20 - 18)(0{,}8) =$ 19,6. Se multiplicó la diferencia entre paréntesis por 0,8, porque corresponde al decimal de 7,8.

$P_{30} = 19{,}6$, es decir, el 30% de las observaciones son valores inferiores a 19,6.

Ejemplo 2: Los pesos, en Kg, de 10 estudiantes son los siguientes: 58, 67, 70, 56, 64, 67, 59, 60, 61, 59. Calcular el percentil 75.

Solución:

Se ordenan los pesos de menor a mayor:

56, 58, 59, 59, 60, 61, 64, 67, 67, 70

$n = 10 = cantidad\ par\ de\ datos$
$k = 75 = percentil\ buscado$

Como la cantidad de observaciones es 10, un número par:

$$\frac{k\text{n}}{100} = \frac{75(10)}{100} = \frac{750}{100} = 7{,}5$$

Al ser la posición un número decimal, se debe interpolar el valor que corresponde a la posición 7 (que es 64 Kg) con la siguiente (que es 67 Kg), como sigue: $64 + (67 - 64)(0{,}5) = 65{,}5$. Se multiplicó la diferencia entre paréntesis por 0,5, porque corresponde al decimal de 7,5.

6. $P_{75} = 65{,}5$, es decir, el 75% de los alumnos tiene una masa corporal menor de 65,5 Kg.

4.2.3.2 Cálculo de percentiles para datos agrupados en intervalos:

La forma de calcular los percentiles es:

$$\boldsymbol{Pk = LI + \frac{k \cdot \frac{n}{100} - fac.ant}{fac - fac.ant} \cdot c}$$

Ejemplo: Las edades de un grupo de 100 personas se muestran en la tabla de distribución de frecuencias siguiente. Se pide calcular P_{60}.

Edades	f_i	F_i
40-45	7	7
45-50	15	22
50-55	20	42
55-60	30	72
60-65	17	89
65-70	12	91
70-75	9	100

$fac =$Como se pide el percentil 60, en la columna F_i de frecuencias acumuladas se busca el valor que supere el 60% del total de personas, o sea, $0{,}60 \cdot 100 = 60$ personas. Ese valor corresponde a la frecuencia absoluta acumulada = 72.

$LI\ =$ Límite inferior del intervalo que corresponde a la frecuencia acumulada que supera el 60% de los datos. Es 55.

$fac.ant$ = Frecuencia absoluta acumulada anterior a la frecuencia absoluta acumulada que supera el 60% de los datos. Es 42.
$n = 100$ (total de personas)

c = Amplitud de los intervalos = 5

$P_{60} = 55 + \frac{60 \cdot \frac{100}{100} - 42}{72-42} \cdot 5 = 55 + \frac{60-42}{30} \cdot 5 = 55 + \frac{18}{30} \cdot 5 = 55 + 3 = 58$ años.

Lo que significa que el 60% de las edades no superan los 58 años.

4.3 ESTADÍGRAFOS DE DISPERSIÓN O DE VARIABILIDAD

Las medidas de dispersión indican cuán cercanos o alejados están los datos de un valor central, por lo general, la media aritmética. En otras palabras, tienen por objetivo estudiar hasta qué punto las medidas de tendencia central calculadas son representativas de la distribución. Por ejemplo, si se quiere estudiar la representatividad de una media aritmética de un conjunto de datos, hay que calcular la desviación de cada uno de los valores de la distribución con respecto a esa media aritmética. Solo si los valores están cercanos a ella, la media aritmética los representará adecuadamente. Cabe señalar, con todo, que algunos estadígrafos de dispersión no están referidos a promedios, por lo que, se suelen clasificar como sigue:

Medidas de dispersión no referentes a promedios: Dos de ellos son:

- Rango o recorrido
- Coeficiente de apertura

Medidas de dispersión referentes a promedios:

- Desviación media
- Varianza
- Desviación estándar o típica

4.3.1 Rango o recorrido:

Es la diferencia entre el mayor valor y el menor valor de una distribución. Es una medida de dispersión no muy significativa, ya que solo indica cuán dispersos están los datos de la distribución entre los valores extremos.

$$\boldsymbol{R = xmax - xmin}$$

Ejemplo: Calcular el rango de los sueldos en pesos de 10 empleados de una empresa: 800.000, 750.000, 1.100.000, 1.000.000, 850.000, 700.000, 1.100.000, 900.000, 750.000, 1.520.000.

Se ordenan los sueldos de menor a mayor, o viceversa:

Sueldo ($)
700.000
750.000
750.000
800.000
850.000
900.000
1.000.000
1.100.000
1.100.000
1.520.000

El sueldo mayor es $1.520.000 y el menor es $700.000, por lo que el rango es: $1.520.000 - 700.000 = \$820.000$.

4.3.2 Coeficiente de apertura:

Es el cociente entre el mayor valor y el menor valor de la distribución. En el ejemplo anterior de los sueldos, el coeficiente de apertura es: $1.520.000/700.000 = 2{,}17$ veces.

4.3.3 Desviación media:

Es la media aritmética entre las diferencias, expresadas en valores absolutos, entre cada valor de la variable y la media aritmética del conjunto de valores. Los valores absolutos se utilizan para evitar que las desviaciones de signo contrario se cancelen mutuamente.

La desviación media para datos no tabulados es:

$$\boldsymbol{DM = \frac{\sum|x_i - \bar{x}|}{n}}$$

donde:

$x_i = Valores\ individuales\ de\ la\ variable$
$\bar{x} = Media\ aritmética$
$n = Totalidad\ de\ datos$

La desviación media para datos tabulados en tablas de frecuencia tipo II, es:

$$\boldsymbol{DM = \frac{\sum|x_i - \bar{x}| \cdot fi}{n}}$$

La desviación media para datos tabulados en tablas de frecuencia tipo III (intervalar), es la misma anterior, solo que ahora los valores x_i son las marcas de clase de los respectivos intervalos.

Ejemplo 1: Calcular la desviación media de los valores: 4, 6, 8, 12, 15.

1) Se calcula la media aritmética:

$$\boldsymbol{\bar{x}} = \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}}$$

$$\bar{x} = \frac{4+6+8+12+15}{5} = \frac{45}{5} = 9$$

2) Se calcula la diferencia entre cada valor y la $\boldsymbol{\bar{x}}$ expresada en valor absoluto (sin signo) y luego se suman los valores obtenidos:

x_i	$x_i - \boldsymbol{\bar{x}}$	$\lvert x_i - \boldsymbol{\bar{x}} \rvert$
4	4-9=-5	5
6	6-9=-3	3
8	8-9=-1	1
12	12-9=3	3
15	15-9=6	6
n=5		$\sum \lvert x_i - \boldsymbol{\bar{x}} \rvert$ =18

3) Se aplica la fórmula:

$$\boldsymbol{DM} = \frac{\sum \lvert \boldsymbol{x_i} - \boldsymbol{\bar{x}} \rvert}{\boldsymbol{n}}$$

$$DM = \frac{18}{5} = 3{,}6$$

Ejemplo 2: Calcular la desviación media de las siguientes notas de un estudiante:

Asignatura	Nota (x_i)
Matemática	5,2
Física	5,7
Ciencias sociales	6,2
Biología	5,4
Química	4,5

1) Se calcula la media aritmética:

$$\boldsymbol{\bar{x}} = \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}}$$

$$\bar{x} = \frac{5{,}2+5{,}7+6{,}2+5{,}4+4{,}5}{5} = \frac{27}{5} = 5{,}4$$

2) Se calcula la diferencia entre cada valor y la Ma expresada en valor absoluto (sin signo) y luego se suman los valores obtenidos:

x_i	$x_i - \boldsymbol{\bar{x}}$	$\lvert x_i - \boldsymbol{\bar{x}} \rvert$
5,2	5,2-5,4=-0,2	0,2
5,7	5,7-5,4=0,3	0,3
6,2	6,2-5,4=0,8	0,8
5,4	5,4-5,4=0,0	0.0
4,5	4,5-5,4=-0,9	0,9
n=5		$\sum \lvert \boldsymbol{xi} - \boldsymbol{\bar{x}} \rvert$ =22

3) Se aplica la fórmula:

$$\boldsymbol{DM = \frac{\sum |x_i - \bar{x}|}{n}}$$

$$DM = \frac{22}{5} = 0{,}44$$

<u>Ejemplo 3</u>: Calcular la desviación media de la siguiente distribución:

x_i	$\boldsymbol{f_i}$
2	3
4	2
5	4
6	6
8	5
9	8
14	2
	n=30

1) Se calcula la $\boldsymbol{\bar{x}}$ y se determina la columna $\boldsymbol{f_i \cdot x_i}$:

$$\boldsymbol{\bar{x}} = \frac{\sum x_i \cdot \boldsymbol{f_i}}{\boldsymbol{n}}$$

$$\bar{x} = \frac{2 \cdot 3 + 4 \cdot 2 + 5 \cdot 4 + 6 \cdot 6 + 8 \cdot 5 + 9 \cdot 8 + 14 \cdot 2}{30} = \frac{210}{30} = 7$$

2) Se determina la columna $|x_i - \boldsymbol{\bar{x}}|$ (en valores absolutos)

3) Se determina la columna $\boldsymbol{f}i|x_i - \boldsymbol{\bar{x}}|$ y se obtiene la suma correspondiente.

x_i	$\boldsymbol{f_i}$	$\lvert x_i - \boldsymbol{\bar{x}} \rvert$	$\boldsymbol{f_i} \lvert x_i - \boldsymbol{\bar{x}} \rvert$
2	3	5	1
4	2	3	6
5	4	2	8
6	6	1	6
8	5	1	5
9	8	2	16
14	2	7	14
	$n = 30$		$\sum \lvert x_i - \bar{x} \rvert \cdot fi = 70$

4) Se aplica la fórmula

$$\boldsymbol{DM = \frac{\sum \lvert x_i - \bar{x} \rvert \cdot f_i}{n}}$$

$$DM = \frac{70}{30} = 2{,}33 \ldots$$

Ejemplo 4: Calcular la desviación media de la siguiente distribución de frecuencia de datos agrupados en intervalos:

x_i	$\boldsymbol{f_i}$
1-5	4
5-9	8
9-13	9
13-17	10
17-21	6
21-25	5
	$n = 42$

Solución:

	Marca de clase x_i	$\boldsymbol{f_i}$	$x_i \cdot \boldsymbol{f_i}$	$\lvert x_i - \bar{x} \rvert$	$\boldsymbol{f_i} \lvert x_i - \bar{x} \rvert$
1-5	3	4	12	10	40
5-9	7	8	56	6	48
9-13	11	9	99	2	18
13-17	15	10	150	2	20
17-21	19	6	114	6	36
21-25	23	5	115	10	50
		$n = 42$	$\sum x_i \cdot \boldsymbol{f_i} = 546$		$\sum \boldsymbol{f_i} \lvert x_i - \bar{x} \rvert$ =212

1) Se calcula la marca de clase xi, se determina la columna $x_i \cdot \boldsymbol{f}i$ y se suman los valores.

$$\overline{\boldsymbol{x}} = \frac{\sum \boldsymbol{x_i} \cdot \boldsymbol{f_i}}{\boldsymbol{n}}$$

$$\bar{x} = \frac{546}{42} = 13$$

2) Se determina la columna $|x_i - \boldsymbol{x}|$.

3) Se determina la columna $\boldsymbol{f_i}|x_i - \boldsymbol{x}|$ y se obtiene la suma.

4) Se aplica la fórmula:

$$DM = \frac{\sum |x_i - x| \cdot fi}{n}$$

$$DM = \frac{212}{42} = 5{,}04$$

4.3.4 Varianza:

Es una medida de dispersión que corresponde al promedio de los cuadrados de las desviaciones de cada valor de la variable con respecto a la media aritmética. A mayor dispersión, mayor valor de la varianza y a menor dispersión, menor valor de la varianza. Se puede definir, también, como el cuadrado de la desviación típica (s^2, para una muestra y σ^2, para una población).

La fórmula para calcular la varianza para datos no agrupados, es:

$$\boldsymbol{s^2} = \frac{\sum (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{\boldsymbol{n}}$$

La fórmula para calcular la varianza para datos agrupados es:

$$\boldsymbol{s^2} = \frac{\sum \boldsymbol{f_i}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{\boldsymbol{n}}$$

donde,

$\boldsymbol{x_i}$ = marca de clase del intervalo
$\boldsymbol{n}$ = totalidad de datos
$\boldsymbol{f_i}$ = frecuencia del intervalo

<u>Ejemplo 1</u>: Calcular la varianza de la siguiente distribución de puntajes de acceso a la universidad correspondientes a 100 estudiantes:

Intervalos	Marca de clase x_i	$\boldsymbol{f_i}$
400-450	425	10
450-500	475	20
500-550	525	30
550-600	575	40
600-650	625	15
650-700	675	10
700-750	725	5
		$n = 130$

Solución:

$Intervalos$	Marca de clase x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
400-450	425	10	4.250	-131	17.161	171.610
450-500	475	20	9.500	-81	6.561	131.220
500-550	525	30	15.750	-31	961	28.830
550-600	575	40	23.000	19	361	14.440
600-650	625	15	9.375	69	4.761	71.415
650-700	675	10	6.750	119	14.161	141.610
700-750	725	5	3.625	169	28.561	142.805
		$n = 130$	72.250			701.930

1) Se calcula la $\boldsymbol{\bar{x}}$. Se determina la columna $\boldsymbol{x_i \cdot f_i}$ y la suma de sus valores.

$$\boldsymbol{\bar{x} = \frac{\sum x_i \cdot f_i}{n}}$$

$$\bar{x} = \frac{72.250}{130} = 555{,}8 \approx 556$$

2) Se calculan las desviaciones $\boldsymbol{x_i - \bar{x}}$ y se elevan al cuadrado, $(\boldsymbol{x_i - \bar{x}})^2$.
3) Se calcula la columna $\boldsymbol{fi〚(x_i - \bar{x})〛\ \hat{}2}$ y la suma $\boldsymbol{\sum fi(xi - x)^2}$.

4) Se aplica la fórmula:

$$\boldsymbol{s^2 = \frac{\sum f_i(x_i - \bar{x}))^2}{n}}$$

$$s^2 = 701.930/130 = 5.399{,}5$$

Ejemplo 2: Calcular la varianza y la desviación estándar de la siguiente distribución de los puntajes en una prueba de matemática correspondientes a 40 estudiantes:

$\boldsymbol{x_i}$	$\boldsymbol{f_i}$
10-20	5
20-30	7
30-40	15
40-50	5
50-60	8
	$\boldsymbol{n = 40}$

<u>Solución</u>:

Intervalos	Marca de clase ($\boldsymbol{x_i}$)	f_i	$\boldsymbol{x_i \cdot f_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$	$\boldsymbol{f_i(x_i - \bar{x})^2}$
10-20	15	5	75	-21	441	2.205
20-30	25	7	175	-11	121	847
30-40	35	15	525	-1	1	15
40-50	45	5	225	9	81	405
50-60	55	8	440	19	361	2.888
		$n = 40$	1.440			6.360

1) Se calcula la Ma. Se determina la columna $\boldsymbol{x_i \cdot f_i}$ y la suma de sus valores.

$$\boldsymbol{\bar{x} = \frac{\sum x_i \cdot f_i}{n}}$$

$$\bar{x} = \frac{1440}{40} = 36$$

2) Se calculan las desviaciones $x_i - \bar{x}$ y se elevan al cuadrado, $(x_i - \bar{x})^2$.

3) Se calcula la columna $f_i(x_i - Ma)^2$ y la suma $\sum f_i(x_i - \bar{x})^2$.

4) Se aplica la fórmula:

$$\boldsymbol{s^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n}}$$

$$s^2 = 6360/40 = 159$$

Si se extrae raíz cuadrada al valor calculado anteriormente, se obtiene la denominada desviación estándar $\boldsymbol{s} = \sqrt{159} = 12{,}6$, que se pasa a explicar a continuación.

4.3.5 Desviación estándar o típica:

Es una medida de dispersión que equivale a la raíz cuadrada del promedio de los cuadrados de las desviaciones de la variable con respecto a la media aritmética. En otras palabras, es la raíz cuadrada de la varianza. Se usa cuando la media aritmética es la medida de tendencia central elegida. Se simboliza por la letra (s) cuando está referida a una muestra y por la letra sigma (σ), cuando corresponde a una población.

$$\boldsymbol{s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}}$$

Si se desarrolla algebraicamente la fórmula anterior, es decir, resolviendo el cuadrado del binomio y teniendo presente que $\bar{\boldsymbol{x}} = \sum x_i/\text{n}$, se llega a una fórmula simplificada para la desviación estándar, que es:

$$\boldsymbol{s = \sqrt{\bar{x}{x_i}^2 - (\bar{\mathrm{x}}x_i)^2}}$$

Es decir, la desviación estándar es la raíz cuadrada de la diferencia entre la media aritmética de los cuadrados de la variable menos la media aritmética de los valores de la variable al cuadrado.

Para datos agrupados en una tabla de frecuencias tipo II, la fórmula es:

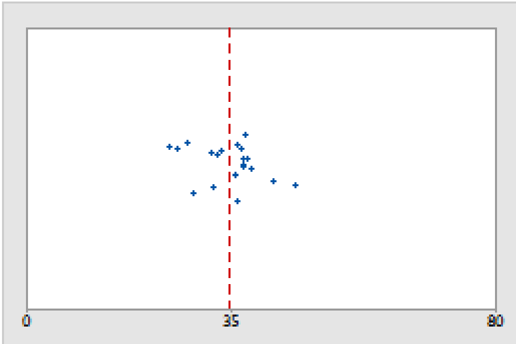
$$\boldsymbol{s} = \sqrt{\frac{\sum f_i(x_i - \bar{\boldsymbol{x}})^2}{\boldsymbol{n}}}$$

Para datos agrupados en una tabla de frecuencias tipo III (intervalar), la fórmula es la misma, solo que xi representa a la marca de clase del intervalo.

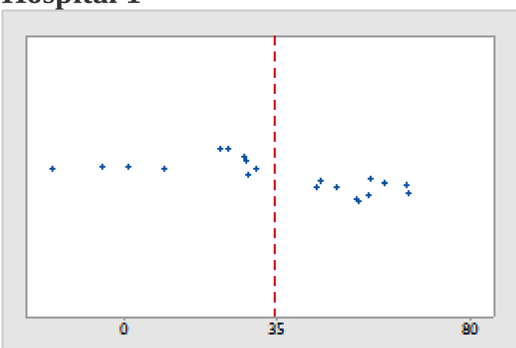
La desviación estándar indica que tan homogéneos son los datos con respecto a su media aritmética o qué tan concentrados a ella se encuentran. A mayor desviación estándar, mayor dispersión de los datos con respecto a la media; al contrario, a menor desviación estándar, menor dispersión.

4.3.5.1 Ejemplo de interpretación de la desviación estándar (tomado de: https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/store-descriptive-statistics/interpret-the-statistics/interpret-the-statistics/

Tiempos de egreso de un hospital:



Hospital 1



Hospital 2

Tiempos de egreso en un hospital:

Los administradores dan seguimiento al tiempo de egreso de los pacientes que son tratados en las áreas de urgencia de dos hospitales. Aunque los tiempos de egreso promedio son aproximadamente iguales (35 minutos), las desviaciones estándar son significativamente diferentes. La desviación estándar del hospital 1 es de aproximadamente 6. En promedio, el tiempo para dar de alta a un paciente se desvía de la media (línea discontinua) aproximadamente 6 minutos. La desviación estándar del hospital 2 es de aproximadamente 20. En promedio, el tiempo para dar de alta a un paciente se desvía de la media (línea discontinua) aproximadamente 20 minutos.

4.3.5.2 Propiedades de la desviación estándar:

- Si se suma un cierto número a cada uno de los elementos de un conjunto, su desviación estándar no varía.
- Si cada uno de los elementos de un conjunto se multiplica por un mismo número, la desviación estándar queda multiplicada por ese número.
- La desviación estándar es siempre positiva o cero. Será cero en el caso que las puntuaciones sean iguales.
- La desviación estándar, al igual que la media aritmética y la varianza, es muy sensible a las puntuaciones extremas.

Ejemplo1: Calcular la desviación típica del siguiente conjunto de datos: 3-8-9-10-11-13

1) Se calcula la $\bar{\boldsymbol{x}}$:

$$\bar{\boldsymbol{x}} = \frac{\sum \boldsymbol{x}_i}{\boldsymbol{n}}$$

$$\bar{x} = \frac{3+8+9+10+11+13}{6} = \frac{54}{6} = 9$$

2) Se calcula la columna $(\boldsymbol{xi} - \bar{\boldsymbol{x}})$, la columna $(x_i - \bar{\boldsymbol{x}})^2$ y se obtiene $\sum(x_i - \bar{\boldsymbol{x}})^2$:

$\boldsymbol{x}_i$	$\boldsymbol{x}_i - \bar{\boldsymbol{x}}$	$(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^2$
3	-6	36
8	-1	1
9	0	0
10	1	1
11	2	4
13	4	16
n=6		$\sum(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^2$=58

3) Se aplica la fórmula:

$$\boldsymbol{s} = \sqrt{\frac{\sum(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^2}{\boldsymbol{n}}}$$

$$s = \sqrt{58/6} = \sqrt{9{,}66} = 3{,}1$$

Es decir, la distancia promedio de los datos de la variable con respecto a su media aritmética es 3,1.

Si se quiere utilizar la fórmula simplificada de la desviación estándar, habría que presentar la información como sigue:

x_i	${x_i}^2$
3	9
8	64
9	81
10	100
11	121
13	169
$n = 6$	544

$$\bar{x}{x_i}^2 = \frac{544}{6} = 91$$

$$\bar{x}_{x_i} = \frac{3+8+9+10+11+13}{6} = \frac{54}{6} = 9$$

$$(\bar{x}_{x_i})^2 = 9 \cdot 9 = 81$$

$$\boldsymbol{s} = \sqrt{\bar{x}{x_i}^2 - (\bar{x}_{x_i})^2}$$

$$s = \sqrt{91 - 81} = \sqrt{10} = 3{,}1$$

Ejemplo2: Calcular la desviación típica de la siguiente distribución:

x_i	f_i
1	5
2	5
4	7
5	4
8	3
13	1
	$n = 25$

Solución:

x_i	f_i	$x_i \cdot f_i$	$x_i - \boldsymbol{\bar{x}}$	$(x_i - \boldsymbol{\bar{x}})^{\mathbf{2}}$	$f_i(x_i - \boldsymbol{\bar{x}})^{\mathbf{2}}$
1	5	5	-3	9	45
2	5	10	-2	4	20
4	7	28	0	0	0
5	4	20	1	1	4
8	3	24	4	16	48
13	1	13	9	81	81
	$n = 25$	$\sum x_i \cdot f_i = \mathbf{100}$			$\sum f_i(x_i - \boldsymbol{\bar{x}})^{\mathbf{2}} = 198$

1) Se calcula la $\boldsymbol{\bar{x}}$:

$$\boldsymbol{\bar{x}} = \frac{\sum \boldsymbol{x_i \cdot f_i}}{\boldsymbol{n}}$$

$$M = \frac{100}{25} = 4$$

2) Se calculan las desviaciones x_i-$\boldsymbol{\bar{x}}$ y se elevan al cuadrado, $(x_i-\boldsymbol{\bar{x}})^2$.

3) Se calcula la columna $f_i(x_i-\boldsymbol{\bar{x}})^2$ y la suma $\sum f_i\ (x_i-\boldsymbol{\bar{x}})^2$.

4) Se aplica la fórmula:

$$s = \sqrt{\frac{\sum f_i (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{\boldsymbol{n}}}$$

$$s = \sqrt{198/25} = \sqrt{7{,}92} = 2{,}81$$

Es decir, la distancia promedio de los datos de la variable con respecto a su media aritmética es 2,81.

Ejemplo 3: En una encuesta, a una parte de la población se le hizo la siguiente pregunta: ¿cuántos hijos desearía tener usted si formara una familia? Las respuestas tabuladas fueron las siguientes. (Fuente: Libro de matemáticas de 3º medio de Saiz y Blumenthal).

Número de hijos x_i	Número de personas f_i
0	12
1	87
2	95
3	46
4	21
5	4

Determinar:

a) La media de la muestra
b) La varianza de la muestra
c) La desviación estándar de la muestra
d) El coeficiente de variación de la muestra
e) Agrega en la tabla una columna que corresponda a la frecuencia relativa de los datos.

Solución:

x_i	f_i	$x_i \cdot f_i$	$x_i - \overline{\boldsymbol{x}}$	$(x_i - \overline{\boldsymbol{x}})^2$	$f_i(x_i - \overline{\boldsymbol{x}})^2$
0	12	0	-1,96	3,8416	46,0992
1	87	87	-0,96	0,9216	80,1792
2	95	190	0,04	0,0016	0,1520
3	46	138	1,04	1,0816	49,7536
4	21	84	2,04	4,1616	87,3936
5	4	20	3,04	9,2416	36,9660
	$n = 265$	$\sum x_i \cdot f_i = 519$		$\sum (xi - \bar{x})^2 = 19{,}2496$	$\sum f_i(x_i - \bar{x})^2$=300,5436

a) La media aritmética $\overline{\boldsymbol{x}}$:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{n}$$

$$M_a = \frac{519}{265} = 1{,}96$$

b) La varianza:

$$\boldsymbol{s^2 = \frac{\sum f_i(x_i - \bar{x})^2}{n}}$$

$$s^2 = \frac{300{,}5436}{265} = 1{,}13$$

c) La desviación estándar:

$$\boldsymbol{s = \sqrt{\frac{\sum f_i(xi - \bar{x})^2}{n}}}$$

$$s = \sqrt{1{,}13} = 1{,}06$$

d) Coeficiente de variación: (ver el concepto más adelante)

$$\boldsymbol{CV = \frac{s}{\bar{x}}}$$

$$CV = \frac{1{,}06}{1{,}96} = 0{,}5408 = 54{,}08\%$$

e) Frecuencia relativa:

x_i	f_i	Frecuencia relativa h_i
0	12	12/265
1	87	87/265
2	95	95/265
3	46	56/265
4	21	21/265
5	4	4/265

<u>Ejemplo 4</u>: Calcular la desviación típica de la siguiente distribución:

$Intervalos$	Marca de clase x_i	f_i
1-5	3	4
5-9	7	8
9-13	11	9
13-17	15	10
17-21	19	6
21-25	23	5
		$n = 42$

Solución:

Intervalos	Marca de clase x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
1-5	3	4	12	-10	100	400
5-9	7	8	56	-6	36	288
9-13	11	9	99	-2	4	36
13-17	15	10	150	2	4	40
17-21	19	6	114	6	36	216
21-25	23	5	115	10	100	500
		$n = 42$	546			1480

1) Se calcula la Ma. Se determina la columna xi*fi y la suma de sus valores.

$$\bar{\boldsymbol{x}} = \frac{\sum x_i \cdot f_i}{\boldsymbol{n}}$$

$$\bar{x} = \frac{546}{42} = 13$$

2) Se calculan las desviaciones $x_i - \bar{\boldsymbol{x}}$ y se elevan al cuadrado, $(x_i - \bar{\boldsymbol{x}})^2$.

3) Se calcula la columna $f_i(x_i - \bar{\boldsymbol{x}})^2$ y la suma $\sum f_i(x_i - \bar{x})^2$.

4) Se aplica la fórmula:

$$s = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{\boldsymbol{n}}}$$

$$s = \sqrt{1480/42} = \sqrt{35{,}238} = 5{,}93$$

Es decir, la distancia promedio de los datos de la variable con respecto a su media aritmética es 5,93.

Ejemplo 5: Calcular la desviación típica de la siguiente distribución de puntajes de acceso a la universidad correspondientes a 100 estudiantes:

$Intervalos$	Marca de clase x_i	f_i
400-450	425	10
450-500	475	20
500-550	525	30
550-600	575	40
600-650	625	15
650-700	675	10
700-750	725	5
		$n = 130$

Solución:

Intervalos	Marca de clase x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{\boldsymbol{x}}$	$(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$	$f_i(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$
400-450	425	10	4250	131	17161	171610
450-500	475	20	9500	81	6561	131220
500-550	525	30	15750	31	961	28830
550-600	575	40	23000	19	361	14440
600-650	625	15	9375	69	4761	71415
650-700	675	10	6750	119	14161	141610
700-750	725	5	3625	169	28561	142805
		$n = 130$	72250			701930

1) Se calcula la Ma. Se determina la columna xi*fi y la suma de sus valores.

$$\bar{\boldsymbol{x}} = \frac{\sum x_i \cdot f_i}{\boldsymbol{n}}$$

$$\bar{x} = \frac{72250}{130} = 555{,}8 \approx 556$$

2) Se calculan las desviaciones $x_i - \bar{\boldsymbol{x}}$ y se elevan al cuadrado, $(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$.

3) Se calcula la columna $f_i(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$ y la suma $\sum f_i(x_i - \bar{x})^2$.

4) Se aplica la fórmula:

$$s = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{\boldsymbol{n}}}$$

$$s = \sqrt{701930/130} = \sqrt{5399{,}5} = 73{,}48$$

Es decir, la desviación de los puntajes de la variable con respecto a su media aritmética es 73,48 puntos.

Ejemplo 6: Las alturas de los jugadores de un equipo de baloncesto vienen dadas por la tabla siguiente:

Calcular:
a) La media de las alturas
b) La varianza
c) La desviación estándar
d) El coeficiente de variación

Fuente: Libro de matemática de 3º medio de Saiz y Blumenthal

$Intervalos$	Marca de clase x_i	f_i
[170-175[	172,5	1
[175-180[	177,5	3
[180-185[	182,5	4
[185-190[	187,5	8
[190-195[	192,5	5
[195-200[	197,5	2

Solución:

Intervalos	Marca de clase x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{\boldsymbol{x}}$	$(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$	$f_i(x_i - \bar{\boldsymbol{x}})^{\mathbf{2}}$
[170-175[	172,5	1	172,5	-14,1	198,81	198,81
[175-180[	177,5	3	532,5	-9,1	82,81	248,43
[180-185[	182,5	4	730,0	-4,1	16,81	67,24
[185-190[	187,5	8	1500	0,9	0,81	6,48
[190-195[	192,5	5	962,5	5,9	34,81	174,05
[195-200[	197,5	2	395,0	10,9	118,81	237,62
		$n = 23$	4292,5			932,63

a) La media de las alturas:

$$\bar{\boldsymbol{x}} = \frac{\sum x_i \cdot f_i}{\boldsymbol{n}}$$

$$\bar{x} = \frac{4292{,}5}{23} = 186{,}6$$

b) La varianza:

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n}$$

$$s^2 = \frac{932{,}63}{23} = 40{,}55$$

c) La desviación estándar:

s = $\sqrt{\frac{\sum \boldsymbol{fi(xi - \bar{x})^2}}{\boldsymbol{n}}}$

s = $\sqrt{\frac{\mathbf{932{,}63}}{\mathbf{23}}} = \mathbf{6{,}37}$

O bien: $\sqrt{\mathbf{40{,}55}} = \mathbf{6{,}37}$

Ejemplo 7: Un director de escuela quiere conocer el nivel de aprendizaje en matemática de los dos 2º años medios de su establecimiento. Para una mayor facilidad en los cálculos se supondrá cursos de solo 5 alumnos. La información de que dispone es la siguiente:

2º medio A	2º medio B
6	10
7	10
8	10
7	2
6	2

El director calcula la media aritmética y observa que en ambos cursos el promedio de las calificaciones es el mismo.

$$M_a\text{A} = \frac{6+7+8+7+6}{5} = \frac{34}{5} = 6{,}8$$

$$M_a\text{B} = \frac{10+10+10+2+2}{5} = \frac{34}{5} = 6{,}8$$

Pero al calcular la desviación estándar observa lo siguiente:

$S_A = 0{,}84$
$S_B = 4{,}38$

Es decir, pese a que ambos cursos alcanzaron un promedio igual a 6,8, los alumnos del 2º medio B presentan una mayor variabilidad en sus calificaciones, comparativamente con el 2º medio A, donde son más homogéneas.

4.3.6 Coeficiente de variación:

$$CV = \frac{s}{\bar{x}}$$

$$V = \frac{6{,}37}{186{,}6} = 0{,}034 = 3{,}41\%$$

Ejercicios propuestos:

Ejercicio: Se realizó una encuesta para determinar el número de libros leídos por un grupo de estudiantes. Dicho número fue el siguiente:

3	14	9	17	1
22	12	8	9	15
16	8	13	20	12
4	15	24	18	10
6	11	18	15	25

Completar la siguiente tabla de distribución con las columnas necesarias para calcular:

a) La media aritmética
b) La mediana
c) La varianza
d) La desviación estándar

Intervalos	Marca de clase x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1-5					

4.3.7 Coeficiente de variabilidad o Coeficiente de variación

Así como fue necesario complementar la información entregada por los estadígrafos de posición (media aritmética, por ejemplo) con los estadígrafos de dispersión (desviación estándar o varianza, por ejemplo), también se hace necesario complementar estos últimos con el denominado Coeficiente de variabilidad. Esto es necesario porque puede darse el caso de dos distribuciones que muestren una misma dispersión en torno a la media, es decir, tengan igual varianza y desviación estándar, pero, en términos relativos, ser una de ellas más homogénea que la otra.

Se define como el cociente entre la desviación estándar de la distribución y su media aritmética. Si los valores provienen de una muestra, la fórmula es:

$$\boldsymbol{CV = \frac{s}{\bar{x}}}$$

Si los valores se originan en una población, se usa σ en vez de s.

Ejemplo: Supóngase una distribución de ingresos A cuya $Ma = 150$ y sus valores extremos sean 100 y 200, y una distribución de ingresos B con una $M_a = 1050$, pero con valores extremos 1000 y 1100. Si ambas distribuciones tuvieran, por ejemplo, una desviación estándar de 60, los respectivos coeficientes de variabilidad serían:

$$CV(A) = \frac{s}{\bar{x}} = \frac{60}{150} = 0{,}40 = 40\%$$

$$CV(B) = \frac{s}{\bar{x}} = \frac{60}{1050} = 0{,}057 = 5{,}7\%$$

Se aprecia, en consecuencia, que la distribución de ingresos B es más homogénea que la A, ya que su variabilidad es 5,7%, menor que el 40% de la distribución A.

5 PUNTUACIONES TÍPICAS

Estas puntuaciones y sus escalas permiten expresar cualquier puntuación en términos de su distancia desde la media aritmética, en unidades de desviación estándar. Se hará mención a dos de ellas, la puntuación sigma (z) (z minúscula) y la puntuación típica Z (Z mayúscula).

5.1 Puntuación sigma z: (letra z en minúscula)

La variable $\boldsymbol{z = \frac{x_i - \bar{x}}{s}}$, que mide la desviación de los valores de la variable con respecto a la media aritmética en unidades de desviación estándar, se llama puntuación sigma o variable normalizada. En otras palabras, los puntajes z son las distancias medidas en desviaciones estándar de cada puntuación con respecto a su promedio. Si el puntaje z tiene signo positivo, significa que la puntuación en la distribución se encuentra sobre la media aritmética, y si tiene signo negativo, que la puntuación se halla bajo esa media.

Cuando los puntajes originales x_i de una variable continua se expresan en términos relativos a la desviación estándar de la distribución, es decir, cuando todas las puntuaciones de una distribución se transforman en sus respectivos puntajes z, se tiene una curva normal tipificada, con media aritmética cero y desviación estándar 1. Se simboliza $\boldsymbol{N(0,1)}$. Los puntajes z son útiles porque reducen puntuaciones de distribuciones diferentes a una unidad de medida común y comparable. Se emplean frecuentemente en evaluaciones del desempeño académico de los estudiantes.

Ejemplo 1: Si en la curva normal de los puntajes de la PSU, que se sabe tiene una media aritmética de 500 puntos y una desviación estándar de 110 (μ=500 u σ=110), un alumno obtiene 610 puntos, el puntaje z equivalente es:

$$\boldsymbol{z = \frac{x_i - \bar{x}}{s}}$$

$$z = \frac{610 - 500}{110} = \frac{110}{110} = 1$$

Que z sea igual a 1 significa que el puntaje de 610 puntos se desvió 110 puntos de la media es una desviación estándar por encima de la media aritmética ($1 \cdot 110 = 110$), que es precisamente el valor de sigma.

Si el puntaje fuera 720 puntos:

$$z = \frac{x_i - \bar{x}}{s}$$

$$z = \frac{720 - 500}{110} = \frac{220}{110} = 2$$

Que z sea igual a 2 significa que el puntaje de 720 puntos se desvió 220 puntos de la media, es decir, dos desviaciones estándar por encima de la media aritmética ($2 \cdot 110 = 220$).

Ejemplo 2: Un estudiante obtuvo en la prueba de Biología un 5,8 y en la de Química un 6,2. Si en la primera el promedio fue un 5,0 con una desviación estándar 0,5 y, en la segunda, la media fue un 5,5 con desviación estándar 0,7, ¿en cuál de las dos pruebas el estudiante estuvo mejor ubicado?

Un primer análisis dice que le fue mejor en Química, ya que 6,2 > 5,8. Ahora, si se tiene en cuenta los promedios en cada asignatura, se diría que en Biología le fue mejor ya que estuvo 0,8 puntos por encima de la media (5,8-5,0=0,8) y en Química 0,7 puntos (6,2-5,5=0,7). Pero estas conclusiones nada dicen de la ubicación del estudiante con respecto al grupo clase, ya que ello depende del grado de concentración o dispersión de las calificaciones con respecto a cada promedio, es decir, de la desviación estándar. Para incluir este factor en la respuesta a la pregunta, se calculan los puntajes z.

El puntaje z en Biología es:

$$zB = \frac{5{,}8 - 5}{0{,}5} = \frac{0{,}8}{0{,}5} = 1{,}6$$

El puntaje z en Química es:

$$zQ = \frac{6{,}2 - 5{,}5}{0{,}7} = \frac{0{,}7}{0{,}7} = 1$$

Es decir, en Biología, la desviación de la nota de 0,8 décimas corresponde a 1,6 sigmas ($\mathbf{1,6 \cdot 0,5 = 0,8}$), es decir, el alumno estuvo 1,6 sigmas por encima del promedio; en Química, en tanto, la desviación de la nota de 0,7 décimas corresponde a 1 sigma ($1 \cdot 0,7 = 0,7$), es decir, el alumno estuvo 1 sigma por encima del promedio. En consecuencia, estuvo mejor ubicado en Biología, pese a que su nota en esa asignatura fue inferior a la conseguida en Química.

Ejemplo 3: Un estudiante obtuvo en la prueba de Ciencias Sociales 82 puntos, siendo el puntaje promedio del curso en esa asignatura de 74 puntos con una desviación estándar de 10. En el examen de matemática su puntaje fue de 88 con una media de 80 puntos y una desviación de 14. ¿En qué asignatura obtuvo una mejor posición relativa?

Solución:

z para Ciencias Sociales:

$$z_{cs} = \frac{x_i - \bar{x}}{s} = \frac{82 - 74}{10} = 0,8$$

z para Matemática:

$$z_m = \frac{x_i - \bar{x}}{s} = \frac{88 - 80}{14} = 0,57$$

El estudiante obtuvo una puntuación de 0,8 de desviación estándar por encima de la media aritmética en ciencias sociales, pero solo de 0,57 de desviación estándar por encima de la media en matemática. Por lo tanto su posición relativa fue superior en ciencias sociales.

Ejercicio propuesto 1: En un examen final de estadística, la puntuación media de un grupo de 150 estudiantes fue de 78 y la desviación típica fue de 8. En álgebra, la media final del grupo fue de 73 y la desviación típica 7,6. En ese grupo, un estudiante obtuvo 75 puntos en estadística y 71 puntos en álgebra ¿En qué asignatura su puntuación relativa fue mayor? Respuesta: en álgebra.

Ejercicio propuesto 2: En un examen final de estadística, la puntuación media fue 4,4 y la desviación típica fue de 0,34. En matemática, la media fue de 5,3 y la desviación típica 0,57. Si el estudiante A obtuvo 6,4 en matemática y B un 5,5 en estadística, ¿quién se desempeñó relativamente mejor? Respuesta: el alumno B.

5.2 Puntuación típica Z (letra Z en mayúscula)

Con el propósito de eliminar los frecuentes puntajes z negativos y fraccionarios, se utilizan para su determinación promedios y desviaciones estándar previamente determinados.

Una de las transformaciones utilizadas habitualmente es:

$$\boldsymbol{Z = 10 \cdot (\frac{x_i - \bar{x}}{s}) + 50}$$

$$\boldsymbol{Z = 10 \cdot z + 50}$$

Donde la media de la distribución es siempre 50 y la desviación estándar es 10.

Ejemplo: Las puntuaciones de un estudiante en dos asignatura fue la siguiente:

Estudiante	Puntuación	Media aritmética	Desviación estándar
Asignatura A	40	47	5
Asignatura B	84	110	20

Se pide calcular la puntuación z y Z.

Solución:

$$z_A = \frac{x_i - \bar{x}}{s} = \frac{40 - 47}{5} = -\frac{7}{5} = -1{,}4$$

$$Z_A = 10 \cdot z + 50 = 10 \cdot (-1{,}4) + 50 = -14 + 50 = 36$$

$$z_B = \frac{x_i - \bar{x}}{s} = \frac{84 - 110}{20} = -\frac{26}{20} = -1{,}3$$

$$Z_B = 10 \cdot z + 50 = 10 \cdot (-1{,}3) + 50 = -13 + 50 = 37$$

Las puntuaciones z y Z en las dos asignaturas muestran un nivel de rendimiento muy similar.
Las puntuaciones Z se utilizan para medir el Coeficiente Intelectual CI, cuya media es 100 y la desviación estándar es 10, y la que se empleaba en la Prueba de Selección Universitaria PSU, donde la media era 500 y la desviación estándar era 110.

6 DISTRIBUCIÓN NORMAL

6.1 Conceptos básicos:

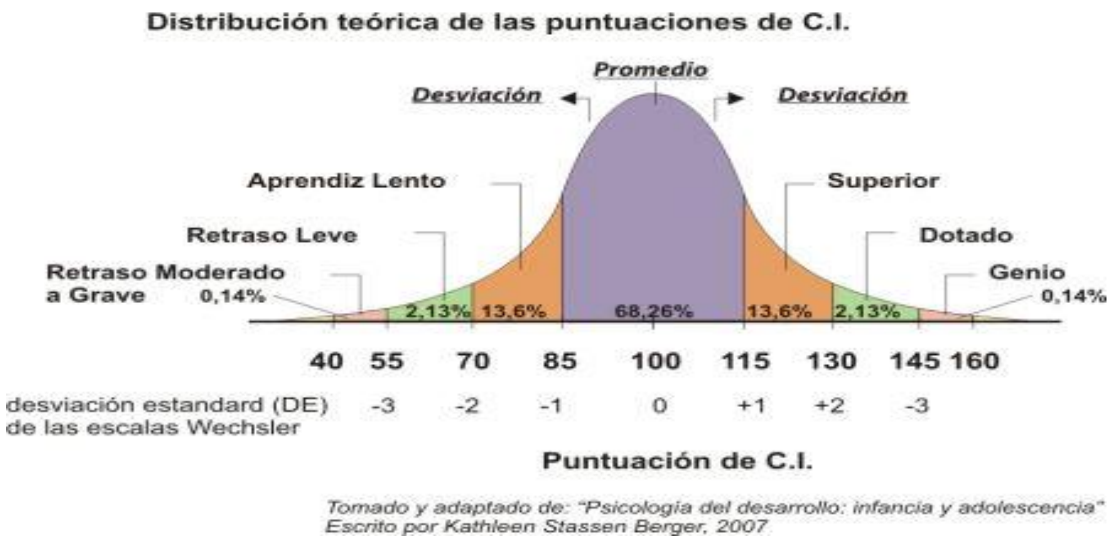
6.1.1 Variable aleatoria continua:

Es la que puede tomar cualquier valor numérico en un intervalo o conjunto de intervalos. Por ejemplo: la estatura o el peso de una persona.

6.1.2 Distribución normal, curva normal o campana de Gauss:

Es una distribución de frecuencias donde la mayoría de los casos se presentan en el centro en torno a la media aritmética y los valores restantes van decreciendo de manera similar hacia ambos lados. Se forma, en definitiva, una curva simétrica.

Es muy frecuente que los valores de las variables en estudio se distribuyan normalmente, es decir, que exista un gran número de distribuciones normales, como los puntajes de la Prueba de Selección Universitaria (PSU), la altura y el peso de las personas, el coeficiente intelectual (IQ o CI), etc. Para este último, la curva normal tiene una media aritmética de 100 y una desviación estándar de 15.

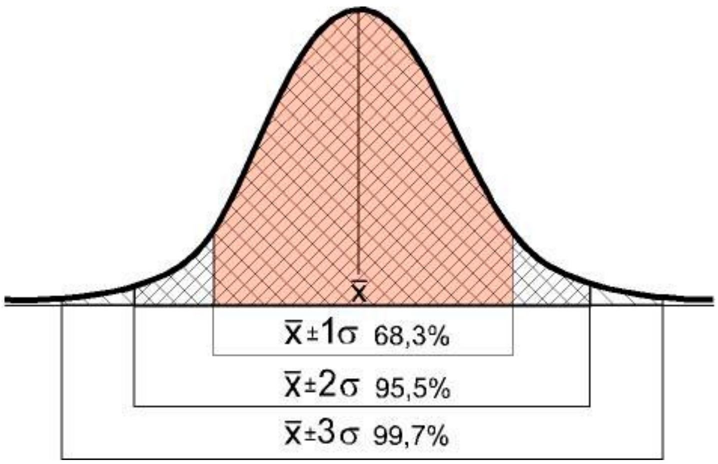


Sin embargo, el proceso conocido como de tipificación, permite reducir las curvas normales a una sola con media aritmética 0 y desviación estándar 1, que se denomina distribución normal estándar o tipificada y se representa por $N(0,1)$.
En términos de probabilidad, el área bajo esta curva es igual a 1 o 100% y, dada la simetría de la curva normal, el área a la derecha de la media es 0,5 o 50% y a la izquierda también 0,5 o 50%; además, el área desde la media a un valor v (v positivo) cualquiera es igual al área desde la media al valor –v (v negativo).

Las características de una distribución normal teórica son las siguientes:

- Es simétrica con respecto al eje y.
- Es un modelo de distribución de frecuencias para variables continuas, por ejemplo, puntajes PSU, coeficiente intelectual, altura de las personas, pesos de las personas, etc.
- Coinciden la media aritmética, la mediana y la moda. Por lo tanto, existe un número igual de puntuaciones a cada lado de la media.
- Las frecuencias disminuyen simétricamente para valores mayores y menores.
- Es asintótica al eje de las abscisas hacia ambos lados (sus extremos se acercan al eje, pero no lo cortan).

- Tiene dos puntos donde la curva cambia de convexa a cóncava que denominan puntos de inflexión. Dichos puntos están situados a una unidad de distancia o de desviación de la media aritmética.



La desviación estándar de la distribución normal es un parámetro de la curva normal con propiedades importantes derivadas de su relación con las puntuaciones y el número de casos. El área achurada de la imagen anterior representa el número total de casos de la distribución. Tomando ambos lados de la curva, el área comprendida entre:

- Entre $\bar{x}$ - 1σ y $\bar{x} + 1\sigma$ quedan comprendidos el 68,26% de los valores de una variable aleatoria normal.
- Entre $\bar{x}$ - 2σ y $\bar{x} + 2\sigma = 95{,}45$%.
- Entre $\bar{x}$ - 3σ y $\bar{x} + 3\sigma = 99{,}73$%.

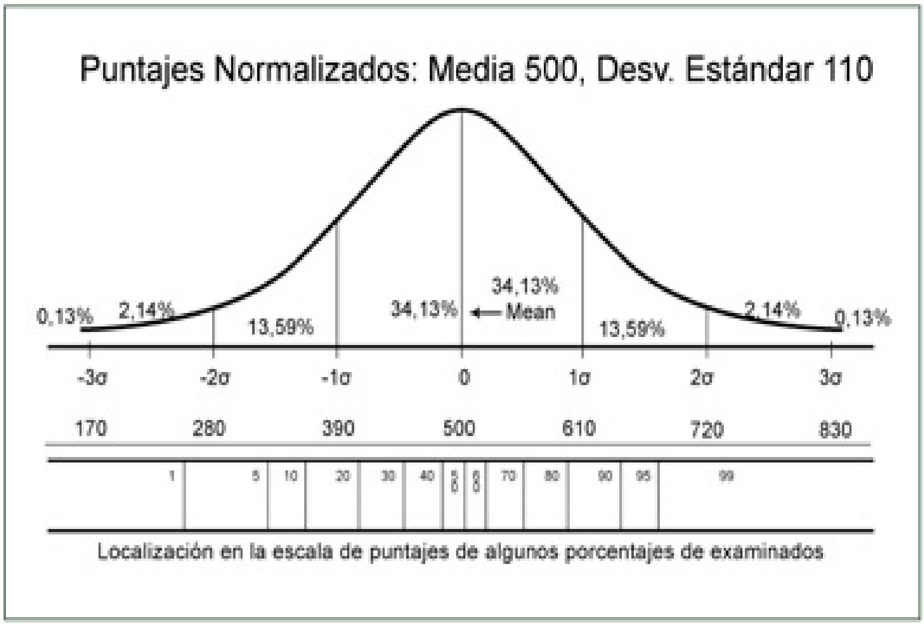
Tomando un solo lado de la curva, ya que es simétrica, se tiene que:

- Entre $\bar{x}$ y $\bar{x} + 1\sigma = 34{,}13$%.

- Entre $\bar{x}$ y $\bar{x} + 2\sigma = 47{,}72$%. $(34{,}13\% + 13{,}59\% = 47{,}72\%)$

- Entre $\bar{x}$ y $\bar{x} + 3\sigma = 49{,}86$%. $(34{,}13\% + 13{,}59\% + 2{,}14\% = 49{,}86\%)$

En la curva de Gauss correspondiente a puntajes PSU normalizados, con una media de 500 y una desviación estándar de 110, se pueden observar los porcentajes bajo la curva señalados anteriormente.



Por ejemplo, si las estaturas de 1.000 personas se distribuyen según una curva normal cuya media es 165 cm y su desviación estándar es 6, entonces, ¿cuántas personas miden entre $\bar{x}-3\sigma$ y $\bar{x}+3\sigma,$ entre $\bar{x}-2\sigma$ y $\bar{x}+2\sigma,$ entre $\bar{x}-1\sigma$ y $\bar{x}+1\sigma$?

$\bar{x}-3\sigma=165-3\sigma=165-18=147$ y $\bar{x}+3\sigma=165+3\sigma=165+18=183$. Por lo tanto, como entre $\bar{x}-3\sigma$ y $\bar{x}+3\sigma$ está el 99,74% de los casos, hay 99,74% de 1000 $(1000\cdot 0{,}9973=997)$ personas que miden entre 147 cm y 183 cm.

$\bar{x}-2\sigma=165-12=153$ y$\bar{x}+2\sigma=165+12=177$. Por tanto, como entre $\bar{x}$ - 2σ y $\bar{x}$ + 2σ está el 95,44% de los casos, hay $95{,}44\%\ de\ 1000=954$ personas que miden entre 153 cm y 177 cm.

$\bar{x}-1\sigma=165-6=159$ y $\bar{x}+1\sigma=165+6=171$. Por lo tanto, como entre $\bar{x}$ - 1σ y $\bar{x}$ + 1σ está el 68,26% de los casos, hay 68,26% de 1.000 $=683$ personas que miden entre 159 cm y 171 cm.

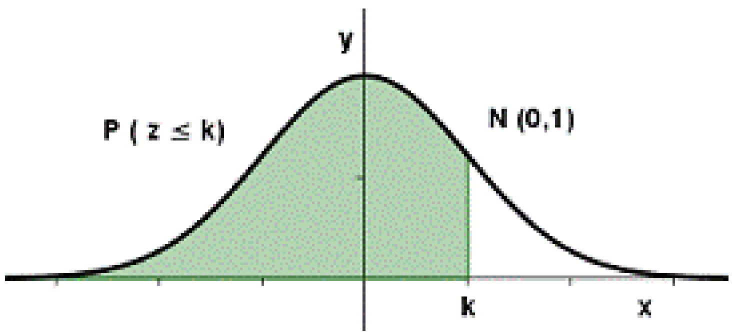
6.2 Cálculo de probabilidades en distribuciones normales estándar o tipificada

La importancia de la curva normal tipificada radica en que es un modelo de probabilidad, donde el área bajo la curva vale 1 o 100%. Como se ha dicho, se la denomina $\boldsymbol{N(0,1)}$, que significa que es una distribución normal cuya media aritmética es igual a 0 y su desviación estándar es igual a 1. Para encontrar áreas específicas bajo la curva se utilizan tablas de distribución normal tipificadas, cuyos valores están expresados en términos de desviaciones estándar de la media. Son tablas de doble entrada en las cuales la primera columna muestra unidades y décimas y la primera fila, las centésimas de las puntuaciones z.

Se utilizan los dos tipos de tablas siguientes:

6.2.1 Tabla de distribución normal tipificada $N(0,1)$, que proporciona, para cada valor de z, el área que queda a su izquierda (Tabla I).

6.2.2 Tabla de distribución normal tipificada $N(0,1)$, que proporciona el área comprendida entre 0 y z (Tabla II).



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8930
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9561	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9934	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9901	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9954	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

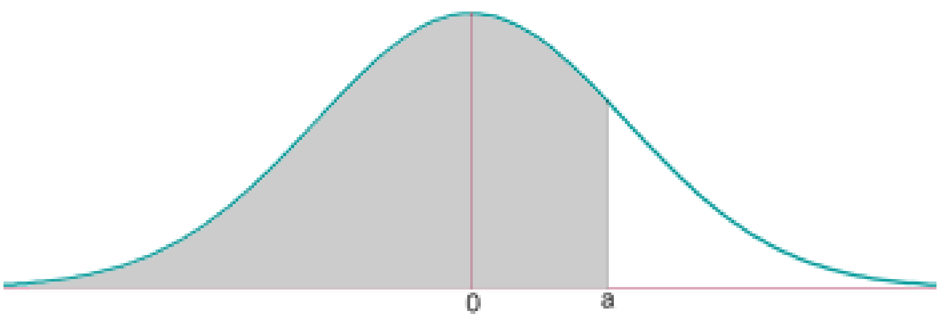
Tabla de distribución normal tipificada $N(0,1)$, que proporciona el área comprendida entre 0 y z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Para los efectos de utilizar las tablas anteriores, hay que tener presente las siguientes relaciones:

Caso 1: $P(z \leq a)$

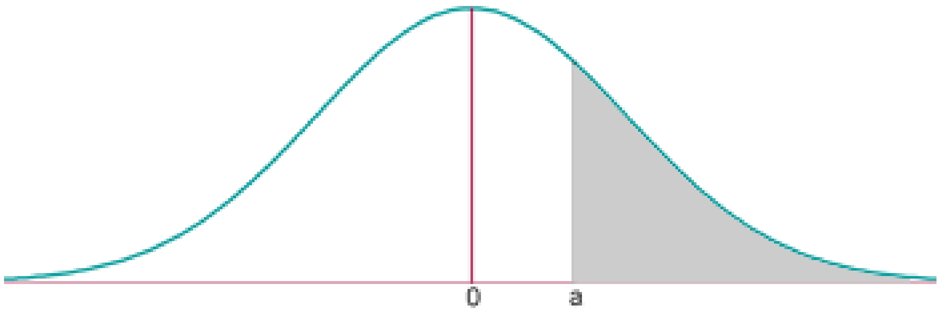


Ejemplo: Haciendo uso de la tabla I, determinar $\boldsymbol{P(z \leq 1,25)}$, es decir, determinar la probabilidad que z sea menor o igual que 1,25.

En la tabla I de distribución normal tipificada, en la primera columna se ubica el valor 1,2 y en la primera fila el valor 0,05, ya que $1{,}2 + 0{,}05 = 1{,}25$. La probabilidad es el número que aparece en la intersección y es 0,8944 u 89,44%.

z	0,01	0,02	0,03	0,04	0,05	0,06	
1,2					0,8944		

Caso 2: $\boldsymbol{P(z > a) = 1 - P(z \leq a)}$



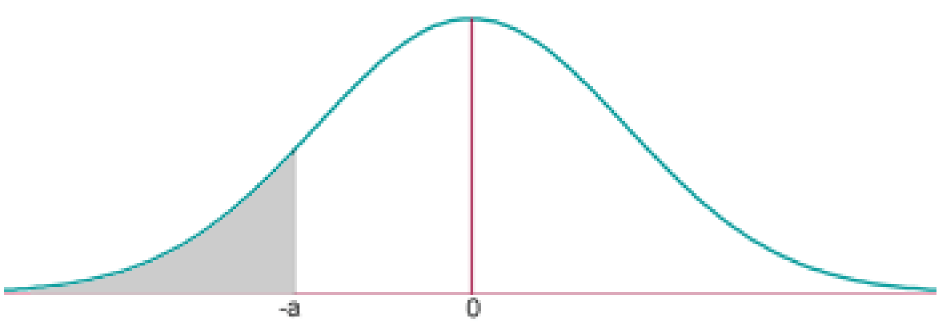
Ejemplo: Determinar $\boldsymbol{P(z \geq 1,43)}$, es decir, determinar la probabilidad que z sea mayor o igual que 1,43.

En este ejemplo, a diferencia del anterior, se está buscando el área de un valor de z que es mayor, no menor a un cierto valor. Para resolver el ejercicio se considera que el área total bajo la curva es 1, por lo que el área a la derecha de z=1,43 se obtiene como diferencia entre 1 – $P(z \leq 1{,}43)$. La probabilidad que $P(z \geq 1{,}43)$ se obtiene en la intersección del valor 1,4 en la primera columna y 0,03 en la primera fila de la tabla I, ya que 1,4+0,03=1,43. La probabilidad de $P(z \leq 1{,}43)$ es 0,9236.

z	0,01	0,02	0,03	0,04	0,05	0,06	
1,4			0,9236				

Y la probabilidad de $\boldsymbol{P(z \geq 1,43) = 1 - P(z \leq 1,43) = 1 - 0,9236 = 0,0764}$ ó $\boldsymbol{7,64}$%

$\boldsymbol{Caso}$ 3: $\boldsymbol{P(z \leq -a) = 1 - P(z \leq a)}$



Ejemplo: Determinar $\boldsymbol{P(z \leq -0,92)}$, es decir, determinar la probabilidad que z sea menor o igual que -0,92.

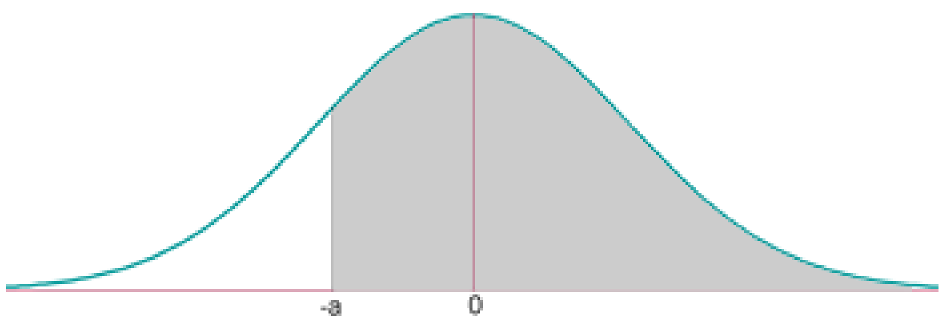
Lo distintivo de este ejemplo es que se está buscando el área de un valor de z que es menor o igual que un valor negativo. Considerando que por simetría de la curva el área a la izquierda de -0,92 es igual al área a la derecha de 0,92:

$\boldsymbol{P}(z \leq -0{,}92) = 1 - P(z \leq 0{,}92) = 1 - 0{,}8212 = 0{,}1788$ ó 17,88%

La probabilidad que $P(z \leq 0{,}92)$ se obtuvo en la intersección del valor 0,9 en la primera columna y 0,02 en la primera fila de la tabla I, ya que $0{,}9 + 0{,}02 = 0{,}92$. La probabilidad de $P(z \leq 0{,}92)$ es 0,8212.

z	0,01	0,02	0,03	0,04	0,05	0,06	
0,9		0,8212					

$\boldsymbol{P}(\boldsymbol{z} > -\boldsymbol{a}) = \boldsymbol{P}(\boldsymbol{z} \leq \boldsymbol{a})$



Ejemplo: Determinar $\boldsymbol{P}(\boldsymbol{z} \geq -\mathbf{2{,}7})$, es decir, determinar la probabilidad que z sea mayor o igual que -2,7.

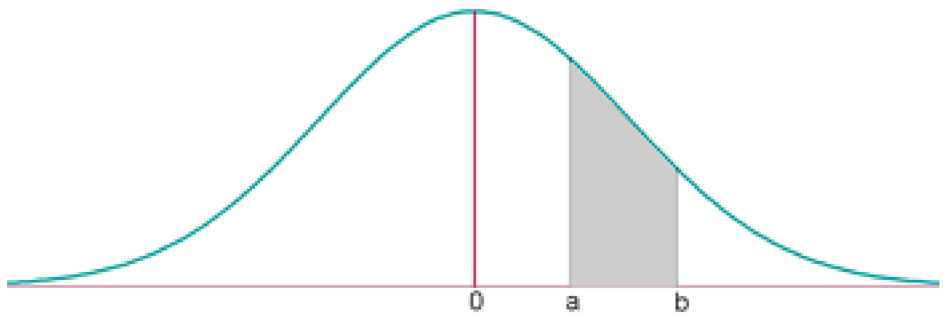
Como la curva es simétrica, el área a la derecha de -2,7 es igual al área a la izquierda de 2,7, entonces,

$P(z \geq -2{,}7) = P(z \leq 2{,}7) = 0{,}9965$

La probabilidad que $\boldsymbol{P}(\boldsymbol{z} \leq \mathbf{2{,}7})$ se obtuvo en la intersección del valor 2,7 en la primera columna y 0,00 en la primera fila, ya que $\mathbf{2{,}7 + 0{,}00 = 2{,}70}$. La probabilidad de $\boldsymbol{P}(\boldsymbol{z} \leq \mathbf{2{,}7})$ es 0,9965 ó 99,65%.

z	0,00	0,01	0,02	0,03			
2,7	0,9965						

$\boldsymbol{Caso}\ 5$: $\boldsymbol{P(a \leq z \leq b) = P(z \leq b) - P(z \leq a)}$



<u>Ejemplo</u>: Determinar $\boldsymbol{P(0,28 \leq z \leq 1,85)}$, es decir, determinar la probabilidad que z sea mayor o igual que 0,28 y menor o igual que 1,85.

El área bajo la curva buscada se puede determinar como la diferencia entre $\boldsymbol{P}(z \leq 1,85)\ \ y\ P(z \leq 0,28)$

$P(0,28 \leq z \leq 1,85)$ = $P(z \leq 1,85) - P(z \leq 0,28)$=0,9678-0,6103=0,3575=35,75%

La probabilidad que $P(z \leq 1,85)$ se obtuvo en la intersección del valor 1,8 en la primera columna y 0,05 en la primera fila, ya que 1,8+0,05=1,85. La probabilidad de $P(z \leq 1,85)$ es 0,9678 ó 96,78%.

La probabilidad que $P(z \leq 0,28)$ se obtuvo en la intersección del valor 0,2 en la primera columna y 0,08 en la primera fila, ya que 0,2+0,08=0,28. La probabilidad de $P(z \leq 0,28)$ es 0,6103 ó 61,03%.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,08
0,2							0,6103
1,8						0,9678	

<u>Ejemplo</u>: Determinar $\boldsymbol{P(0,22 < z < 1,35)}$, es decir, determinar la probabilidad que z sea mayor que 0,22 y menor que 1,35. (libro IV medio, p. 291).

El área bajo la curva buscada se puede determinar como la diferencia entre $P(< 1,35)\ \ y\ P(z < 0,22)$

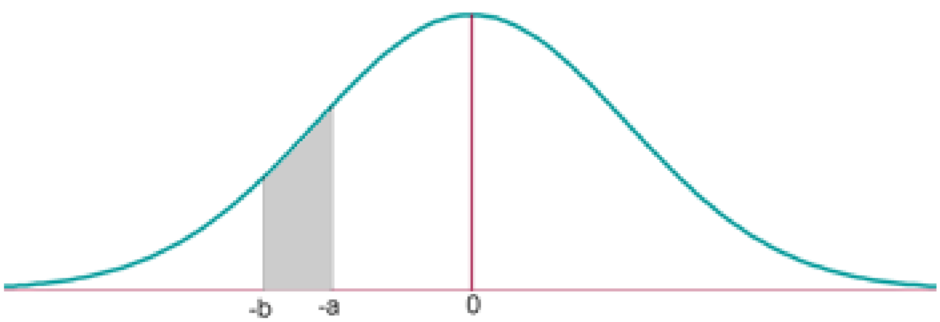
$P(0,22 < z < 1,35)$ = $P(z < 1,35) - P(z < 0,22) = 0,91149 - 0,58706 = 0,32443 =$ 32,44%

La probabilidad que $\boldsymbol{P(z < 1,35)}$ se obtuvo en la intersección del valor 1,3 en la primera columna y 0,05 en la primera fila, ya que 1,3+0,05=1,35. La probabilidad de $P(z \leq 1,35)$ es 0,91149 o 91,15%.

La probabilidad que $P(z < 0,22)$ se obtuvo en la intersección del valor 0,2 en la primera columna y 0,02 en la primera fila, ya que, 0,2+0,02=0,22. La probabilidad de $P(z < 0,22)$ es 0,58706 o 58,71%.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,08
0,2			0,58706				
1,3						0,91149	

$\boldsymbol{Caso\ 6}$: $\boldsymbol{P(-b < z \leq -a) = P(a < z \leq b)}$

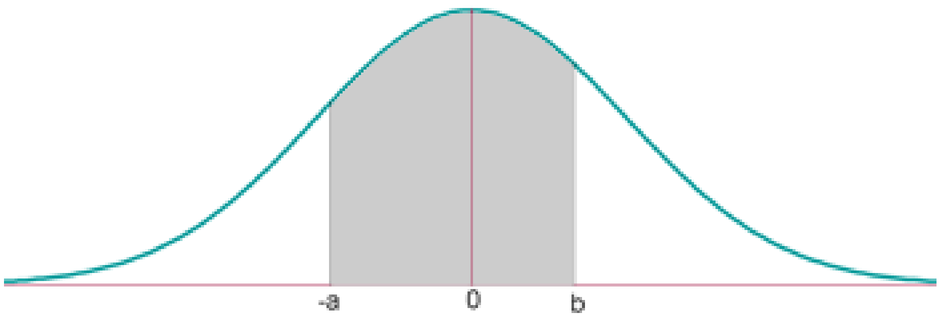


Ejemplo: Determinar $\boldsymbol{P(-1,39 < z \leq -0,44)}$, es decir, determinar la probabilidad que z sea mayor que -1,39 y menor o igual que -0,44.

Por la simetría de la curva de Gauss,

$P(-1,39 < z \leq -0,44)=$ $P(0,44 < z \leq 1,39)= P(z \leq 1,39) - P(z < 0,44) = 0,9177 - 0,67 = 0,2477$

$\boldsymbol{caso\ 7}$: $\boldsymbol{P(-a < z \leq b) = P(z \leq b) - [1 - P(z \leq a)]}$



<u>Ejemplo</u>: Determinar $\boldsymbol{P(-0,83 \leq z \leq 1,32)}$, es decir, determinar la probabilidad que z sea mayor o igual que -0,83 y menor o igual que 1,32.

El área bajo la curva buscada se puede determinar como la diferencia entre $P(z \leq 1{,}32)$ $y\ P(z \leq -0{,}83)$

$P(-0{,}83 \leq z \leq 1{,}32)$ = $P(z \leq 1{,}32) - P(z \leq -0{,}83)$

Pero, por el $\boldsymbol{Caso\ 3}$: $\boldsymbol{P(z \leq -a) = 1 - P(z \leq a)}$

Entonces,

$P(-0{,}83 \leq z \leq 1{,}32)$ = $P(z \leq 1{,}32) - [1 - P(z \leq 0{,}83)$

$P(-0{,}83 \leq z \leq 1{,}32)$ = 0,9066-1+0,7967= 0,7033

La probabilidad que $P(z \leq 1{,}32)$ se obtuvo en la intersección del valor 1,3 en la primera columna y 0,02 en la primera fila, ya que 1,3+0,02=1,32. La probabilidad de $P(z \leq 1{,}85)$ es 0,9066 ó 90,66%.

La probabilidad que $P(z \leq 0{,}83)$ se obtuvo en la intersección del valor 0,8 en la primera columna y 0,03 en la primera fila, ya que 0,8+0,03=0,83. La probabilidad de $P(z \leq 0{,}83)$ es 0,7967 ó 79,67%.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,08
0,8				0,7967			
1,3			0,9066				

Algunos ejemplos utilizando la tabla de distribución normal tipificada II:

Ejemplo 1: Haciendo uso de la tabla II, determinar $\boldsymbol{P(z \leq 0,22)}$, es decir, determinar la probabilidad que z sea menor o igual que 0,22.

En este caso, dado que la tabla II proporciona el área comprendida entre 0 y z, lo primero que hay que establecer es una probabilidad del tipo $P(0 < z < a)$, es decir, $P(0 < z \leq 0{,}22)$. En la tabla II de distribución normal tipificada, en la primera columna se ubica el valor 0,2 y en la primera fila el valor 0,02, ya que $0{,}2 + 0{,}02 = 0{,}22$. La probabilidad entre 0 y 0,22 es el número que aparece en la intersección y es 0,08706.

z	0,01	0,02	0,03	0,04	0,05	0,06	
0,2		0,0871					

El valor recién calculado se adiciona al valor bajo la curva a la izquierda de su valor central que, por definición de la curva normal, es 0,5. Por lo tanto:

$P(z \leq 0{,}22) = 0{,}5 + 0{,}0871 = 0{,}5871$ *o* 58,71%

Ejemplo 2: Haciendo uso de la tabla II, determinar $\boldsymbol{P(z < -1,80)}$, es decir, determinar la probabilidad que z sea menor que -1,80.

Como $P(z < -1{,}8)$ = $P(z > 1{,}8)$, en este caso, la tabla II proporciona el área comprendida entre 0 y Z, lo primero que hay que establecer es una probabilidad del tipo $P(0 < z < a)$, es decir, $P(0 < z < 1{,}8)$. En la tabla II de distribución normal tipificada, en la primera columna se ubica el valor 0,2 y en la primera fila el valor 0,00, ya que 1,8+0,00=1,80. La probabilidad entre 0 y 1,80 es el número que aparece en la intersección y es 0,4641.

z	0,00	0,01	0,02	0,03	0,04	0,05	
1,8	0,4641						

Como el valor bajo la curva a la derecha del punto central es 0,5, a dicho valor hay que restarle el valor recién determinado entre 0 y 1,80=0,4641.

Por lo tanto:

$P(z < 1{,}80) = \ 0{,}5 - 0{,}4641 = 0{,}0359$ *o* 3,59%

Ejemplo 3: Haciendo uso de la tabla II, determinar $\boldsymbol{P(z > 1,0092)}$, es decir, determinar la probabilidad que z sea mayor que 1,0092.

Se aproxima 1,0092 a 1,01, ya que la tabla II entrega valores con solo dos decimales. Como la tabla II proporciona el área comprendida entre 0 y z, lo primero que hay que establecer es una probabilidad del tipo $P(0 < Z < a)$, es decir, $P(0 < z < 1{,}01)$. En la tabla II de distribución normal tipificada, en la primera columna se ubica el valor 1,0 y en la primera fila el valor 0,01, ya que 1,0+0,01=1,01. La probabilidad entre 0 y 1,01 es el número que aparece en la intersección y es 0,3438.

z	0,00	0,01	0,02	0,03	0,04	0,05	
1,0		0,3438					

Como el valor bajo la curva a la derecha del punto central es 0,5, a dicho valor hay que restarle el valor recién determinado entre 0 y 1,01=0,3438.

Por lo tanto:

$\boldsymbol{P(z > 1{,}0092)}$= 0,5-0,3438=0,1562 ó 15,62%

<u>Ejemplo 4</u>: Haciendo uso de la tabla II, determinar $P(z > -1{,}61)$, es decir, determinar la probabilidad que z sea mayor que -1,61.

Según el Caso 4: $P(z > -a) = P(z \leq a)$,

Entonces, $P(z > -1{,}61) = P(z < 1{,}61)$

Dado que la tabla II proporciona el área comprendida entre 0 y z, lo primero que hay que establecer es una probabilidad del tipo $\boldsymbol{P(0 < Z < a)}$, es decir, $\boldsymbol{P(0 < z < 1{,}61)}$. En la tabla II de distribución normal tipificada, en la primera columna se ubica el valor 1,0 y en la primera fila el valor 0,01, ya que 1,6+0,01=1,61. La probabilidad entre 0 y 1,61 es el número que aparece en la intersección y es 0,4463.

z	0,00	0,01	0,02	0,03	0,04	0,05	
1,6		0,4463					

Considerando que el valor bajo la curva a la izquierda del punto central es 0,5, a dicho valor hay que sumarle el valor recién determinado entre 0 y 1,61=0,4463. Por tanto,

$P(z > -1{,}61) = 0{,}5 + 0{,}4463 = 0{,}9463$ ó 94,63%

Ejemplo 5: Haciendo uso de la tabla II, determinar $\boldsymbol{P(-2{,}06 < z \leq -0{,}24)}$ es decir, determinar la probabilidad que z sea mayor que -2,06 y menor o igual que -0,24.

Según el **Caso 6:** $\boldsymbol{P(-b < z \leq -a) = P(a < z \leq b)}$

Por la simetría de la curva de Gauss,

$P(-2{,}06 < z < -0{,}24) =$ $P(2{,}06 < z < 0{,}26) = P(z < 2{,}06) - P(z < 0{,}24) = 0{,}4803 - 0{,}0948 = 0{,}3855 = 38{,}55\%$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
0,2					0,0948		
2,0							0,4803

Ejemplo 6: Haciendo uso de la tabla II, determinar $\boldsymbol{P(-0{,}02 \leq Z \leq 1{,}70)}$ es decir, determinar la probabilidad que Z sea mayor 0 igual que -0,02 y menor o igual que 1,17.

Se calcula la probabilidad entre -0,02 y 0, teniendo presente que, por la simetría de la curva de Gauss, $P(Z \geq -0{,}02) = P(Z \leq 0{,}02)$. Luego se determina $P(Z \leq 1{,}70)$.

Por lo tanto, para encontrar la $P(-0{,}02 \leq Z \leq 1{,}70)$, hay que sumar
$P(Z \leq 1{,}70) + P(Z \leq 0{,}02) = 0{,}4554 + 0{,}0080 = 0{,}4634 = 46{,}34\%$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
0,0			0,0080				
1,7	0,4554						

Ejemplos de cálculo de probabilidades bajo la curva de Gauss que requieren determinar previamente el valor del puntaje z:

Ejemplo 1: ¿Cuál es la probabilidad de obtener entre 670 y 770 puntos en la PSU?

Se recuerda que la distribución normal de los puntajes de la PSU tenía una media aritmética de 500 y una desviación estándar de 110.

Para $x = 670$, $z = \frac{x_i - \bar{x}}{s} = \frac{670-500}{110} = \frac{170}{110} = 1{,}55$

Para $x = 770$, $z = \frac{x_i - \bar{x}}{s} = \frac{770-500}{110} = \frac{270}{110} = 2{,}45$

En la tabla I de la distribución normal tipificada, para:

$z = 1{,}55$, el área es: 0,9394

$z = 2{,}45$, el área es: 0,9929

$P(670 \leq x \leq 770) = P(1{,}55 \leq z \leq 2{,}45)$

$P(670 \leq x \leq 770 = P(z \leq 2{,}45) - P(z \leq 1{,}55)$

$P(670 \leq x \leq 770 = 0{,}9929 - 0{,}9394 = 0{,}0535 = 5{,}35\%$

<u>Ejemplo 2</u>: ¿Cuál es la probabilidad de obtener entre 450 y 550 puntos en la PSU?

Se recuerda que la distribución normal de los puntajes de la PSU tiene una media aritmética de 500 y una desviación estándar de 110.

Para $x = 450$, $z = \frac{x_i - \bar{x}}{s} = \frac{450-500}{110} = \frac{-50}{110} = -0{,}45$

Para $x = 550$, $z = \frac{x_i - \bar{x}}{s} = \frac{550-500}{110} = \frac{50}{110} = +0{,}45$

$P(450 \leq x \leq 550) = P(-0{,}45 \leq z \leq 0{,}45)$

$P(450 \leq x \leq 550 = P(z \leq 0{,}45) + P(z \geq -0{,}45)$

Como la curva normal es simétrica, el área bajo la curva a la derecha de -0,45 es igual al área a la izquierda de 0,45, entonces:

$P(450 \leq x \leq 550) = P(z \leq 0{,}45) + P(z \leq 0{,}45)$

En la tabla de la distribución normal tipificada, para:

z=0,45, el área es: 0,6736

$P(450 \leq x \leq 550) = 0{,}6736 + 0{,}6736 = 1{,}3347$

<u>Ejemplo 3</u>: Las calificaciones de los 500 aspirantes que se presentaron a un examen para un puesto de trabajo se distribuyen normalmente con una media aritmética de 6,5 puntos y una varianza 4.

a) Calcular la probabilidad de que un aspirante obtenga más de 8 puntos.
b) Determinar la proporción de aspirantes con calificaciones inferiores a 5 puntos.
c) ¿Cuántos aspirantes obtuvieron calificaciones comprendidas entre 5 y 7,5 puntos?

Solución:

a) Se trata de una distribución normal $N(6{,}5, \sqrt{4}) = N(6{,}5, 2)$. Se recuerda que la desviación estándar es la raíz cuadrada de la varianza.

Tipificando la calificación de 8 puntos: $z = \frac{x_i - \bar{x}}{s} = \frac{8-6{,}5}{2} = \frac{1{,}5}{2} = 0{,}75$

Utilizando la tabla de distribución normal tipificada I, se tiene que la probabilidad solicitada, que es la que está a la derecha de z, se calcula por la diferencia entre $\mathbf{1} - P(z \leq 0{,}75) = 1 - 0{,}7734 = 0{,}2266 = 22{,}66\%$.

b) Tipificando la calificación de 5 puntos: $\boldsymbol{z = \frac{xi - \bar{x}}{s} = \frac{5-6{,}5}{2} = \frac{-1{,}5}{2} = -0{,}75}$

Teniendo presente que por simetría de la curva normal, $\boldsymbol{P(z \leq -0{,}75) = P(z \leq 0{,}75)}$ y utilizando la tabla de distribución normal tipificada I, se tiene que la probabilidad solicitada se calcula por la diferencia entre $1 - P(z \leq 0{,}75) = 1 - 0{,}7734 = 0{,}2266 = 22{,}66\%$.

c) Tipificando la calificación de 5 puntos: $z = \frac{x_i - \bar{x}}{s} = \frac{5-6{,}5}{2} = \frac{-1{,}5}{2} = -0{,}75$

Tipificando la calificación de 7,5 puntos: $z = \frac{x_i - \bar{x}}{s} = \frac{7{,}5-6{,}5}{2} = \frac{1}{2} = 0{,}5$

Teniendo presente que por simetría $P(z > -0{,}75) = P(z < 0{,}75)$

$P(-0{,}75 < z < 0{,}5) = P(z > -0{,}75) + P(z < 0{,}5) = 0{,}2734 + 0{,}1915 = 0{,}4649$

Multiplicando esta probabilidad por el total de aspirantes al puesto que son 500, se determina que los aspirantes que tienen entre 5 y 7,5 puntos son: $0{,}4648 \cdot 500 = 232$.

Ejemplo 4: La media aritmética de los pesos de 500 estudiantes es de 151 libras y la desviación típica de 15. Suponiendo que los pesos se distribuyen normalmente, ¿Cuántos estudiantes pesan: a) Entre 120 y 155 libras y b) más de 185 libras? Hacer uso de la tabla II.

Respuesta a):

$$z_1 = \frac{x_i - \bar{x}}{s} = \frac{120 - 151}{15} = -2{,}06$$

$$z_2 = \frac{x_i - \bar{x}}{s} = \frac{155 - 151}{15} = 0{,}26$$

El área bajo la curva normal, entre $z = 0$ y $z = -2{,}06$ es: 0,4803 y el área entre $z = 0$ y $z = 0{,}26$ es 0,1026. Por lo tanto, el porcentaje de estudiantes entre 120 y 155 libras es: $0{,}4803 + 0{,}1026 =$

$0{,}5829 = 58{,}29\%$. Como los estudiantes son 500, los que están entre los pesos señalados son: $0{,}5829 \cdot 500 = 291$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
0,2							0,1026
2,0	0,4554						0,4803

Respuesta b):

$$z = \frac{x_i - \bar{x}}{s} = \frac{185 - 151}{15} = \frac{34}{15} = 2{,}26$$

El área bajo la curva normal, entre Z=0 y Z=2,26 es: 0,4881. Por lo tanto, el área hacia la derecha de 185 libras se obtiene restando ese valor del 50%, es decir, 0,50-0,4881=0,0119 o 1,19%. De esta manera, los estudiantes que pesan más de 185 libras es: 500*0,119=5,95=6

<u>Ejemplo 5</u>: Las edades de los alumnos de un curso siguen aproximadamente una distribución normal, con media 25 años y desviación típica 5. Calcular:

a) La proporción de alumnos con más de 29 años.
b) La proporción de alumnos que tienen edades entre 22 y 26 años.

<u>Solución a)</u>:

Se trata de una distribución normal $N(25, 5)$.

Tipificando la edad de 29 años:

Para $x = 29$, $z = \frac{x_i - \bar{x}}{s} = \frac{29-25}{5} = \frac{4}{5} = 0{,}8$

Utilizando la tabla de distribución normal tipificada I, que proporciona el área que queda a la izquierda de x=29 años, se tiene que para z=0,8 el área es: 0,7881. Por lo tanto, como el área total bajo la curva es 1, el área que está a la derecha de 29 es: $1 - P(z \leq 0{,}8) = 1 - 0{,}7781 = 0{,}2119 = 21{,}2\%$.

Alternativamente, se puede llegar al mismo resultado utilizando la tabla de distribución normal tipificada II, que proporciona el área entre 0 y un valor determinado de z, en este caso, entre 0 y 0,8. Tal área es: 0,2881. Como con esta tabla se está utilizando la mitad del área bajo la curva, la proporción de alumnos con más de 29 años es: $0{,}5 - P(z \leq 0{,}8) = 0{,}5 - 0{,}2881 = 0{,}2119 = 21{,}2\%$.

Solución b):

$$z_1 = \frac{x_i - \bar{x}}{s} = \frac{26 - 25}{5} = \frac{1}{5} = 0{,}20$$

$$z_2 = \frac{x_i - \bar{x}}{s} = \frac{22 - 25}{15} = -\frac{3}{5} = -0{,}6$$

Usando la tabla de distribución tipificada II:

$P(-0{,}60 \leq z \leq 0{,}20) = P(0 \leq z \leq 0{,}2) + P(-0{,}6 \leq z \leq 0)$

$P(-0{,}60 \leq z \leq 0{,}20) = P(0 \leq z \leq 0{,}2) + P(0{,}6 \geq z \geq 0)$

$P(-0{,}60 \leq z \leq 0{,}20) = 0{,}0793 + 0{,}2257 = 0{,}3050$ o 30,5%

Usando la tabla de distribución tipificada I:

$P(-0{,}60 \leq z \leq 0{,}20) = P(z \leq 0{,}2) - P(z \leq -0{,}6)$

$P(-0{,}60 \leq z \leq 0{,}20) = P(z \leq 0{,}2) - P(z \geq 0{,}6)$

$P(-0{,}60 \leq z \leq 0{,}20) = P(z \leq 0{,}2) - [1 - P(z \leq 0{,}6)]$

$P(-0{,}60 \leq z \leq 0{,}20) = 0{,}5793 - [1 - 0{,}7257]$

$P(-0{,}60 \leq z \leq 0{,}20) = 0{,}5793 - 1 + 0{,}7257 = 0{,}3050$ o 30,5%

6.3 Ejercicios sobre la curva normal

Ejercicio: ¿Cuál es la probabilidad de un coeficiente intelectual (CI) entre 95 y 107, sabiendo que para esta distribución la media aritmética es 100 y la desviación típica es 10?

Solución:

a) Hay que calcular previamente los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:

$$z_{95} = \frac{x_i - \bar{x}}{\sigma} = \frac{95 - 100}{10} = -0{,}50$$

$$z_{107} = \frac{x_i - \bar{x}}{\sigma} = \frac{107 - 100}{10} = 0{,}70$$

b) Plantear la probabilidad:

$P(-0{,}50 < z < 0{,}70)$

Teniendo en mente el uso de la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda a su izquierda:

$P(-0{,}50 < z < 0{,}70) = P(z < 0{,}70) - P(z < -0{,}50)$

Para evitar el valor negativo, usar la característica de simetría de la curva normal. Por lo que $P(z < -0{,}50) = P(z > 0{,}50)$

$P(-0{,}50 < z < 0{,}70) = P(z < 0{,}70) - P(z > 0{,}50)$

Pero hay que trabajar con valores "<", es decir, "menores que" para poder usar la tabla mencionada. Por ello, hay que usar $P(z > 0{,}50) = 1 - P(z < 0{,}50)$.

$P(-0{,}50 < z < 0{,}70) = P(z < 0{,}70) - [1 - P(z < 0{,}50)]$

Usando los valores de esa tabla:

$P(-0{,}50 < z < 0{,}70) = 0{,}7580 - [1 - 0{,}6915]$

$P(-0{,}50 < z < 0{,}70) = 0{,}7580 - [0{,}3085] = 0{,}4495 = 44{,}95\%$

Ejercicio: ¿Cuál es la probabilidad de un coeficiente intelectual (CI) entre 112 y 122, sabiendo que para esta distribución la media aritmética es 100 y la desviación típica es 10?

Solución:

c) Hay que calcular previamente los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:
d)

$$z_{112} = \frac{x_i - \bar{x}}{\sigma} = \frac{112 - 100}{10} = 1{,}2$$

$$z_{120} = \frac{x_i - \bar{x}}{\sigma} = \frac{120 - 100}{10} = 2{,}2$$

e) Plantear la probabilidad:

$$P(1{,}2 < z < 2{,}2)$$

Teniendo en mente el uso de la tabla de distribución normal tipificada $N(0{,}1)$ que proporciona para cada valor de x el área que queda a su izquierda:

$P(1{,}2 < z < 2{,}2) = P(z < 2{,}2) - P(z < 1{,}2)$

Como ya las probabilidades que incluyen valores "< " (y positivos) propios de esa tabla, solo resta ir a ella y encontrar los valores correspondientes a los puntajes z:

$P(1{,}2 < z < 2{,}2) = 0{,}9861 - 0{,}8849 = 0{,}1012 = 10{,}12\%$

Ejercicio: ¿Cuál es la probabilidad de obtener entre 670 y 770 puntos en la PAA (anterior a la PSU), sabiendo que para esta distribución la media aritmética era 500 y la desviación típica era 100?

Solución:

a) Hay que calcular previamente los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:

$$z_{670} = \frac{x_i - \bar{x}}{\sigma} = \frac{670 - 500}{100} = 1{,}70$$

$$z_{770} = \frac{x_i - \bar{x}}{\sigma} = \frac{770 - 500}{100} = 2{,}70$$

b) Plantear la probabilidad:

$P(1{,}7 < z < 2{,}7)$

Teniendo en mente el uso de la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda <u>a su izquierda:</u>

$P(1{,}7 < z < 2{,}7) = P(z < 2{,}7) - P(z < 1{,}7)$

Como las probabilidades ya incluyen valores "< " propios de esa tabla (y positivos), solo resta ir a ella y encontrar los valores correspondientes a los puntajes z:

$P(1{,}7 < z < 2{,}7) = 0{,}9965 - 0{,}9554 = 0{,}0411 = 4{,}11\%$

5) ¿Cuál era la probabilidad de obtener entre 450 y 550 puntos en la PAA (anterior a la PSU), sabiendo que para esta distribución la media aritmética es 500 y la desviación típica es 100?

Solución:

f) Hay que calcular previamente los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:

$$z_{450} = \frac{x_i - \bar{x}}{\sigma} = \frac{450 - 500}{100} = -0{,}50$$

$$z_{550} = \frac{x_i - \bar{x}}{\sigma} = \frac{550 - 500}{100} = 0{,}50$$

g) Plantear la probabilidad:

$$P(-0{,}50 < z < 0{,}50)$$

Teniendo en mente el uso de la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda a su izquierda:

$$P(-0{,}50 < z < 0{,}50) = P(z < 0{,}50) - P(z < -0{,}50)$$

Para evitar el valor negativo, usar la característica de simetría de la curva normal. Por lo que $P(z < -0{,}50) = P(z > 0{,}50)$

$$P(-0{,}50 < z < 0{,}50) = P(z < 0{,}50) - P(z > 0{,}50)$$

Pero hay que trabajar con valores "<", es decir, "menores que" para poder usar la tabla mencionada. Por ello, hay que usar $P(z > 0{,}50) = 1 - P(z < 0{,}50)$

$$P(-0{,}50 < z < 0{,}50) = P(\mathrm{z} < 0{,}50) - [1 - P(z < 0{,}50)]$$

Usando los valores de esa tabla:

$$P(-0{,}50 < z < 0{,}50) = 0{,}6915 - [1 - 0{,}6915]$$

$$P(-0{,}50 < z < 0{,}50) = 0{,}6915 - [0{,}3085] = 0{,}383 = 38{,}3\%$$

Ejercicio: Las temperaturas mínimas registradas durante un mes en Punta Arenas distribuyen como una normal con media de -5ºC y desviación estándar de 3ºC. Se sabe que la probabilidad de que la temperatura sea menor que -2ºC es 0,16 y que la probabilidad de que la temperatura sea mayor que -7ºC es de 0,52. ¿Cuál es la probabilidad de que la temperatura tome valores entre los -7ºC y los -2ºC? (Ejercicio de puntajenacional.cl).

A) 0,32
B) 0,36
C) 0,48
D) 0,68
E) 0,81

Solución:

a) Como en este ejercicio se dan las probabilidades, no es necesario calcular previamente los puntajes sigma (puntajes z), por lo que los datos que se dan de la media y la desviación estándar no son necesarios.

b) Plantear la probabilidad:

$P(-7 < T < -2)$

$P(-7 < T < -2) = P(T < -2) - P(z < -7)$

$P(-7 < T < -2) = P(T > 2) - P(z > 7)$

$P(-7 < T < -2) = 1 - P(T < 2) - [1 - P(T < 7)]$

$P(-7 < T < -2) = 1 - 0{,}16 - [1 - 0{,}52] = 0{,}84 - 0{,}48 = 0{,}36$

Ejercicio: La estatura de los integrantes de una delegación de 180 deportistas se describe aproximadamente con una distribución normal $N(184{,}9)$. ¿Cuál es aproximadamente el porcentaje de deportistas cuya estatura es mayor que 175 cm?

Solución:

a) Hay que calcular previamente el respectivo puntaje sigma (puntaje z) para la puntuación 175:
$$z_{175} = \frac{x_i - \bar{x}}{\sigma} = \frac{175 - 184}{9} = -\frac{9}{9} = -1$$

b) Plantear la probabilidad:

$P(z > -1)$

Para evitar el valor negativo, usar la característica de simetría de la curva normal. Por lo que $P(z > -1) = P(z < 1)$.

Usando la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda <u>a su izquierda:</u>

$P(z < 1) = 0{,}8413$

Solución alternativa:

Si se calcula el intervalo $(Ma - 1\sigma, Ma + 1\sigma) = (184 - 9, 184 + 9) = (175, 193)$ se llega a las puntuaciones del ejercicio, y se sabe que en una distribución normal dicho intervalo agrupa el 68,3% de las puntuaciones.

Como el área bajo la curva es igual a 1, hay un 15,85% menor que 175 cm y un 15,85% mayor a 193 cm. Como interesa la puntuación mayor que175 cm, hay que sumar a los 68,3% los 15,85% que están a la derecha de la puntuación 193 cm, es decir, 68,3% + 15,85%= 84,15%, valor similar al determinado en la primera solución.

6) La estatura de los integrantes de una delegación de 180 deportistas se describe aproximadamente con una distribución normal N(184,9). ¿Cuántos deportistas aproximadamente son de estatura mayor que 157 cm y menor que 211 cm?

Solución:

a) Hay que calcular previamente los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:

$$z_{157} = \frac{x_i - \bar{x}}{\sigma} = \frac{157 - 184}{9} = -\frac{27}{9} = -3$$

$$z_{211} = \frac{x_i - \bar{x}}{\sigma} = \frac{211 - 184}{9} = \frac{27}{9} = 3$$

b) Plantear la probabilidad:

$$P(-3 < z < 3)$$

Teniendo en mente el uso de la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda a su izquierda:

$$P(-3 < z < 3) = P(z < 3) - P(z < -3)$$

Para evitar el valor negativo, usar la característica de simetría de la curva normal. Por lo que $P(z < -3) = P(z > 3)$

$$P(-3 < z < 3) = P(z < 3) - P(z > 3)$$

Pero hay que trabajar con valores “<”, es decir, “menores que” para poder usar la tabla mencionada. Por ello, hay que usar $P(z > 3) = 1 - P(z < 3)$

$$P(-3 < z < 3) = P(z < 3) - [1 - P(z < 3)]$$

Usando los valores de esa tabla:

$$P(-3 < z < 0{,}3) = 0{,}9987 - [1 - 0{,}9987]$$

$$P(-3 < z < 3) = 0{,}9987 - [0{,}0013] = 0{,}9974 = 99{,}7\%$$

Solución alternativa:

Si se calcula el intervalo:

$(Ma - 3\sigma, Ma + 3\sigma) = (184 - 3 * 9, 184 + 3 * 9) = (184 - 27, 184 + 27) = (157, 211)$ se llega a las puntuaciones del ejercicio, y se sabe que en una distribución normal dicho intervalo agrupa el 99,7% de las puntuaciones. Los deportistas son, entonces, 0,997*180=179 aproximadamente.

Ejercicio: Se mide la masa en gramos de los huevos producidos en un gallinero. Si dichas masas se distribuyen de forma normal con una media aritmética igual a 54 y una desviación típica de 16., ¿cuál es la probabilidad de que al elegir aleatoriamente menor que 52 gramos?

Solución:

a) Hay que calcular previamente el respectivo puntaje sigma (puntaje z) para la puntuación 52:

$$z_{52} = \frac{x_i - \bar{x}}{\sigma} = \frac{52 - 54}{16} = -\frac{2}{16} = -0{,}125$$

b) Plantear la probabilidad:

$$P(x < 52) = P(z < -0{,}125)$$

Recordando que $P(z < -a) = P(z > a)$ por simetría de la curva de Gauss

$$P(z < -0{,}125) = P(z > 0{,}125)$$

Pero como se usará la tabla de distribución normal tipificada N(0,1) que proporciona para cada valor de x el área que queda <u>a su izquierda,</u> se debe operar con un valor "<". Recordando que: $P(z > a) = 1 - P(z < a)$

$$P(z < -0{,}125) = P(z > 0{,}125) = 1 - P(z < 0{,}125)$$

Como el valor 0,125 no se encuentra directamente en la tabla, hay que promediar el que se da para las columnas 0,02 y 0,03, es decir:

$$\frac{0{,}5470 + 0{,}5517}{2} = 0{,}54975$$

$$P(z < -0{,}125) = P(z > 0{,}125) = 1 - 0{,}54975 = 0{,}45025 = 45\%$$

Ejercicio: Los sueldos de 50 personas se distribuyen $N(315, 85)$ en miles de pesos. a) ¿ Cuál es la probabilidad de que al elegir una persona reciba más de $390.000?, b) ¿Qué porcentaje de personas reciben menos de $260.000?, c) ¿Cuántas personas, aproximadamente, reciben entre 230 y 400 miles de pesos?

Solución a):

- Hay que calcular previamente el respectivo puntaje sigma (puntaje z) para la puntuación 390.000:
  $$z_{390.000} = \frac{x_i - \bar{x}}{\sigma} = \frac{390 - 315}{85} = \frac{75}{85} = 0{,}88$$

- Plantear la probabilidad:

  $$P(x > 390.000) = P(z > 0{,}88) = 1 - P(z < 0{,}88)$$

  $$P(z > 0{,}88) = 1 - 0{,}8106 = 0{,}1894 = 18{,}94\%$$

Solución b):

- Hay que calcular previamente el respectivo puntaje sigma (puntaje z) para la puntuación 260.000:
  $$z_{260.000} = \frac{x_i - \bar{x}}{\sigma} = \frac{260 - 315}{85} = -\frac{55}{85} = -0{,}65$$

- Plantear la probabilidad:

  $$P(x < 260.000) = P(z < -0{,}65) = P(z > 0{,}65)$$
  $$P(z > 0{,}65) = 1 - P(z < 0{,}65) = 1 - 0{,}7422 = 0{,}2578 = 25{,}78\%$$

Solución c):

- Hay que calcular previamente los respectivos puntaje sigma (puntaje z) para las puntuaciones 230 y 400 miles de pesos.

  $$z_{230.000} = \frac{x_i - \bar{x}}{\sigma} = \frac{230 - 315}{85} = -\frac{85}{85} = -1$$

  $$\mathrm{z}_{400.000} = \frac{\mathrm{x_i} - \bar{\mathrm{x}}}{\sigma} = \frac{400 - 315}{85} = \frac{85}{85} = 1$$

- Plantear la probabilidad:

  $$\mathrm{P}(230.000 < \mathrm{x} < 400.000) = \mathrm{P}(-1 < \mathrm{z} < 1)$$

$$P((-1 < z < 1) = P(z < 1) - P(z < -1)$$

$$P((-1 < z < 1) = P(z < 1) - P(z > 1)$$

$$P((-1 < z < 1) = P(z < 1) - [1 - P(z < 1)] = 0{,}8413 - (1 - 0{,}8413)$$
$$= 0{,}8413 - 0{,}1587 = 0{,}6826$$

Por lo tanto, la cantidad de personas que reciben entre 230 y 400 miles de pesos son: $0{,}6826 \cdot 50 \approx 34$.

Ejercicio: Un alumno obtuvo 245 puntos en una prueba de aptitud verbal donde la media aritmética fue de 220 puntos y la desviación típica de 50 puntos. En la prueba de aptitud matemática logró 175 puntos donde la media fue de 150 y la desviación típica de 25 puntos. ¿En cuál de las dos pruebas obtuvo un mejor rendimiento?

Solución:

Hay que calcular los respectivos puntajes sigma (puntajes z) para cada de las puntuaciones:

$$z_{av} = \frac{x_i - \bar{x}}{\sigma} = \frac{245 - 220}{50} = 0{,}50$$

$$z_{am} = \frac{x_i - \bar{x}}{\sigma} = \frac{175 - 150}{25} = 1$$

dado que $z_{am} > z_{av}$, el mejor rendimiento lo obtuvo en la prueba de matemática. El rango percentil de este alumno según la tabla de distribución normal tipificada $N(0,1)$, que proporciona, para cada valor de x, el área que queda a su izquierda es de solo 69,15% para la prueba verbal y 84,13% para la prueba de matemática.

<u>Relaciones muy importantes</u>: Si $a \in \mathbb{R}$

i) $P(z \leq -a) = P(z \geq a)$

ii) $P(z \geq a) = 1 - P(z \leq a)$

iii) $P(-a \leq z \leq a) = p(z \leq a) - P(z \leq -a)$

7 REFERENCIAS BIBLIOGRÁFICAS

Grassau, E. (1964). Elementos de estadística. Santiago: Editorial Universitaria S.A.

Molina, C. (1998). Introducción a la metodología de la investigación. Serie de ciencias de la educación 12. Santiago: CPEIP.

Núñez del Prado, A. (1985). Estadística básica para planificación. México: Siglo XXI Editores.

Ovalle de Gómez, R. (1978). Estadística a su alcance. Serie de conocimientos prácticos. Bogotá: Editorial Norma.

Tapia, O., Ormazábal, M., Olivares, J. López, D. (2006). Manual de preparación matemática. Santiago: Ediciones UC.

Noviembre 2024