

Universidad Nacional de Jujuy



Facultad de Ciencias Económicas

Guía práctica para Resolver problemas de aplicación

ESTADÍSTICA DESCRIPTIVA

Temas:

Organización y presentación de datos univariantes

- Tabulación, análisis, interpretación y representación gráfica de datos estadísticos.

Series simples y series de frecuencias

- Para variables discretas y continuas.
- Distribuciones de frecuencias absolutas, relativas y acumulativas. Gráficos.

Medidas de tendencia central

- Media aritmética simple y ponderada. Mediana. Moda o modo. Media geométrica y media armónica.

Medidas de dispersión

- Rango o recorrido. Desviación media y desviación intercuartilar. Variancia y desviación estándar. Coeficiente de variación

Medidas de asimetría y kurtosis

Profesora: Lydia María Llanos

Estadística Descriptiva

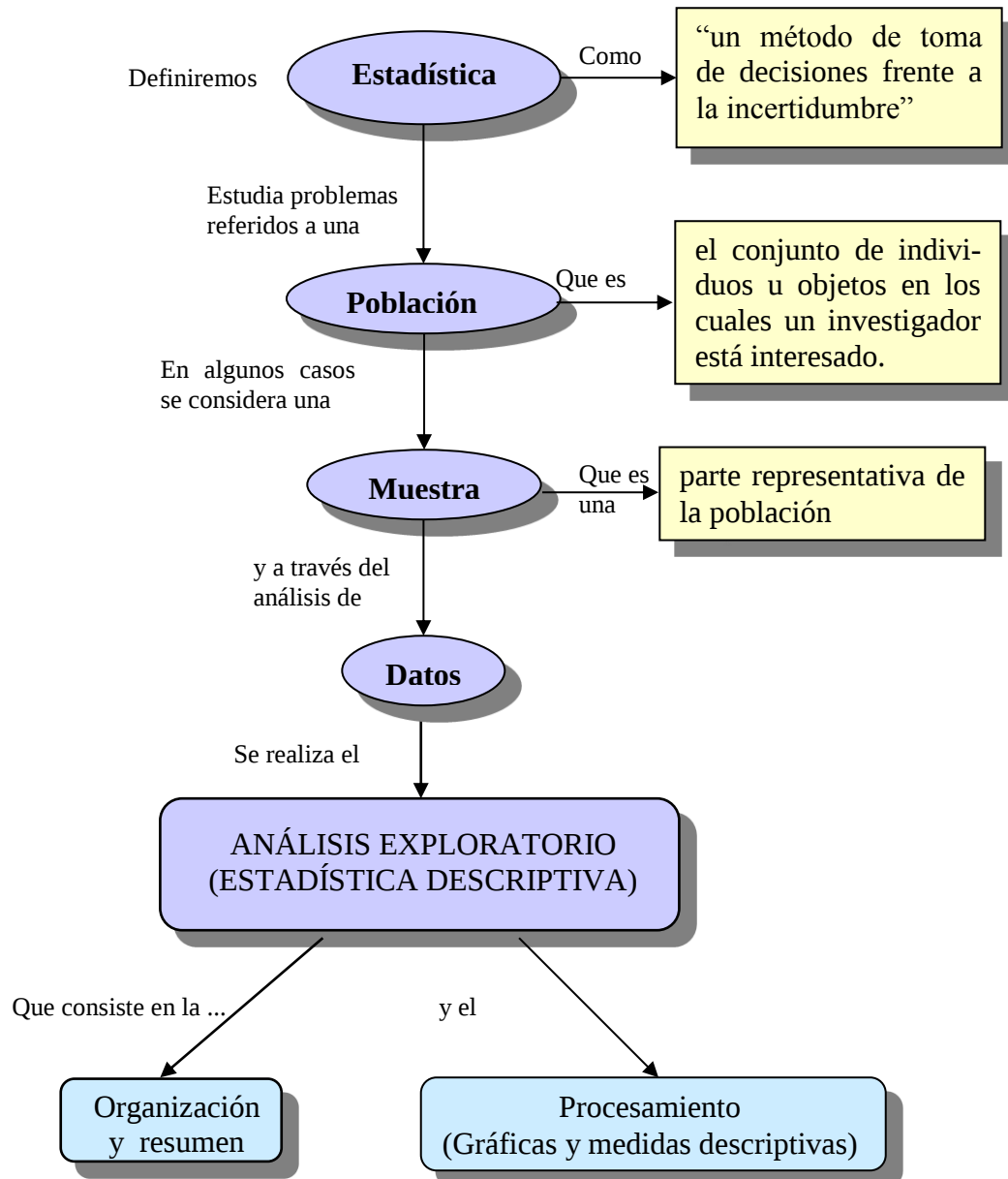


Diagrama 1

I.- VARIABLES: Clasificación. Escalas de medición

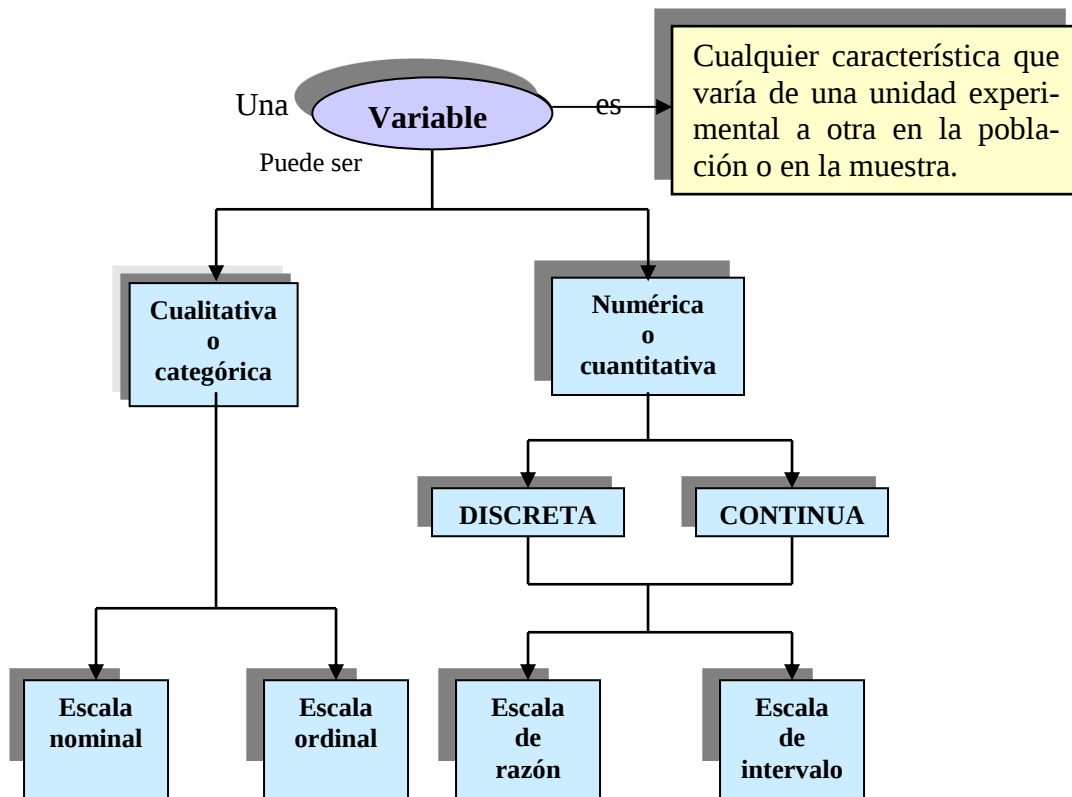


Diagrama 2

Clasificar las siguientes variables e indicar la escala de medición

- Cantidad de alumnas egresadas del Colegio de Huerto durante los últimos diez años.
- Temperatura diaria de Jujuy durante el mes de Julio de 2000.
- Lugar de nacimiento de las personas que viven en San Salvador de Jujuy.
- Nivel educacional de los padres de los alumnos de la Facultad de Ciencias Económicas.

Para la clasificación de las variables debemos considerar los valores que toma la variable X

1) Si los valores son numéricos, la variable es **numérica** o **cuantitativa**. Si el número proviene de contar es **discreta**, si resulta de efectuar mediciones es **continua**.

- X: Cantidad de alumnas egresadas del Colegio de Huerto durante los últimos diez años.

La cantidad de egresadas puede ser 42, 58, etc. Son cantidades enteras que provienen de contar a las alumnas egresadas. Entonces la variable es **cuantitativa discreta**.

b) X: Temperatura diaria de Jujuy.

La temperatura diaria puede ser 12°, 23° o puede estar comprendida entre 21° y 22° etc. Estas cantidades provienen de realizar mediciones de la temperatura diaria. Entonces la variable es **cuantitativa continua**.

Para reconocer la escala de medición, en una variable cuantitativa observamos:

Si la variable tiene un cero natural, es decir si el cero indica la ausencia de la característica, la escala es de **Razón**.

Para el caso a) 0 indicaría la ausencia de alumnas egresadas, entonces la **escala es de Razón**.

Para el caso c) de las mediciones de la temperatura 0° no indica la ausencia de temperatura, el 0 considerado es un 0 convencional, en este caso la escala de medición es **de intervalo**.

II) Para el caso en que los valores de la variable están asociados a distintas categorías o cualidades, la variable es **categorica o cualitativa**.

c) X: Lugar de nacimiento de personas que viven en Jujuy.

Los lugares de nacimiento podrían ser El Carmen, Palpalá, Humahuaca, etc. Cada uno de los lugares de nacimiento constituye una categoría o una cualidad, entonces X es **categorica**. En este caso las categorías o cualidades no están relacionadas, no tienen un orden preestablecido, en este caso la escala de medición es **nominal**.

d) X: Nivel educacional de los padres de los alumnos de la Facultad de Ciencias Económicas.

En este caso los valores de la variable pueden agruparse en categorías ordenadas: primario, medio, terciario, etc.. entonces la variable es categorica o cualitativa y la escala de medición es **ordinal**.

II) Organización y resumen de datos

Problema 1: Un ingeniero agrónomo visita 50 plantaciones de naranja en la zona del ramal y en cada una anotó el número de plantas atacadas por cierto hongo, de lo cual resultaron los siguientes datos:

15	16	18	18	15	15	17	20	17	20
20	17	18	18	16	16	17	20	17	20
25	18	19	19	15	21	18	21	16	21
15	20	16	18	16	21	19	19	19	21
18	18	17	19	20	17	19	19	17	21

- a. Defina la variable considerada. Clasifíquela e indique la escala de medición.
- b. Construya una tabla de distribución de frecuencias, interprételas y represéntelas gráficamente.
- c. ¿Cuántas plantaciones tienen a lo sumo 19 plantas atacadas por hongos?.
- d. ¿Cuántas plantaciones tienen por lo menos 17 plantas atacadas por hongos?.
- e. ¿Qué proporción de las plantaciones bajo estudio tienen 18 plantas atacadas por hongos?
- f. ¿Qué proporción de las plantaciones bajo estudio tienen 18 o menos plantas atacadas por hongos?
- g. ¿Qué proporción de las plantaciones bajo estudio tienen 18 o más plantas atacadas por hongos?

4

Solución

a) Definición de la variable

X: Cantidad de plantas atacadas por cierto hongo.

Clasificación

Cómo cada valor que toma la variable proviene de contar las plantas, entonces la variable es **cuantitativa discreta**.

Escala de medición

Si teóricamente X tomara el valor 0 , entonces 0 indica la ausencia de plantas atacadas por hongos, entonces la escala de medición **es de razón**.

b) Construya una tabla de distribución de frecuencias, interprételas y represéntelas gráficamente.

El conjunto de observaciones supera los 25 datos, en este caso, la forma conveniente de examinar los datos es presentarlos en forma de resumen mediante la construcción de tablas y gráficas apropiadas.

Para la organización de los datos se considera el tamaño de la muestra. A este tamaño designaremos con **n**. En nuestro problema $n = 50$ y la variable es cuantitativa discreta. La información recogida puede ser resumida en una tabla de frecuencias.

Para este tipo de variable, es una tabla en la que se asocia cada valor de la variable con la cantidad de veces que se observa dicho valor. Se pueden calcular también frecuencias relativas, porcentuales y acumuladas.

Tabla I.- Tabla genérica de distribución de frecuencias de una variable cuantitativa discreta para una muestra de tamaño = n

Variable Diferentes valores ob- servados x_i	Frecuencia Absoluta f_i	Frecuencia Relativa r_i	Frecuencia Relativa Porcentual $r_i \%$	Frecuencia Absoluta Acumulada Menor que $F_i \downarrow$	Frecuencia Absoluta Acumulada Mayor que $F'_i \uparrow$	Frecuencia Relativa Acumulada Menor que $R_i \downarrow$	Frecuencia Relativa Acumulada Mayor que $R'_i \uparrow$
x_1	f_1	r_1	$100r_1$	F_1	$F'_1 = n$	R_1	$R'_1 = 1$
x_2	f_2	r_2	$100r_2$	F_2	F'_2	R_2	R'_2
.....							
x_m	f_m	r_m	$100r_m$	$F_m = n$	F'_m	$R_m = 1$	R'_m
Totales	$\sum f_i = n$	$\sum r_i = 1$	100				

Tabla II.- Número de plantas de naranjas atacadas por un cierto hongo

Variable Diferentes valores ob- servados x_i	Frecuencia Absoluta f_i	Frecuencia Relativa r_i	Frecuencia Relativa Porcentual $r_i \%$	Frecuencia Absoluta Acumulada Menor que $F_i \downarrow$	Frecuencia Absoluta Acumulada Mayor que $F'_i \uparrow$	Frecuencia Relativa Acumulada Menor que $R_i \downarrow$	Frecuencia Relativa Acumulada Mayor que $R'_i \uparrow$
15	5	0.1	10%	5	50	0.1	1
16	6	0.12	12%	11	45	0.22	0.9
17	8	0.16	16%	19	39	0.38	0.78
18	9	0.18	18%	28	31	0.56	0.62
19	8	0.16	16%	36	22	0.72	0.44
20	7	0.14	14%	43	14	0.86	0.28
21	6	0.12	12%	49	7	0.98	0.14
25	1	0.02	2%	50	1	1	0.02
Totales	50	1	100				

1) Frecuencia absoluta:

Las frecuencias absolutas se obtienen observando las veces que se repite cada valor de la variable.

En nuestro problema, los valores que toma la variable X son: 15, 16, 17, 25 plantas atacadas por el hongo.

15 plantas atacadas aparecen en 5 plantaciones.

16 plantas atacadas aparecen en 6 plantaciones.

17 plantas atacadas aparecen en **8** plantaciones.

.....

21 plantas atacadas aparecen en **6** plantaciones.

25 plantas atacadas aparecen en **1** plantación.

Interpretación:

$f_1 = 5$, significa: 5 plantaciones de las observadas tienen 15 plantas de naranjas atacadas por cierto hongo.

$f_2 = 6$, significa: 6 plantaciones de las observadas tienen 16 plantas de naranjas atacadas por cierto hongo.

.....

$f_7 = 1$, significa que 1 plantación de las observadas tiene 25 plantas de naranjas atacadas por cierto hongo.

2) Frecuencia relativa:

La frecuencia relativa de un valor de la variable es la proporción de veces que ocurre dicho valor.

La frecuencia relativa $r_1 = f_1 / n = 5 / 50 = 0.1$

La frecuencia relativa $r_2 = f_2 / n = 6 / 50 = 0.12$

.....

La frecuencia relativa $r_8 = f_8 / n = 1 / 50 = 0.01$

Interpretación

0.1 es la proporción de plantaciones de la muestra que tiene 15 plantas atacadas.

0.12 es la proporción de plantaciones de la muestra que tiene 16 plantas atacadas.

.....

0.02 significa la proporción de plantaciones de la muestra que tiene 25 plantas atacadas.

3) Frecuencia relativa porcentual

Es la frecuencia relativa multiplicada por 100.

$r_1 \times 100 = 0.1 \times 100 = 10\%$

$r_2 \times 100 = 0.12 \times 100 = 12\%$

.....

$r_8 \times 100 = 0.02 \times 100 = 2\%$

Interpretación

El **10%** de las plantaciones de la muestra tiene 15 plantas atacadas.

El **12%** de las plantaciones de la muestra tiene 16 plantas atacadas.

.....

El **2%** de las plantaciones de la muestra tiene 1 planta atacada.

Frecuencias acumuladas

La frecuencia acumulada para un valor dado de la variable es la suma de las frecuencias (absolutas o relativas) de los valores menores o iguales al valor que se está considerando.

7

$$F_1 = f_1 = 5$$

$$F_2 = F_1 + f_2 = 5 + 6 = 11$$

$$F_3 = F_2 + f_3 = 11 + 8 = 19$$

.....

$$F_8 = F_7 + f_8 = 49 + 1 = 50$$

Interpretación

11 significa que: 11 plantaciones poseen 16 plantas atacadas o menos.

19 significa que: 19 plantaciones poseen a lo sumo 17 plantas atacadas.

.....

5) Frecuencia absoluta acumulada “Mayor que”

$$F'_1 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 = n = 50$$

$$F'_2 = f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 = 6 + 8 + 9 + 8 + 7 + 6 + 1 = 45$$

$$F'_3 = f_3 + f_4 + f_5 + f_6 + f_7 + f_8 = 8 + 9 + 8 + 7 + 6 + 1 = 39$$

.....

$$F'_8 = f_8 = 1$$

Otra forma

$$F'_1 = n = 50$$

$$F'_2 = F'_1 - f_1 = 50 - 5 = 45$$

$$F'_3 = F'_2 - f_2 = 45 - 6 = 39$$

$$F'_4 = F'_3 - f_3 = 39 - 8 = 31$$

.....

$$F'_8 = F'_7 - f_7 = 7 - 6 = 1$$

Interpretación

F'_2 = 45 significa que: 45 plantaciones de la muestra tienen 16 o más plantas atacadas por el hongo.

F'_4 = 31 significa que: 31 plantaciones de la muestra tienen 18 o más plantas atacadas por el hongo.

En forma similar se calculan las frecuencias relativas acumuladas.

6) Frecuencias relativas acumuladas “Menor que”

$$R_1 = r_1 = 0.1$$

$$R_2 = R_1 + r_2 = 0.1 + 0.12 = 0.22$$

$$R_3 = R_2 + r_3 = 0.22 + 0.16 = 0.38$$

.....

$$R_8 = R_7 + r_8 = 0.98 + 0.02 = 1$$

Interpretación

0.22 significa que: 0.22 es la proporción de plantaciones de la muestra que tienen 16 o menos plantas atacadas.

0.38 significa que: 0.38 es la proporción de plantaciones de la muestra que tienen 17 o menos plantas atacadas.

.....

7) Frecuencias relativas acumuladas “Mayor que”

$$R'_1 = r'_1 + r'_2 + r'_3 + r'_4 + r'_5 + r'_6 + r'_7 + r'_8 = 1$$

$$R'_2 = r'_2 + r'_3 + r'_4 + r'_5 + r'_6 + r'_7 + r'_8 = 0.12 + 0.16 + 0.18 + 0.16 + 0.14 + 0.12 + 0.02 = 0.9$$

$$R'_3 = r'_3 + r'_4 + r'_5 + r'_6 + r'_7 + r'_8 = 0.16 + 0.18 + 0.16 + 0.14 + 0.12 + 0.02 = 0.78$$

$$R'_4 = r'_4 + r'_5 + r'_6 + r'_7 + r'_8 = 0.18 + 0.16 + 0.14 + 0.12 + 0.02 = 0.62$$

.....

$$R'_8 = r_8 = 0.02$$

Otra forma

$$R'_1 = 1$$

$$R'_2 = R'_1 - r'_1 = 1 - 0.1 = 0.9$$

$$R'_3 = R'_2 - r'_2 = 0.9 - 0.12 = 0.78$$

$$R'_4 = R'_3 - r'_3 = 0.78 - 0.16 = 0.62$$

.....

$$R'_8 = R'_7 - r'_7 = 0.14 - 0.12 = 0.02$$

Interpretación

0.9 significa que: 0.9 es la proporción de las plantaciones de la muestra que tienen 16 o más plantas atacadas.

0.78 significa que: 0.78 es la proporción de las plantaciones de la muestra que tienen 17 o más plantas atacadas.

Representación gráfica

Por tratarse de una variable discreta, la distribución de las frecuencias absolutas y relativas se realiza a través del gráfico de bastones; las frecuencias acumuladas a través del gráfico de escalones.

Gráfico de bastones:

El gráfico de bastones se usa para representar a las distintas distribuciones de frecuencias correspondientes a datos discretos.

Sobre el eje de abscisas o de las “x” se representan los valores de la variable; en nuestro problema la cantidad de plantas atacadas. Sobre el eje de las ordenadas o de las “y”, las frecuencias absolutas, las frecuencias relativas o las porcentuales.

Sobre cada valor de la variable se levanta un segmento vertical de longitud igual a la frecuencia absoluta (o la relativa, o la porcentual) del valor correspondiente.

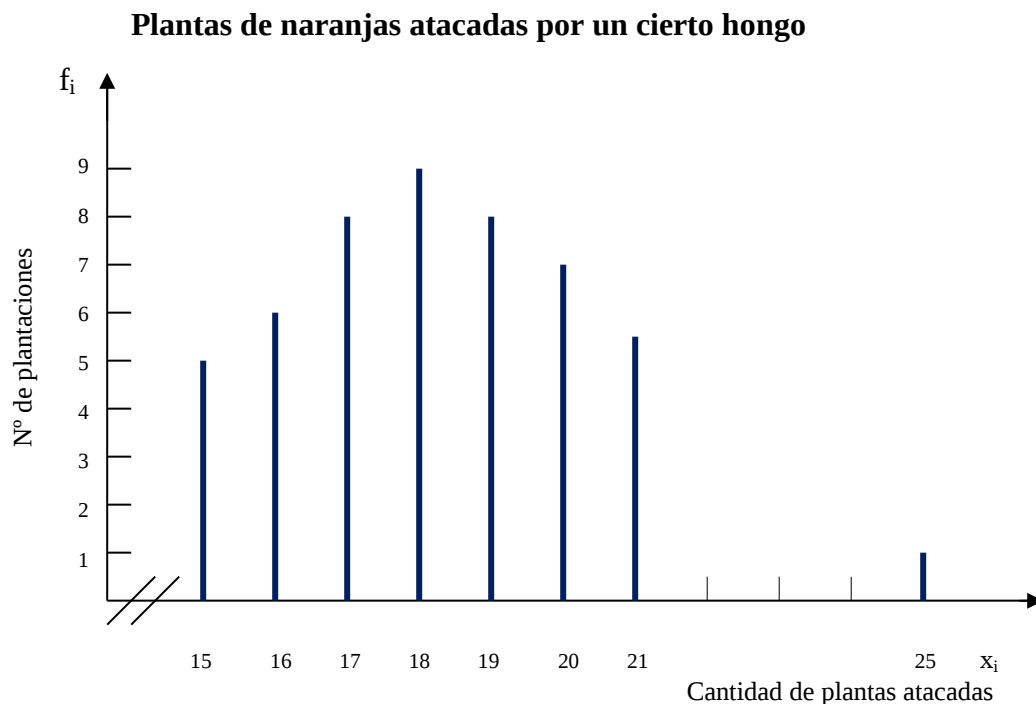
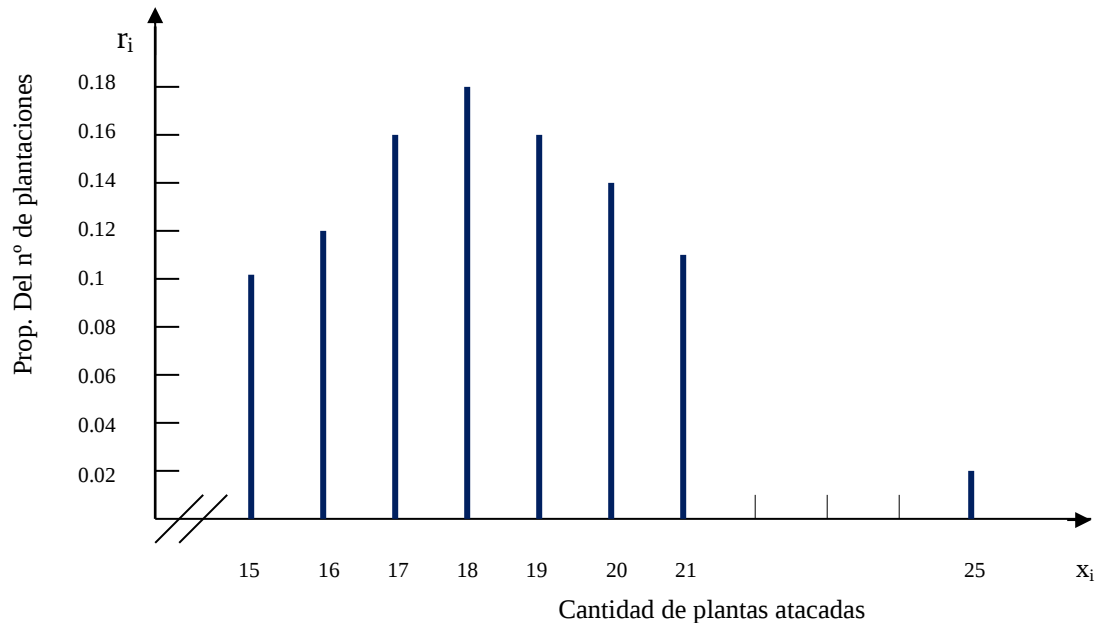
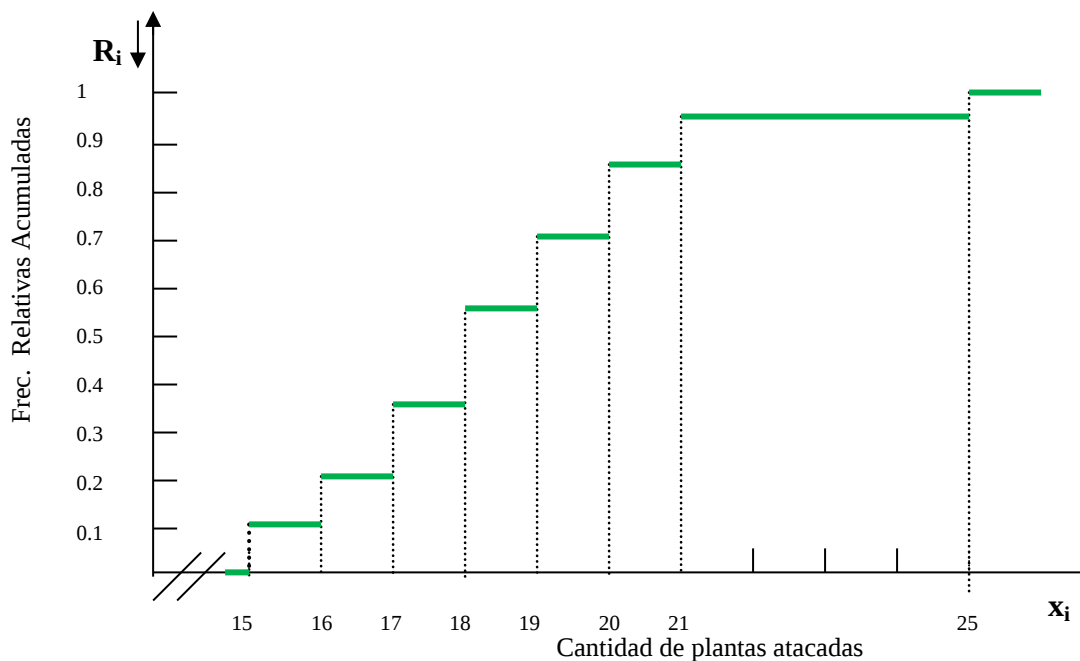


Gráfico 1: Gráfico a bastones de las frecuencias absolutas

**Gráfico 2:** Gráfico a bastones de las frecuencias relativasGráfico de escalones

Las frecuencias acumuladas (absolutas, relativas o porcentuales) “Mayor que” y “Menor que” de una variable discreta se representan por el gráfico de escalones.

Sobre el eje horizontal de la “x” se representan cada uno de los valores de la variable, en nuestro problema: cantidad de plantas atacadas. Por cada uno de estos valores se levanta un segmento vertical igual a la frecuencia acumulada y se completan con tramos horizontales hasta el valor inmediato siguiente.

**Gráfico 3:** Gráfico a escalones de las frecuencias relativas acumuladas “Menor que”

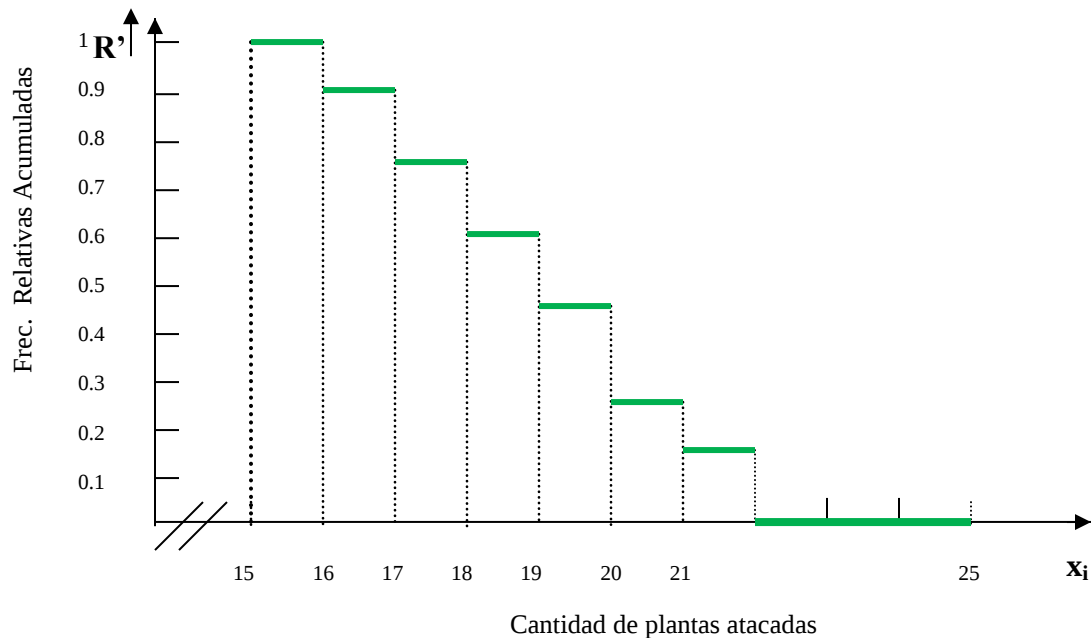


Gráfico 4: Gráfico a escalones de las frecuencias relativas acumuladas “Mayor que”

c) ¿Cuántas plantaciones tienen a lo sumo 19 plantas atacadas por hongos?.

Respuesta: 36 plantaciones tienen a lo sumo 19 plantas atacadas.

Esta respuesta se la puede leer en la tabla o en el gráfico de escalones. Este número se encuentra en la tabla II, en la intersección de la fila correspondiente al valor de 19 plantas atacadas con la columna de frecuencia absoluta acumulada “menor que”.

De acuerdo a la interpretación de las frecuencias absolutas menor que: 36 plantaciones poseen 19 plantas afectadas o menos.

d) ¿Cuántas plantaciones tienen por lo menos 17 plantas atacadas por hongos?.

En esta caso se utiliza la columna de frecuencia absoluta acumulada “mayor que” y la fila correspondiente a 17 plantas afectadas. En la intersección tenemos 39 plantaciones.

Respuesta: 39 plantaciones tienen por lo menos 17 plantas atacadas.

Problema 2

Suponga que se han registrado 50 observaciones referentes a los pesos de 50 lingotes de acero producidos por SIDER, la muestra fue obtenida de la producción semanal y las unidades están dadas en Kg.

94.5	93.0	95.5	95.3	92.4	94.4	92.8	93.2	93.6	95.5
92.9	93.6	95.7	93.8	94.8	93.9	92.7	91.6	93.6	93.7
94.2	95.7	94.7	94.3	92.7	94.5	96.2	95.4	93.7	91.9
94.7	92.7	95.0	93.0	92.9	93.7	92.7	93.3	94.6	96.4
94.1	93.7	94.2	93.7	94.0	93.9	93.6	94.6	92.3	94.4

- a) Defina la variable considerada. Clasifíquela e indique la escala de medición.
- b) Organizar las observaciones realizadas en una tabla de **distribución de frecuencias** con 5 clases de igual amplitud.
- c) Interprete las distintas frecuencias y construya las gráficas correspondientes.
- d) ¿Cuántos lingotes de acero pesan 94.5 Kg. o más?
- e) ¿Cuántos lingotes de acero pesan más de 93.5 Kg.?
- f) ¿Qué porcentaje de lingotes de acero pesan menos de 94.5 Kg. ?

Solución

a) Definición de la variable

La variable X es: Peso de lingotes de acero expresado en Kg.

Clasificación

Los valores que toma la variable son valores numéricos que resultan de realizar mediciones, entonces se trata de una variable *numérica o cuantitativa continua*.

Escala de medición

El valor de 0 Kg. indicaría la ausencia de peso, es un cero natural, por lo que la escala de medición corresponde a la *de razón*.

b) Distribución de frecuencias

Para construir una tabla de distribución de frecuencias recordaremos que: “**Una distribución de frecuencias es una tabla de resumen en la que los datos se agrupan o arreglan en clases o categorías ordenadas en forma numérica, establecidas de modo conveniente**”.¹

Como se trata de una variable continua, y que teóricamente la magnitud de la variable puede tomar una infinidad de valores y que entre dos pesos de la variable puede existir otro peso, la organización de los datos o tabla de distribución de frecuencias contiene intervalos de clase.

“Al construir la tabla de distribución de frecuencias, se debe prestar atención a:

1. Seleccionar el número apropiado de clases para la tabla.
2. Obtener un **intervalo de clase** o “ancho” apropiados para cada clase.
3. **Establecer los límites** y las **fronteras** de cada clase para evitar superposición.”⁴

^{1 y 4} Berenson, M.L.- Levine, D.M. Estadística para Administración y Economía. Conceptos y Aplicaciones.

Una tabla de distribución de frecuencias de una variable continua es como la que sigue.

Tabla III.- Tabla de distribución de frecuencias de una variable continua para una muestra de tamaño **n**.

Intervalos de clase $L_i - L_s$	Marca de clase x'_i	Frec. Abs. f_i	Frec. Abs. Acumul. Menor que $F_i \downarrow$	Frec. Abs. Acumul. Mayor que $F'_i \uparrow$	Frecuenc. Relativa r_i	Frec. Rel. Acumul. Menor que $R_i \downarrow$	Frec. Rel. Acumul. Mayor que $R'_i \uparrow$	Frec. Porcent. $r_i \%$	Frec. Porc. Acumul. Menor que $R_i \% \downarrow$	Frec. Porc. Acumul. Mayor que $R'_i \% \uparrow$
$L_0 - L_1$	x'_1	f_1	F_1	$F'_1 = n$	r_1	R_1	$R'_1 = 1$	$r_1 \%$	$R_1 \%$	$R'_1 \% = 100$
$L_1 - L_2$	x'_2	f_2	F_2	F'_2	r_2	R_2	R'_2	$r_2 \%$	$R_2 \%$	$R'_2 \%$
.....		...						...	...	
$L_{m-1} - L_m$	x'_m	f_m	$F_m = n$	$F'_m = f_m$	r_m	$R_m = 1$	$R'_m = r_m$	$r_m \%$	$R_m \% = 100$	$R'_m \% = r_m$
Totales		n			1			100		

Al observar los datos del problema, puede verse que el mayor valor que toma la variable peso de los lingotes de acero es de 96.4 Kg. y el menor valor es 91.6 Kg.

El rango o recorrido $r = x_i \text{ máx.} - x_i \text{ mín.} = 96.4 \text{ Kg.} - 91.6 \text{ Kg.} = 4.8 \text{ Kg.}$, que es la amplitud total.

“El rango se define como la diferencia que existe entre el mayor y el menor valor observado que toma la variable en estudio”.²

La decisión sobre la cantidad de intervalos y la amplitud de cada uno de ellos es arbitraria, depende del problema en cuestión y en última instancia del investigador. Los intervalos de clase pueden ser de igual o distinta amplitud dependiendo del problema o el criterio del investigador. Generalmente se utiliza entre 5 y 15 clases.

- En nuestro problema está establecido que debemos considerar 5 intervalos de igual amplitud.

La amplitud o ancho del intervalo = amplitud total o rango / n° de intervalos

$$\text{Amplitud} = 4.8 \text{ Kg.} / 5 = 0.96 \cong 1$$

Trabajar con 0.96 resultaría dificultoso por lo que es conveniente trabajar con el entero más próximo.

Ampliamos el rango o recorrido de datos original [91.6 , 96.4], por otro que lo contenga y que sea múltiplo de 5. El rango ampliado y que contiene al original es [91.5 , 96.5].

Determinación de los límites de clase

Para el primer intervalo:

Se toma el valor mínimo del rango ampliado, en nuestro caso 91.5 y este es el límite inferior. El límite superior se obtiene sumando 1 que es el ancho o amplitud para los datos de nuestro problema:

$$91.5 + 1 = 92.5. \text{ Entonces el primer intervalo es } 91.5 - 92.5$$

² Universidad Nacional de Córdoba. Curso de Posgrado. “ESTADÍSTICA APLICADA A LA INVESTIGACIÓN”

Segundo intervalo:

El límite superior del primero pasa ser límite inferior del segundo, sumándole a éste el ancho del intervalo se obtiene el límite superior: $92.5 - 93.5$.

Con el mismo procedimiento se obtienen los demás intervalos.

Tabla IV.- Tabla de distribución de frecuencias de los pesos de los 50 lingotes de acero

Intervalo de clase $L_i - L_s$	Marca de clase x'_i	Frec. Abs. f_i	Frec. Abs. Acumul. Menor que $F_i \downarrow$	Frec. Abs. Acumul. Mayor que $F'_i \uparrow$	Frec. Rel. r_i	Frec. Rel. Acumul. Menor que $R_i \downarrow$	Frec. Rel. Acumul. Mayor que $R'_i \uparrow$	Frec. Porcent. $r_i \%$	Frec. Porc. Acumul. Menor que $R_i \% \downarrow$	Frec. Porc. Acumul. Mayor que $R'_i \% \uparrow$
91.5 – 92.5	92	4	4	50	0.08	0.08	1	8	8	100
92.5 – 93.5	93	11	15	46	0.22	0.30	0.92	22	30	92
93.5 – 94.5	94	20	35	35	0.40	0.70	0.70	40	70	70
94.5 – 95.5	95	9	44	15	0.18	0.88	0.30	18	88	30
95.5 – 96.5	96	6	50	6	0.12	1	0.12	12	100	12
Totales		50			1			100		

Ubicación de los pesos en los intervalos correspondientes

El peso del lingote igual a 94.5 Kg. ¿en qué intervalo considerarlo? ¿ en el tercer o cuarto intervalo?. Se puede ubicar en cualquiera de los dos, pero el criterio adoptado se debe considerar en todos los casos y para todo el estudio.

- En nuestro problema y en todos los que se consideren, el criterio es:

91.5 – 92.5 : peso igual o mayor que 91.5 pero menos de 92.5

92.5 – 93.5 : peso igual o mayor que 92.5 pero menos de 93.5

93.5 – 94.5 : peso igual o mayor que 93.5 pero menos de 94.5

94.5 – 95.5 : peso igual o mayor que 94.5 pero menos de 95.5

95.5 – 96.5 : peso igual o mayor que 95.5 pero menos de 96.5

- Entonces 94.5 ubicaremos en el cuarto intervalo.

Para el caso del peso 92.9, éste pertenece al segundo intervalo y así se continua con los demás datos.

Frecuencias

Se pueden calcular, además de las frecuencias absolutas, las relativas o las porcentuales, según el problema en estudio requiera proporciones o porcentajes respectivamente.

- En nuestro problema calcularemos todas las frecuencias ya que no se especifica cual se prefiere.

El cálculo de frecuencias relativas y acumuladas se realiza del mismo modo que se trabajó con

Para las frecuencias absolutas se consideran la cantidad de lingotes de acero contenidos en cada intervalo.

c) Interpretaciones

Frecuencia absoluta

- 4:** 4 lingotes de acero pesan menos de 92.5
11: 11 lingotes pesan 92.5 Kg. o más, pero menos de 93.5
20: 20 lingotes pesan 93.5 Kg. o más, pero menos de 94.5
9 : 9 lingotes pesan 94.5 Kg. o más, pero menos de 95.5
6 : 6 lingotes pesan 95.5 Kg. o más, pero menos de 96.5

Frecuencia Absoluta acumulada “Menor que”

- 4 :** 4 es el número de lingotes de acero que pesan menos de 92.5 Kg.
15: 15 es el número de lingotes de acero que pesan menos de 93.5 Kg.
35: 35 lingotes de acero pesan menos de 94.5 Kg.
44: 44 lingotes de acero pesan menos de 95.5 Kg.

.....

Frecuencia Absoluta acumulada “Mayor que”

- 50:** 50 lingotes de acero pesan 91.5 kg. o más.
46: 46 lingotes de acero pesan 92.5 kg. o más
o 46 lingotes de acero pesan por lo menos 92.5 kg.
35: 35 lingotes de acero pesan 93.5 kg. o más.
15: 15 lingotes de acero pesan 94.5 kg. o más.
6: 6 lingotes de acero pesan 95.5 kg. o más.

Frecuencia relativa

Se divide cada frecuencia absoluta por el tamaño de la muestra. En nuestro problema se divide por 50.

- 0.08:** 0.08 es la proporción de lingotes de acero que pesan 91.5 Kg. o más pero menos de 92.5 Kg.
0.22: 0.22 es la proporción de lingotes de acero que pesan 92.5 Kg. o más pero menos de 93.5 Kg.

.....

- 0.12:** 0.12 es la proporción de lingotes de acero que pesan 95.5 Kg. o más pero menos de 96.5 Kg.

Frecuencia Absoluta acumulada “Menor que”

- 0.08:** 0.08 es la proporción de lingotes de acero que pesan menos de 92.5 Kg.

.....

- 0.70:** 0.70 es la proporción de lingotes de acero que pesan menos de 94.5 Kg.

0.88: 0.88 es la proporción de lingotes de acero que pesan menos de 95.5 Kg.

Frecuencia Absoluta acumulada “Mayor que”

0.92: 0.92 es la proporción de lingotes de acero que pesan 92.5 Kg o más.

o 0.92 es la proporción de lingotes de acero que pesan por lo menos 92.5 Kg.

0.70: 0.70 es la proporción de lingotes de acero que pesan 93.5 Kg o más.

o es la proporción de lingotes de acero que pesan por lo menos 93.5 Kg.

16

.....
0.12: 0.12 es la proporción de lingotes de acero que pesan 95.5 Kg o más.

o es la proporción de lingotes de acero que pesan por lo menos 95.5 Kg.

Frecuencia porcentual

8: El 8% de los lingotes de acero pesan 91.5 Kg o más pero menos de 92.5 Kg.

22: El 22% de los lingotes de acero pesan 92.5 Kg o más pero menos de 93.5 Kg.

.....
18: El 18% de los lingotes de acero pesan 94.5 Kg o más pero menos de 95.5 Kg.

Frecuencia porcentual acumulada “Menor que”

30: El 30% de los lingotes de acero pesan menos de 93.5 Kg.

70: El 70% de los lingotes de acero pesan menos de 94.5 Kg.

88: El 88% de los lingotes de acero pesan menos de 95.5 Kg.

Frecuencia porcentual acumulada “Mayor que”

92: El 92% de los lingotes de acero pesan 92.5 Kg. o más.

o el 92% de los lingotes de acero pesan por lo menos 92.5 Kg.

.....
30: El 30% de los lingotes de acero pesan 94.5 Kg. o más

o el 92% de los lingotes de acero pesan por lo menos 94.5 Kg.

La marca de clase

La **marca de clase** es el punto medio del intervalo de clase, es el valor utilizado para representar todos los datos resumidos en el intervalo en particular.

Por ejemplo, para el primer intervalo:

$$\frac{91.5 + 92.5}{2} = 92 \quad \text{y así se calcula para el resto de los intervalos.}$$

Los **histogramas** y los **polígonos** de frecuencias se utilizan para representar gráficamente las distribuciones de frecuencias absolutas, relativas y porcentuales de datos cuantitativos continuos agrupados en clase. Los histogramas son gráficas de barras verticales.

Para las frecuencias acumuladas “Menor que” y “Mayor que” se usan polígonos llamados **ojivas**.

Distribución del peso de 50 lingotes de acero

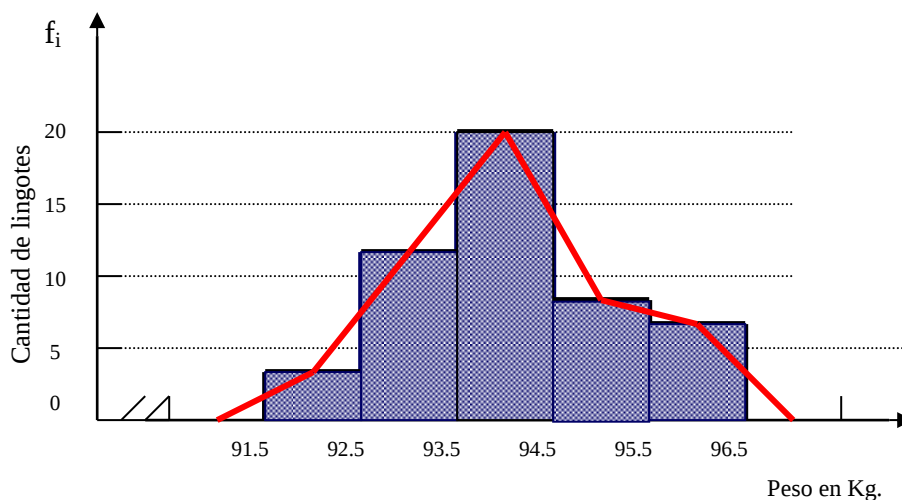


Gráfico 5: Histograma y polígono de frecuencias absolutas del peso de los lingotes de acero.

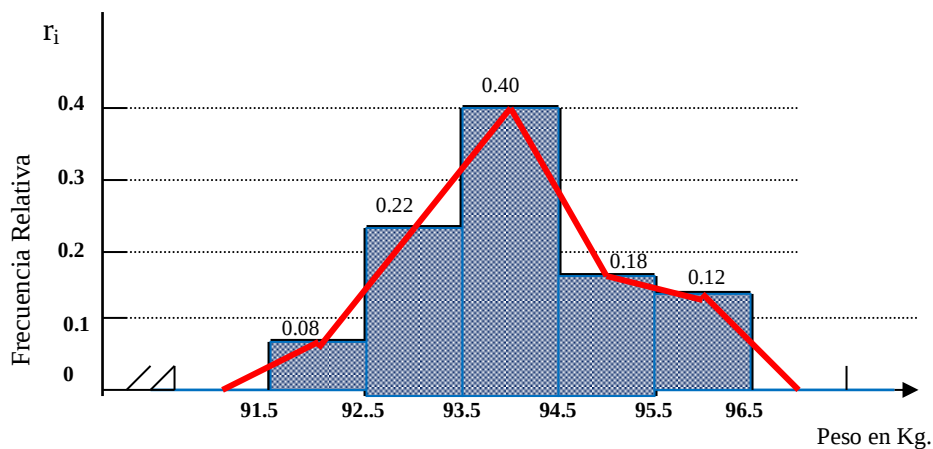


Gráfico 6: Histograma y polígono de frecuencia relativa de los pesos de lingotes de acero

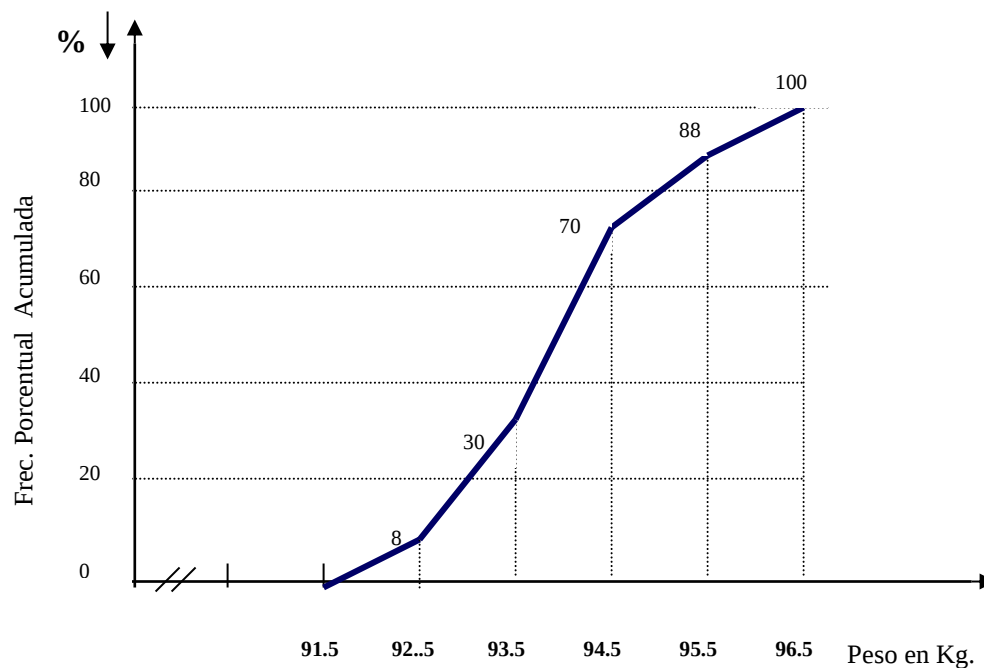
Polígonos de frecuencias acumuladas. Ojiva.

Gráfico 7: Frecuencias porcentuales acumuladas "Menor que"

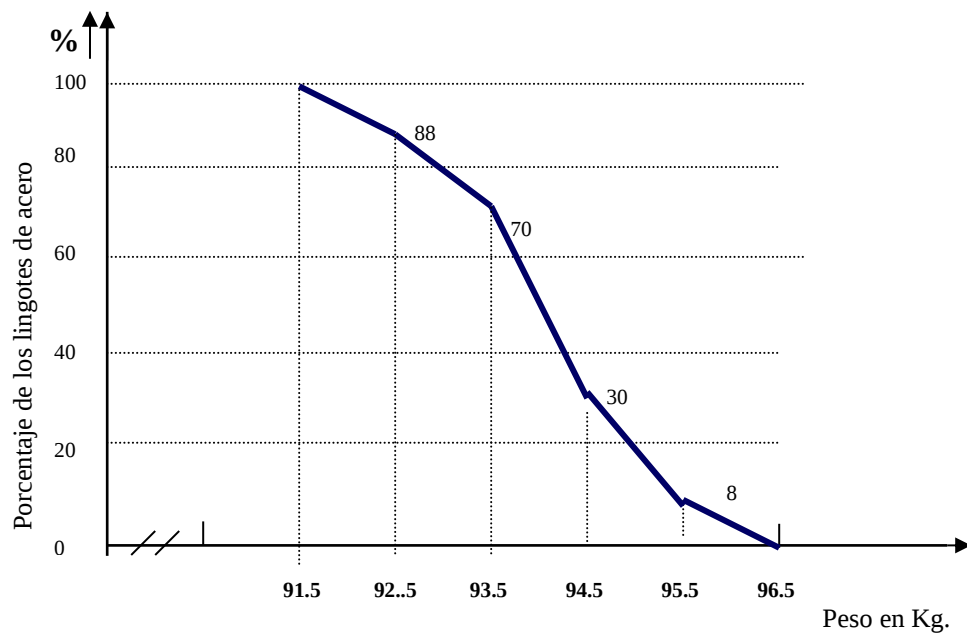


Gráfico 8: Frecuencias porcentuales acumuladas "Mayor que"

Diagrama resumen para la gráfica de datos

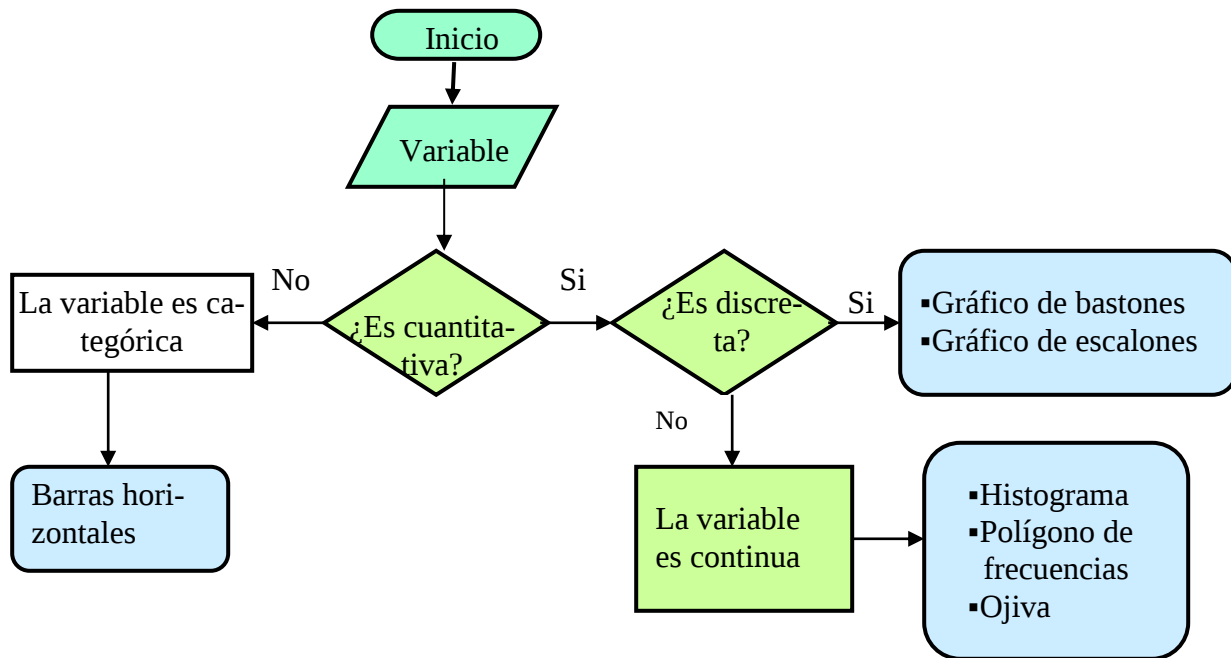


Diagrama 3

Problema 3

En una encuesta efectuada en la Facultad de Ciencias Económicas se preguntó a los alumnos de primer año acerca de su preferencia en cuanto a las tres áreas fundamentales de la carrera: Matemática, Contabilidad y Economía. En una muestra de 40 alumnos se obtuvieron las siguientes respuestas:

Alumno N°	Preferencia	Alumno N°	Preferencia
1	Economía	21	Matemática
2	Matemática	22	Economía
3	Matemática	23	Contabilidad
4	Contabilidad	24	Contabilidad
5	Economía	25	Economía
6	Contabilidad	26	Matemática
7	Contabilidad	27	Economía
8	Matemática	28	Economía
9	Economía	29	Contabilidad
10	Economía	30	Contabilidad
11	Contabilidad	31	Matemática
12	Economía	32	Matemática
13	Matemática	33	Contabilidad
14	Economía	34	Economía
15	Contabilidad	35	Economía
16	Matemática	36	Contabilidad
17	Matemática	37	Matemática
18	Economía	38	Economía
19	Economía	39	Economía
20	Economía	40	Contabilidad

- Identifique la variable en estudio, a que tipo pertenece y la escala de medición.
- Resuma la información en una tabla de frecuencias.
- Represente las frecuencias mediante un gráfico.
- ¿Cuál es el área de preferencia predominante de los alumnos de primer año de la Facultad de Ciencias Económicas?.

Solución

- Identificación de la variable

X: Área de preferencia

Tipo de variable

Los valores que toma la variable en este caso son cualitativos o también podemos decir que se pueden establecer categorías: Matemática, Contabilidad o Economía. La variable es entonces **cualitativa o categórica**.

La escala de medición es **nominal**.

- Una tabla de frecuencias para variables categóricas es una tabla que asocia cada categoría de la variable con el número de veces que se repite dicha categoría. Esta tabla también se denomina tabla de distribución de frecuencias.

Los datos cualitativos o categóricos se ordenan en frecuencias absolutas, frecuencias relativas y relativas porcentuales.

La tabla de distribución de frecuencias adoptará la siguiente forma:

Tabla V: Tabla de distribución de frecuencias categóricas.

Variable (categorías)	Frecuencias absolutas	Frecuencias relativas	Frecuencia relativa porcentual
Categoría A	f_A	r_A	$100 r_A \%$
Categoría B	f_B	r_B	$100 r_B \%$
.....			
Categoría Z	f_Z	r_Z	$100 r_Z \%$
Totales	N	1	100%

Tabla VI: Tabla de distribución de frecuencias de a variable área de preferencia

Categoría: área de preferencia	Frecuencia absoluta f_i	Frecuencia relativa r_i	Frecuencia relativa Porcentual
Economía	17	0.425	42.5
Matemática	11	0.275	27.5
Contabilidad	12	0.30	30
Total	40	1	100

c) La tabla de frecuencia anterior se puede representar por medio de un gráfico de barras horizontales.

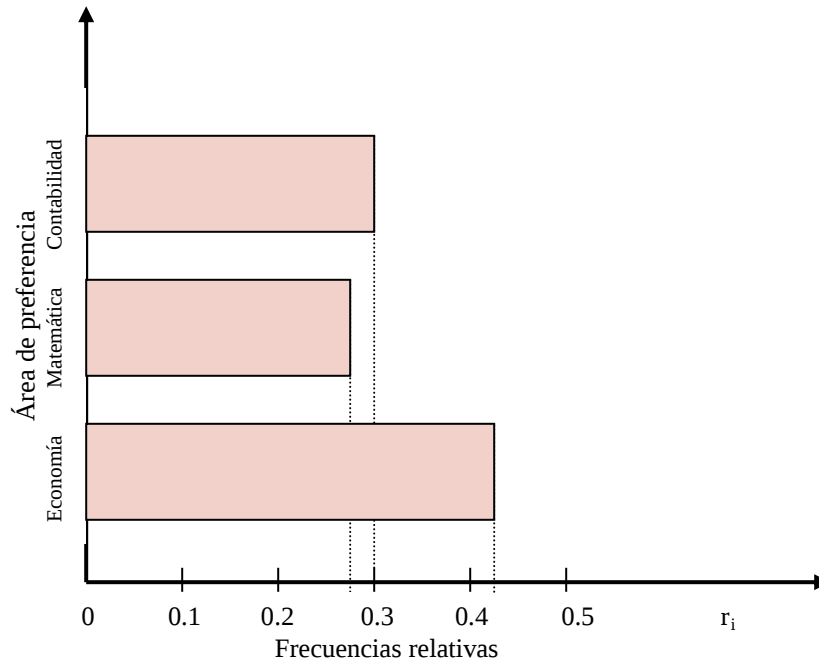


Gráfico 9: Frecuencias relativas de la variable área de preferencia

d) Para contestar esta pregunta observamos la tabla de frecuencias o el gráfico y se concluye que el área de preferencia que predomina es el de Economía.

III Medidas descriptivas

Hemos visto que es posible representar un conjunto de datos mediante tablas y gráficas dependiendo de la naturaleza de las variables. Para completar la descripción de datos vamos a determinar ahora valores numéricos que permitan caracterizar al conjunto de observaciones.

En todo conjunto de datos, se pueden calcular medidas descriptivas que representan las propiedades de posición, dispersión y forma.

La característica más importante que describe o resume un grupo de datos es su posición. La primera cuestión que se plantea en el análisis de las series de frecuencia es la de la representación de los valores de la serie, mediante un valor central. Para cualquier conjunto particular de datos, es posible seleccionar un valor típico para que lo describa o lo resuma. Este valor típico descriptivo se llama *promedio*. Las medidas comunmente utilizadas son: **media aritmética**, **moda** y **mediana**. Además de estas medidas, existen otras para una posición “no central”. Estas medidas se llaman **cuantiles** y las más utilizadas son los **cuartiles**, los **deciles** y los **percentiles**.

La segunda característica más importante que describe un conjunto de datos, es la dispersión.

“La dispersión es la cantidad de variación, desperdigamiento o diseminación en los datos”.³ Estas medidas son el *recorrido* o *rango*, el *recorrido intercuartílico*, la *desviación media*, la *desviación mediana*, la *varianza*, la *desviación estándar* y el *coeficiente de variación*.

Una tercera característica importante de un conjunto de datos es su forma, es decir la manera en la cual están distribuidos los datos. Estas características se llaman: *asimetría* y *curtosis* o *apuntamiento*.

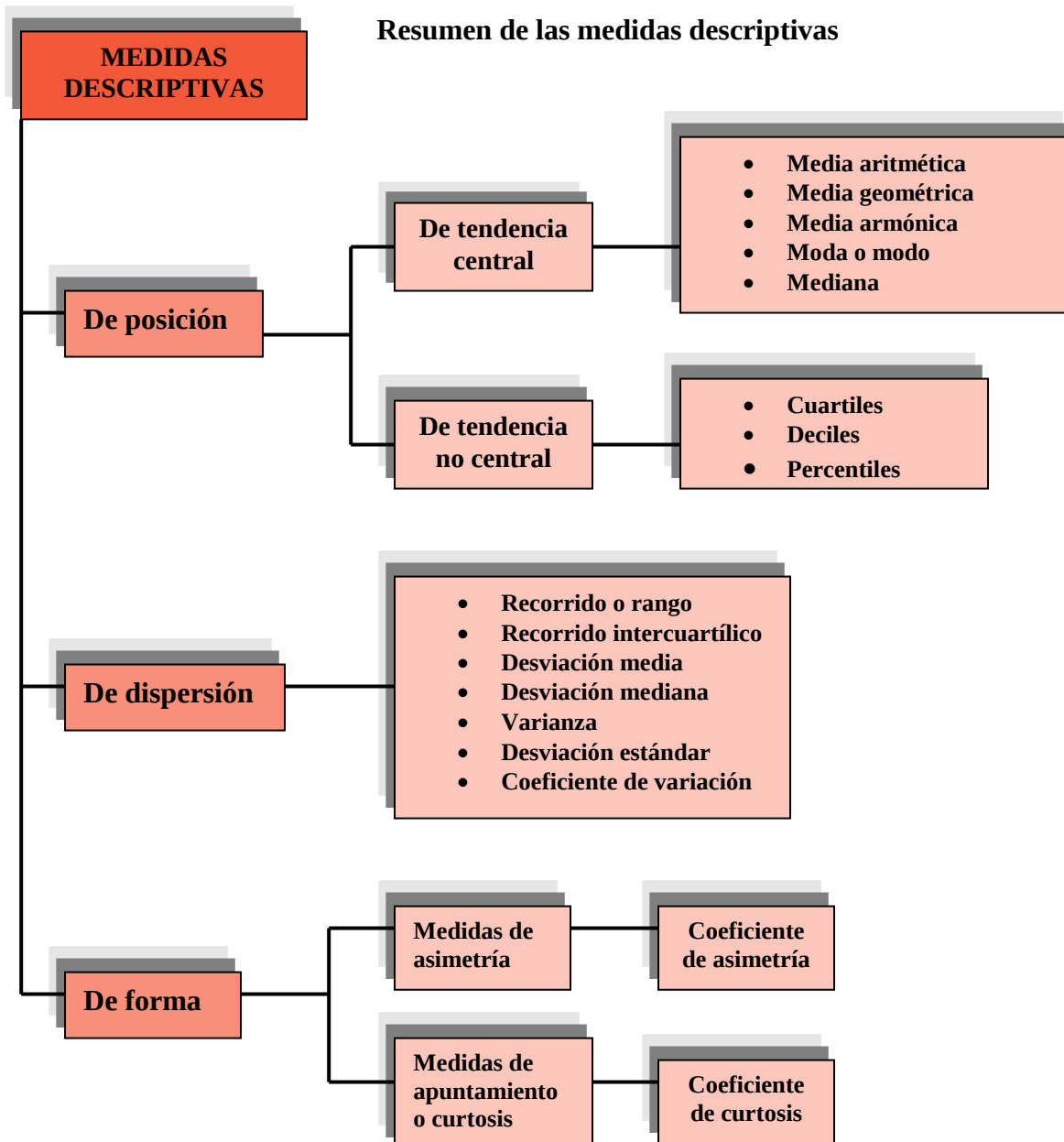


Diagrama 4

³ Berenson, M.L.- Levine, D.M. Estadística para Administración y Economía. Conceptos y Aplicaciones.

Problema 4

En un experimento se probó la eficacia de cierto insecticida. Para ello se consideraron 50 parcelas de un cultivo atacado por un insecto en particular. Al cabo de un cierto tiempo de aplicación del insecticida, se contó el número de insectos vivos por parcela y los datos se agruparon en la siguiente tabla de distribución de frecuencias:

Cantidad de insectos vivos por parcela

Cantidad de insectos	Cantidad de parcelas
5	3
6	9
7	10
8	12
9	8
10	5
11	3

En relación a este experimento :

- ¿Qué tipo de variable se está analizando?. Indique la escala de medición.
- Calcule la media aritmética de la cantidad de insectos vivos por parcela e interprete la misma.
- Determine la mediana y la moda de la distribución en forma analítica y gráfica. Interprete ambas medidas.
- Responda:
 - Por debajo de que cantidad de insectos vivos, se encuentra el 75% de las parcelas?.
 - Por debajo de que cantidad de insectos vivos, se encuentra el 25% de las parcelas?.
- Calcule e interprete el rango y una medida de dispersión apropiada.
- Calcule e interprete el coeficiente de asimetría y el de curtosis.

Solución

- ¿Qué tipo de variable se está analizando?. Indique la escala de medición.

Identificación de la variable

X: Cantidad de insectos vivos por parcela.

Como cada valor que toma la variable proviene de contar los insectos, entonces la variable es *cuantitativa discreta*. La escala de medición es *de razón*.

- Calcule la media aritmética de la cantidad de insectos vivos por parcela e interprete la misma.

Cálculo de la media aritmética

“La media actúa como *punto de equilibrio* o balanceo, de modo que las observaciones que son

mayores equilibran a las que son menores”.⁴

-En nuestro problema la variable es discreta y los datos están agrupados.

La fórmula para calcular la media aritmética para variables discretas con datos agrupados en una tabla

de frecuencias es: $\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{n}$ donde **m** es la cantidad de valores que toma la variable.

En este caso para hallar la media aritmética se agrega a la tabla de frecuencias una columna auxiliar que facilite el cálculo.

24

Tabla VII : Cantidad de insectos vivos por parcela

Cantidad de insectos x_i	Cantidad de parcelas f_i	$x_i f_i$
5	3	15
6	9	54
7	10	70
8	12	96
9	8	72
10	5	50
11	3	33
Totales	50	390

$$\bar{x} = \frac{\sum_{i=1}^m x_i f_i}{n} = 390 / 50 = 7,8 \cong 8 \text{ insectos vivos por parcela.}$$

El resultado matemático obtenido , no es un posible valor de la variable, por lo que tomamos el entero próximo: $\bar{x} = 8$ insectos vivos por parcela.

Interpretación: En promedio se observan 8 insectos vivos por parcela.

c. Determine en forma analítica y gráfica, la moda y la mediana de la distribución. Interprete ambas medidas.

Cálculo de la moda

“La moda es aquel valor de la variable que ocurre con más frecuencia.”⁵

-En la columna de frecuencias absolutas, la mayor frecuencia observada es 12 parcelas, y a 12 parcelas le corresponde **8 insectos** en la columna de los valores de la variable.

Tabla VIII : Cantidad de insectos vivos por parcela

Cantidad de insectos x_i	Cantidad de parcelas f_i
5	3
6	9
7	10
8	12
9	8
10	5
11	3
Totales	50

Entonces la **Mo** = 8 insectos vivos por parcela.

Interpretación: 8 insectos vivos es el número que se observa en la mayor cantidad de parcelas.

⁴ Berenson, M.L.- Levine, D.M. Estadística para Administración y Economía. Conceptos y Aplicaciones.

⁵ Universidad Nacional de Córdoba. Curso de Posgrado. “ESTADÍSTICA APLICADA A LA INVESTIGACIÓN”

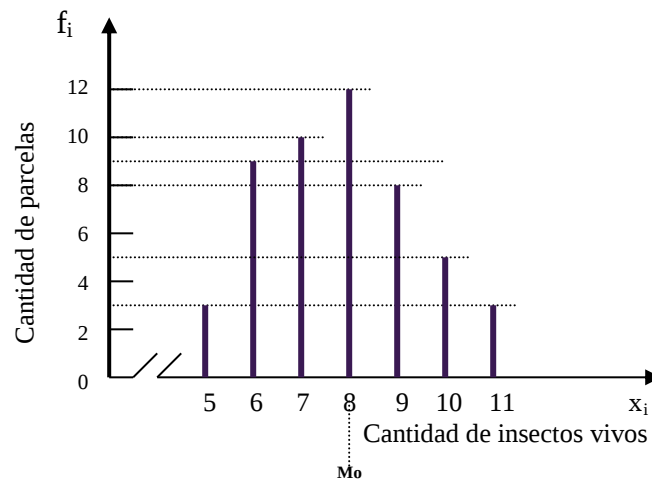


Gráfico 10: Gráfico de bastones para la distribución de frecuencias absolutas

Cálculo de la Mediana

“La mediana de un conjunto de observaciones es un valor de la variable que divide a este conjunto (ordenado de menor a mayor) en dos subconjuntos que contienen la misma cantidad de datos”.⁶

La mediana de un conjunto de n valores es el valor que no es superado ni supera a más de la mitad de las n observaciones.

Para el caso de datos agrupados

Tabla IX: Cantidad de insectos vivos por parcela

Cantidad de insectos x_i	Cantidad de parcelas f_i	Frec. Absol. Acumulada $F_i \downarrow$
5	3	3
6	9	12
7	10	22
8	12	34
9	8	42
10	5	47
11	3	50
Totales	50	

$M_e \rightarrow$

-Se calculan las frecuencias absolutas acumuladas de la forma “menor que”.

-Se calcula $n / 2 = 50 / 2 = 25$

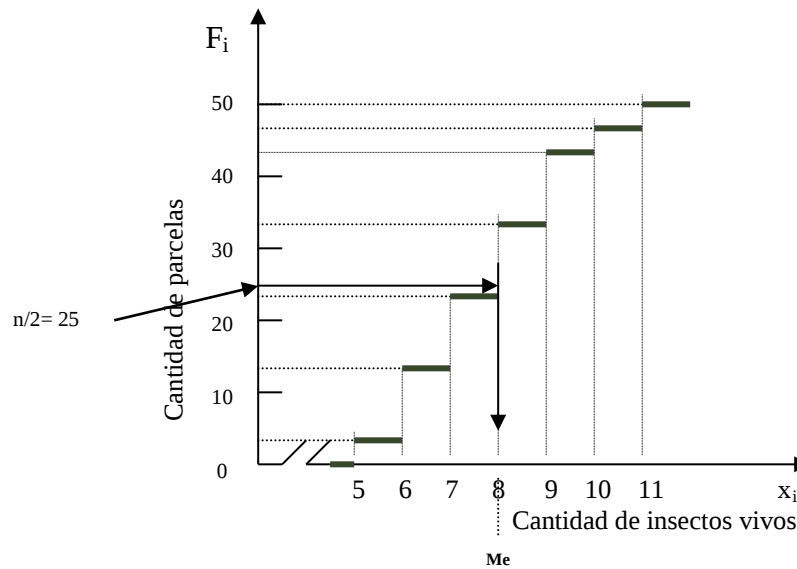
-En la columna de frecuencias acumuladas se busca la primera que supera a 25. Esa frecuencia es 34 y a 34 le corresponde el valor $x_4 = 8$, por lo que 8 es la mediana de este conjunto de datos.

$M_e = 8$ insectos vivos por parcela.

Interpretación: El 50% de las parcelas posee 8 insectos vivos o menos.

El 50% de las parcelas posee 8 insectos vivos o más.

⁶ Universidad Nacional de Córdoba. Curso de Posgrado. “ESTADÍSTICA APLICADA A LA INVESTIGACIÓN”

Método gráfico**Gráfico 11:** Gráfico de escalones de las frecuencias absolutas acumuladas

-Como se trata de una variable discreta se construye el gráfico de escalones “menor que” de las frecuencias absolutas acumuladas F_i (o de las relativas acumuladas).

-A una altura $n/2 = 25$ (o 0.5 si el gráfico de escalones corresponde a frecuencias relativas acumuladas), sobre el eje vertical de las frecuencias acumuladas se traza una paralela al eje de abscisas hasta cortar el gráfico de escalones.

-Por el punto de intersección, se baja una vertical que corta al eje de las abscisas en un punto que corresponde a 8 insectos. *Este punto es la mediana.*

Observación: Siempre que en un conjunto de datos, esté presente un valor extremo o atípico, es más conveniente utilizar la mediana como medida de tendencia central, en lugar de la media aritmética.

d. Responda:

- i) Por debajo de que cantidad de insectos vivos, se encuentra el 25% de las parcelas?.
- ii) Por debajo de que cantidad de insectos vivos, se encuentra el 75% de las parcelas?.

Las medidas que permiten responder a estas preguntas son los cuartiles.

Los cuartiles de una distribución, son tres valores de la variable Q_1 , Q_2 , y Q_3 que dividen al conjunto de datos ordenados en cuatro grupos iguales.

Tabla X: Cantidad de insectos vivos por parcela

	Cantidad de insectos x_i	Cantidad de parcelas f_i	Frec. Absol. Acumulada F_i ↓
	5	3	3
	6	9	12
Q_1 →	7	10	22
	8	12	34
Q_2 →	9	8	42
Q_3 →	10	5	47
	11	3	50
	Totales	50	

Para Q_1

-Se calculan las frecuencias absolutas acumuladas “menor que”.

-Se calcula $n / 4 = 50 / 4 = 12.5$

-En la columna de frecuencias acumuladas se busca la primera que supera a 12.5. Esa frecuencia es 22 y a 22 le corresponde el valor de $x_3 = 7$, por lo que 7 es el primer cuartil de este conjunto de datos.

$Q_1 = 7$ insectos vivos por parcela

Interpretación: El 25% de las parcelas posee 7 insectos vivos o menos y el 75% restante tiene 7 insectos vivos o más.

Para Q_2

El segundo cuartil coincide con la mediana, es decir $Q_2 = M_e = 8$ insectos vivos por parcela. La interpretación es igual a la de la mediana.

Para Q_3

-Se calculan las frecuencias absolutas acumuladas “menor que”.

-Se calcula $3n / 4 = 150 / 4 = 37.5$

-En la columna de frecuencias acumuladas se busca la primera que supera a 37.5. Esa frecuencia es 42 y a 42 le corresponde el valor de $x_5 = 9$, por lo que 9 es el tercer cuartil de este conjunto de datos.

$Q_3 = 9$ insectos vivos por parcela.

Interpretación: El 75% de las parcelas posee 9 insectos vivos o menos y el 25% restante tiene 9 insectos vivos o más.

Determinación gráfica de los cuartiles

El método gráfico para la determinación de los cuartiles es similar al visto para la mediana.

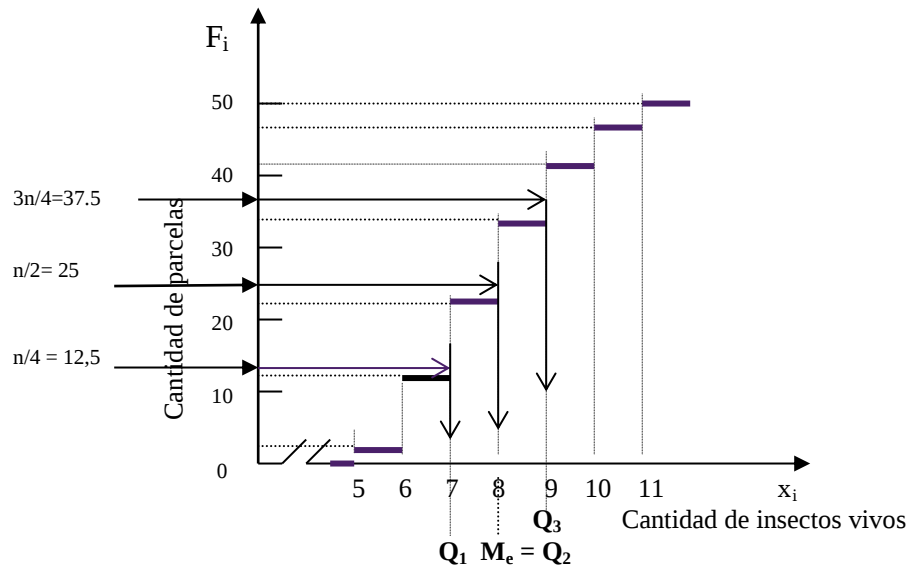


Gráfico 12: Gráfico de escalones

e) Calcule e interprete el rango y una medida de dispersión apropiada.

Rango se ha definido en la pág. 15.

En nuestro problema: rango = 11 insectos – 5 insectos = 6 insectos

Interpretación: la variabilidad o el recorrido total de la variable es de 6 insectos vivos por parcela.

Observación: La medida de dispersión a elegir depende de la medida de posición central que se ha calculado. Si se ha calculado la media aritmética como medida de posición, conviene utilizar la **desviación media** como medida de dispersión; en el caso de que la medida de posición utilizada es la mediana, conviene trabajar con la **desviación mediana**. Otra medida de dispersión ampliamente conocida es la **varianza**, que también considera las desviaciones con respecto a la media.

En nuestro caso se han calculado las tres medidas de posición central por lo que, y a modo de ejemplificar, calcularemos las distintas medidas de dispersión.

Cálculo de la varianza

La expresión para el cálculo de la varianza es:
$$S^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{n - 1}$$

Para hallar analíticamente la varianza agregaremos a la tabla de frecuencias otras columnas que faciliten el cálculo.

-En nuestro problema $\bar{x} = 7.8$ insectos vivos por parcela.

Tabla XI: Cantidad de insectos vivos por parcela

Cantidad de insectos x_i	Cantidad de parcelas f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
5	3	5 - 7.8 = -2.8	7.84	7.84 x 3 = 23.52
6	9	6 - 7.8 = -1.8	3.24	3.24 x 9 = 29.16
7	10	7 - 7.8 = -0.8	0.64	0.64 x 10 = 6.4
8	12	8 - 7.8 = 0.2	0.04	0.48
9	8	9 - 7.8 = 1.2	1.44	11.52
10	5	10 - 7.8 = 2.2	4.84	24.2
11	3	11 - 7.8 = 3.2	10.24	30.72
Totales	50			$\sum = 126$

$$S^2 = \frac{126}{49} = 2.57 \text{ insectos}^2$$

Dado que la variable queda elevada al cuadrado, entonces se emplea como medida de dispersión la raíz cuadrada de la varianza, la que se conoce con el nombre de **desviación estándar**.

$$S = \sqrt{S^2} = \sqrt{2.57} = 1.60 \cong 2 \text{ insectos vivos.}$$

Interpretación: La dispersión promedio de cómo varían las cantidades de insectos vivos por parcela mayores que la cantidad promedio y las cantidades inferiores al promedio es de 2 insectos vivos por parcela.

Otra expresión para hallar la varianza está dada por la fórmula de trabajo o expresión abreviada. Su utilización elimina el cálculo de los desvíos, por lo que el uso de la calculadora resulta mucho más práctico:

$$S^2 = \frac{n \sum_{i=1}^m x_i^2 f_i - \left(\sum_{i=1}^m x_i f_i \right)^2}{n(n-1)}$$

Disponemos los datos en el siguiente cuadro:

x_i	f_i	$x_i^2 f_i$	$x_i f_i$
5	3	25x3=75	15
6	9	36x9=324	54
7	10	490	70
8	12	768	96
9	8	648	72
10	5	500	50
11	3	363	33
Totales	50	3168	390

$$S^2 = \frac{50 \times 3168 - 152100}{50 \times 49} = 2.57, \quad S = 1.60$$

A modo de ejemplo calcularemos otras medidas de dispersión.

Desviación media

La expresión para el cálculo de la desviación media para datos agrupados, como es el caso

del problema planteado, está dada por:
$$DM = \frac{\sum_{i=1}^m |x_i - \bar{x}| f_i}{n}$$

Podemos ordenar los valores en una tabla de la siguiente manera:

Tabla XII : Cantidad de insectos vivos por parcela

Cantidad de insectos x_i	Cantidad de parcelas f_i	$x_i - \bar{x}$	$ x_i - \bar{x} $	$ x_i - \bar{x} f_i$
5	3	$5 - 7.8 = -2.8$	2.8	$2.8 \times 3 = 8.4$
6	9	$6 - 7.8 = -1.8$	1.8	$1.8 \times 9 = 16.2$
7	10	$7 - 7.8 = -0.8$	0.8	8
8	12	$8 - 7.8 = 0.2$	0.2	2.4
9	8	$9 - 7.8 = 1.2$	1.2	9.6
10	5	$10 - 7.8 = 2.2$	2.2	11
11	3	$11 - 7.8 = 3.2$	3.2	9.6
Totales	50			$\sum = 65.2$

$$DM = \frac{65.2}{50} = 1.304 \cong 1 \text{ insecto vivo por parcela.}$$

Interpretación: En promedio, los desvíos de las observaciones con respecto a la media es de un insecto vivo por parcela.

Desviación mediana

Para su cálculo debemos trabajar en forma similar al de la desviación media, pero considerando en lugar de la media, la mediana.

Entonces la expresión que nos permite su cálculo es:
$$DMe = \frac{\sum_{i=1}^m |x_i - Me| f_i}{n}.$$

Como la $Me = 8$ insectos vivos, entonces $DMe = 1.28 \cong 1$ insecto vivo por parcela.

Realice los cálculos para verificar el resultado.

f) Calcule e interprete el coeficiente de asimetría y el de curtosis.

Además de resumir los datos con las medidas de posición y dispersión calculadas, es necesario considerar también la forma de los datos.

En primer lugar consideraremos la simetría de la distribución

Resumiendo : $\bar{x} = 7.8$ insectos vivos por parcela.

$Me = Mo = 8$ insectos vivos por parcela.

Como $\bar{x} < Me = Mo$, $\Rightarrow$ podemos considerar que la distribución de los datos es casi *simétrica*. Observando el gráfico nº 10 correspondiente a la distribución de frecuencias, podemos comentar que no hay valores extremos en una dirección particular, de modo que los valores bajos y los valores altos tienden a equilibrarse o a balancearse entre sí.

En distribuciones de frecuencias unimodales, un coeficiente de asimetría está dado por:

$$As = \frac{\bar{x} - Mo}{S} \quad \text{con } -1 < As < 1.$$

En nuestro problema $As = \frac{7.8 - 8}{1.61} = -0.12$, tiende a cero. Por lo que la distribución es casi simétrica como ya lo habíamos anticipado. Presenta una leve asimetría negativa.

- Para el caso de que **As** resulte negativo, la distribución presenta asimetría negativa o se dice sesgada a la izquierda.
- Para el caso de que **As** sea igual a cero, la distribución es simétrica.
- Para el caso de **As** positivo, la distribución presenta asimetría positiva o se dice sesgada a la derecha.

Curtosis o “kurtosis” o apuntamiento

El apuntamiento de una distribución de frecuencias indica la mayor o menor altura del máximo central, con respecto a la altura de la curva normal con media y desviación típica correspondiente a la distribución que se estudia.

El grado de apuntamiento o curtosis está dado por el coeficiente de curtosis **K**.

$$\text{Un coeficiente está dado por: } K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \quad \text{con } 0 < K < 0.5$$

Los cuartiles Q_3 y Q_1 ya están calculados (pág. 30).

Necesitamos, entonces calcular P_{90} y P_{10} que son los percentiles 90 y 10. Su cálculo es similar al de los cuartiles.

Tabla XIII: Cantidad de insectos vivos por parcela

	Cantidad de insectos x_i	Cantidad de parcelas f_i	Frec. Absol. Acumulada F_i
	5	3	3
P_{10} →	6 ←	9	12
	7	10	22
	8	12	34
	9	8	42
P_{90} →	10 ←	5	47
	11	3	50
	Totales	50	

Cálculo de P_{10}

$$10 \frac{n}{100} = 10 \frac{50}{100} = 5$$

----El valor inmediato superior a 5 en la columna de frecuencias acumuladas es 12, y a 12 le corresponde $x_2 = 6$.

$P_{10} = 6$ insectos vivos por parcela.

Interpretación: El 10% de las parcelas posee 6 insectos vivos o menos y el 90% restante tiene 6 insectos vivos o más.

32

Cálculo de P_{90}

$$90 \frac{n}{100} = 90 \frac{50}{100} = 45$$

-El valor inmediato superior a 45 en la columna de frecuencias acumuladas es 47, y a 47 le corresponde $x_6 = 10$. Entonces $P_{90} = 10$ insectos vivos por parcela.

Interpretación

El 90% de las parcelas posee 10 insectos vivos o menos y el 10% restante tiene 10 insectos vivos o más.

Entonces $K = \frac{9-7}{2(10-6)} = \frac{2}{8} = 0.25$ en este caso decimos que la distribución es **mesocúrtica**

(tiene el apuntamiento de la curva normal).

Si K tiende a 0 la distribución es **platicúrtica** (más aplanada que la normal).

Si K tiende a 0.5 la distribución es **leptocúrtica** (más puntiaguda que la normal).

Otras medidas de tendencia central1) **La media geométrica**

Esta medida descriptiva se usa cuando se desea promediar tasas de cambios, proporciones, índices.

“La aplicación más útil de la media geométrica es para promediar tasas de cambio, porque el promedio de tasas de cambios sólo puede medirse correctamente por este método”.⁷

⁷ Chou, Y.L. Análisis estadístico.

Media geométrica para datos no agrupados

Si tenemos n observaciones $x_1, x_2, \dots, x_n$ positivos, la media geométrica está dada por:

$$G = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 x_2 \dots x_n}$$

Media geométrica para datos agrupados

Si los datos están agrupados en m clases, la media geométrica está dada por:

$$G = \sqrt[n]{\prod_{i=1}^m x_i^{f_i}} = \sqrt[n]{x_1^{f_1} x_2^{f_2} \dots x_m^{f_m}}$$

Problema 5: Crecimiento de una cuenta de ahorros

Suponga que se depositaron \$100 inicialmente y que se acumulan los intereses a tasas variables de 7, 8, 10, 12 y 18% anual, durante 5 años. Halle el factor de crecimiento promedio anual.

Solución

El factor de crecimiento es la cantidad por la cual se deben multiplicar los ahorros al comienzo del año para obtener los ahorros al final del año.**

Disponemos los datos en la siguiente tabla:

Tabla XIV: Factor de crecimiento

Año	Tasa de interés %	*Factor de crecimiento x_i
1	7	1.07
2	8	1.08
3	10	1.10
4	12	1.12
5	18	1.18

*Para el cálculo del factor de crecimiento se trabaja con la expresión: $1 + \frac{\text{tasa}}{100}$

Así para 7%, tenemos $1 + \frac{7}{100} = 1.07$.

Es decir que \$1, más el interés que produce a una tasa del 7% anual al cabo de un año asciende a \$1.07. Por tratarse de tasas de interés, para obtener el promedio de crecimiento pedido, usaremos la media geométrica.

$$G = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[5]{1.07 \times 1.08 \times 1.10 \times 1.12 \times 1.18} = \sqrt[5]{1.679965} = 1.1093 \text{ veces cada año.}$$

Respuesta: 1.1093 veces cada año es el factor de crecimiento promedio.

$$1.1093 = 1 + 0.1093 = 1 + \frac{10.93}{100} \text{ que corresponde a una tasa de interés promedio del 10.93\% anual. } ($$

0.1093 es la tasa de crecimiento).

Como para ilustrar la aplicación de la media geométrica, hagamos una comparación de los montos que se obtienen con las tasas variables de interés dadas y con la tasa de interés constante igual a la tasa promedio.

Tabla XV: Crecimiento de un depósito de \$100 con tasas variables

Año	Tasa de interés %	*Factor de crecimiento x_i	**Ahorros al final del año
1	7	1.07	\$100 x 1.07 = \$107
2	8	1.08	107 x 1.08 = 115.56
3	10	1.10	115.56 x 1.10 = 127.12
4	12	1.12	127.12 x 1.12 = 142.37
5	18	1.18	142.37 x 1.18 = \$168

Esta tabla muestra que el crecimiento real es \$168.

Si el banco pagara intereses a una tasa constante de 10.93% anual al cabo de 5 años, tendríamos:

Tabla XVI: Crecimiento de un depósito de \$100 a tasa constante.

Año	Tasa de interés %	*Factor de crecimiento x_i	**Ahorros al final del año
1	10.93	1.1093	\$100 x 1.1093 = 110.93
2	10.93	1.1093	110.93 x 1.1093 = 123.054649
3	10.93	1.1093	123.054649 x 1.1093 = 136.5045221
4	10.93	1.1093	136.5045221 x 1.1093 = 151.4244663
5	10.93	1.1093	151.4244663 x 1.1093 = 167.9751604 = \$168

El crecimiento real es igual al mostrado en la tabla XV, esto nos dice que el promedio calculado en base a la media geométrica es el correcto.

Este monto real también se puede encontrar utilizando la fórmula de monto a interés compuesto:

$$C_n = C_o(1 + i)^n$$

C_n : cantidad al final del periodo

C_o : cantidad al inicio del periodo

i : tasa de interés compuesto o tasa de cambio

n : número de periodos

Apliquemos a nuestro caso

$$C_0 : \$100$$

$$i : 0.1093. \text{ La tasa promedio es } 10.93\%, \text{ la tasa de cambio} = 10.93/100$$

$$n : 5$$

$$C_n = 100(1+0.1093)^5 = \$168$$

2) Media armónica.

“[...], la media armónica es el recíproco de la media aritmética de los recíprocos de las observaciones”.⁸

Dado un conjunto de n observaciones: $x_1, x_2, \dots, x_n$.

Para datos no agrupados o serie simple

$$H = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{1}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Para datos agrupados en m clases

$$H = \frac{1}{\frac{1}{x'_1} f_1 + \frac{1}{x'_2} f_2 + \dots + \frac{1}{x'_m} f_m} = \frac{1}{\sum_{i=1}^m \frac{1}{x'_i} f_i} = \frac{n}{\sum_{i=1}^m \frac{1}{x'_i} f_i} \quad \text{con } i = 1, 2, \dots, m$$

La media armónica se aplica cuando se tiene términos para cuyos recíprocos se quiere calcular la media; también cuando se presenta una relación inversa entre las variables implícitas.

Problema 6

a) Suponga que ha gastado \$1 por 3 docenas de naranjas en un negocio, \$1 por 4 docenas de naranjas en otro negocio, y \$1 por 5 docenas de naranjas en otro. Determine el precio promedio por docena de naranjas.

El precio por docena en cada negocio está dado por:

$$x_1 = \frac{1}{3} \text{ es el precio por docena en el primer negocio.}$$

⁸ Chou, Y.L. Análisis estadístico.

$x_2 = \frac{1}{4}$ es el precio por docena en el segundo negocio.

$x_3 = \frac{1}{5}$ es el precio por docena en el tercer negocio.

Lo que queremos calcular es la media de los recíprocos de los números 3, 4, 5. La media armónica es el promedio que mejor se ajusta.

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{3}{\frac{1}{3} + \frac{1}{4} + \frac{1}{5}} = \frac{3}{\frac{47}{60}} = 0.25$$

36

Respuesta: El precio promedio por docena es de \$0.25

Comprobación

En total se compraron 12 docenas.

Pagando a razón de \$0.25 por docena $\rightarrow 12 \times \$0.25 = \3

Si se hubiera tomado la media aritmética tendríamos:

$$\bar{x} = \frac{\frac{1}{3} + \frac{1}{4} + \frac{1}{5}}{3} = \frac{\frac{47}{60}}{3} = \frac{47}{180} = 0.261. \text{ El precio promedio por docena es } \$0.261$$

Comprobación

Pagando a razón de \$0.261 por docena $\rightarrow 12 \times \$0.261 = \3.132 , más de lo que realmente se gastó, entonces la utilización de la media aritmética en este caso es incorrecta.

b) Cinco mecanógrafas tienen las siguientes velocidades: 40, 36, 28, 40 y 35 palabras por minuto. Si cada una de ellas escribe un mismo texto, calcule la velocidad media.

Solución

Las variables consideradas en este enunciado son: **velocidad** (palabras por minuto) y **tiempo** para escribir un mismo texto.

Las variables velocidad y el tiempo están en relación inversa, a mayor velocidad menor tiempo empleado para escribir el texto. Entonces el promedio está dado por la media armónica.

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{5}{\frac{1}{40} + \frac{1}{36} + \frac{1}{28} + \frac{1}{40} + \frac{1}{35}} = \frac{5}{\frac{358}{2520}} = 35.196 \quad \text{Palabras por minuto.}$$

Problema 7

En un experimento industrial donde se desea estudiar la capacidad de una máquina para producir piezas dentro de especificaciones, se han medido 100 tornillos (registrándose su longitud en mm.), los resultados se presentan en la siguiente tabla de frecuencias:

Longitud de los tornillos (Medidos en mm.)	Cantidad de tornillos
6-7	11
7-8	9
8-9	14
9-10	11
10-11	22
11-12	14
12-13	7
13-14	5
14-15	4
15-16	3

37

- Calcule la longitud promedio de los tornillos fabricados por la máquina.
- Determine la mediana y la moda de la distribución de los datos analítica y gráficamente. Interprete las mismas.
- Calcule el rango, la desviación estándar. Interprete las mismas.
- Calcule el recorrido intercuartílico e interprételo.
- Calcule los coeficientes de asimetría y de curtosis.

Solución

Observación: En todo problema, primero se debe definir la variable de estudio, clasificarla y reconocer la escala de medición.

Variable en estudio

X: Longitud de los tornillos producidos por una máquina

Clasificación de la variable

Se trata de una variable **numérica continua**.

La escala de medición es de razón.

- Calcule la longitud promedio de los tornillos fabricados por la máquina.

La medida que nos da la respuesta es la media aritmética. Por tratarse de una **variable numérica continua**, y como los datos están agrupados en intervalos de clase, la expresión para calcular esta medida es similar a la utilizada en el problema anterior y es la siguiente:

$$\bar{x} = \frac{\sum_{i=1}^m x'_i f_i}{n}, \text{ en la que } m \text{ es la cantidad intervalos y } x'_i \text{ es la marca de clase de los intervalos.}$$

Para su cálculo se organizan los datos en una tabla de frecuencias absolutas, en la que se agrega una columna con las marcas de clases y otra con los productos de cada marca de clase por la correspondiente frecuencia absoluta

Tabla XVII: Longitud de los tornillos producidos por una máquina

Intervalos de clase	Marca de clase x'_i	Frecuencia f_i	$x'_i f_i$
6-7	6.5	11	71.5
7-8	7.5	9	67.5
8-9	8.5	14	119.0
9-10	9.5	11	104.5
10-11	10.5	22	231.0
11-12	11.5	14	161.0
12-13	12.5	7	87.5
13-14	13.5	5	67.5
14-15	14.5	4	58.0
15-16	15.5	3	46.5
Total		100	1014.0

$$\bar{x} = \frac{1014.4}{100} = 10.144 \text{ mm.}$$

Interpretación: En promedio el proceso fabrica tornillos de 10.14 mm. de longitud.

b) Determine la moda y la mediana de la distribución de los datos analítica y gráficamente. Interprete las mismas.

Cálculo de la moda para datos agrupados en intervalos

1.-Método gráfico

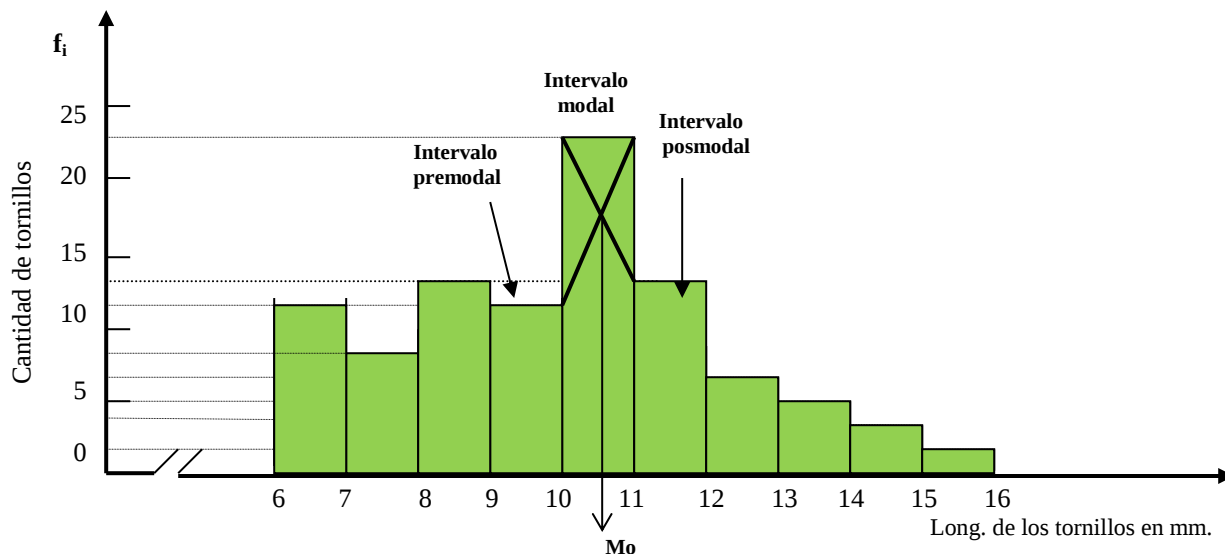


Gráfico 13: Histograma de frecuencias absolutas de la variable longitud de los tornillos fabricados por una máquina.

- Se construye el histograma de frecuencias absolutas.
- Como todos los intervalos tienen el mismo ancho, el rectángulo de mayor altura corresponde al intervalo modal.
- Se unen los vértices superiores del intervalo modal con los vértices superiores del intervalo premodal y del intervalo posmodal.
- Por el punto de intersección de los segmentos considerados, se traza una vertical que corta al eje de las abscisas en un punto. Este punto corresponde a la moda.

2.- Método analítico

Para el cálculo se utiliza la siguiente fórmula: $Mo = L_i + \frac{d_1}{d_1 + d_2} a_i$

Consideraremos la siguiente tabla:

Tabla XVIII: Longitud de los tornillos producidos por una máquina.

Intervalos de clase	Frecuencia absoluta f_i	
6-7	11	
7-8	9	
8-9	14	
9-10	11	→ Frecuencia absoluta premodal
10-11	22	→ Frecuencia absoluta modal
11-12	14	→ Frecuencia absoluta posmodal
12-13	7	
13-14	5	
14-15	4	
15-16	3	
Total	100	

Observación:

Se puede trabajar con las frecuencias relativas, en lugar de las frecuencias absolutas.

- La mayor frecuencia absoluta es 22. Esta frecuencia es la que corresponde al intervalo 10-11.

Entonces 10-11 es el *intervalo modal*.

- L_i es el límite inferior del intervalo modal. En nuestro caso $L_i = 10$.
- d_1 es la diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia absoluta del intervalo premodal.

$$d_1 = 22 - 11 = 11$$

- d_2 es la diferencia entre la frecuencia absoluta del intervalo modal y la frecuencia absoluta del intervalo posmodal.

$$d_2 = 22 - 14 = 8$$

- a_i es la amplitud del intervalo modal. $a_i = 1$

Reemplazando estos valores en la fórmula tenemos:

$$Mo = 10 + \frac{11}{11+8} \cdot 1 = 10 + \frac{11}{19} = 10.58 \text{ mm.}$$

Interpretación: La longitud de tornillos que se presenta con mayor frecuencia en el proceso de fabricación es de 10.58 mm.

Cálculo de la mediana para el caso de datos agrupados en intervalos

1.- Método analítico

Para el cálculo tenemos en cuenta que se trata de una **variable continua**, y que los **datos están agrupados** en intervalos de clase. La expresión que se utiliza es la siguiente:

$$Me = L_i + \frac{\frac{n}{2} - F_{(i-1)}}{f_j} \cdot a_j$$

Tabla XIX.-Tabla de frecuencias de una variable continua para una muestra de tamaño n.

Intervalos de clase $L_i - L_s$	Frec. Abs. f_i	Frec. Abs. Acumul. Menor que $F_i \downarrow$
$L_0 - L_1$	f_1	F_1
$L_1 - L_2$	f_2	F_2
.....		
$L_{j-2} - L_{j-1}$	f_{j-1}	F_{j-1}
$L_{j-1} - L_j$	f_j	F_j
.....		
$L_{m-1} - L_m$	f_m	$F_m = n$
Totales	n	

- Se consideran las frecuencias acumuladas “menor que” $F_i \downarrow$.
(se puede considerar las frecuencias relativas acumuladas)

- Se calcula $n/2$ y se ubica la menor frecuencia acumulada F_j que supere a este valor $n/2$.

- A F_j le corresponde el intervalo medial $L_{j-1} - L_j$ y la frecuencia absoluta f_j .

L_{j-1} es el límite inferior del intervalo medial y L_j el límite superior del intervalo.

Con a_i se designa la amplitud de los intervalos, entonces a_j , es

la amplitud del intervalo medial.

Volviendo a nuestro problema y al igual que en el problema 4 (Tabla IX), consideraremos la distribución de las frecuencias absolutas acumuladas “menor que”.

Tabla XX.- Longitud de los tornillos fabricados por una máquina.

Intervalos De clase	Frecuencia f_i	Frec. absol. acumuladas F_i
6-7	11	11
7-8	9	20
8-9	14	34
9-10	11	45
<u>10-11</u>	<u>22</u>	<u>67</u>
11-12	14	81
12-13	7	88
13-14	5	93
14-15	4	97
15-16	3	100
Total	100	

-Se calculan las frecuencias acumuladas “menos que” F_i ↓

- Se calcula $n / 2 = 100 / 2 = 50$

- En la columna F_i , se ubica la primera frecuencia que supera a $n / 2$, en nuestro caso que supere a 50. Esa frecuencia es $F_j = F_5 = 67$,

- A 67 la corresponde el intervalo 10-11, denominado intervalo medial, 10 es el límite inferior del intervalo medial y $f_j = f_5 = 22$ la frecuencia absoluta del intervalo

medial.

- En nuestro problema todos los intervalos son iguales, entonces $a_5 = a = 1$ es la amplitud del intervalo medial. (con a_5 designamos la amplitud del quinto intervalo)

- La $F_{(j-1)}$ es $F_4 = 45$

Reemplazando los valores obtenidos en la expresión $Me = L_{j-1} + \frac{\frac{n}{2} - F_{(j-1)}}{f_j} \cdot a_j$

Tenemos que: $Me = 10 + \frac{50 - 45}{22} \cdot 1 = 10 + \frac{5}{22} = 10.23\text{mm}$

Interpretación

- El 50% de los tornillos producidos por una máquina tienen una longitud de 10.23 mm. o menos.
- El 50% de los tornillos producidos por una máquina tienen una longitud de 10.23 mm. o más.

2.-Método gráfico

Para el cálculo gráfico de la mediana se procede de la siguiente manera.

-Como se trata de una **variable continua** se construye la **ojiva “menor que”** de las frecuencias absolutas acumuladas F_i (o de las relativas acumuladas R_i).

-A una altura $n/2 = 50$ (o 0.5 si la ojiva corresponde a las frecuencias relativas acumuladas), sobre el eje vertical de las frecuencias acumuladas se traza una paralela al eje de abscisas hasta cortar la ojiva.

-Por el punto anterior, se baja una vertical que corta al eje de las abscisas en un punto que corresponde aproximadamente a 10.23 mm. *Este punto es la mediana.*

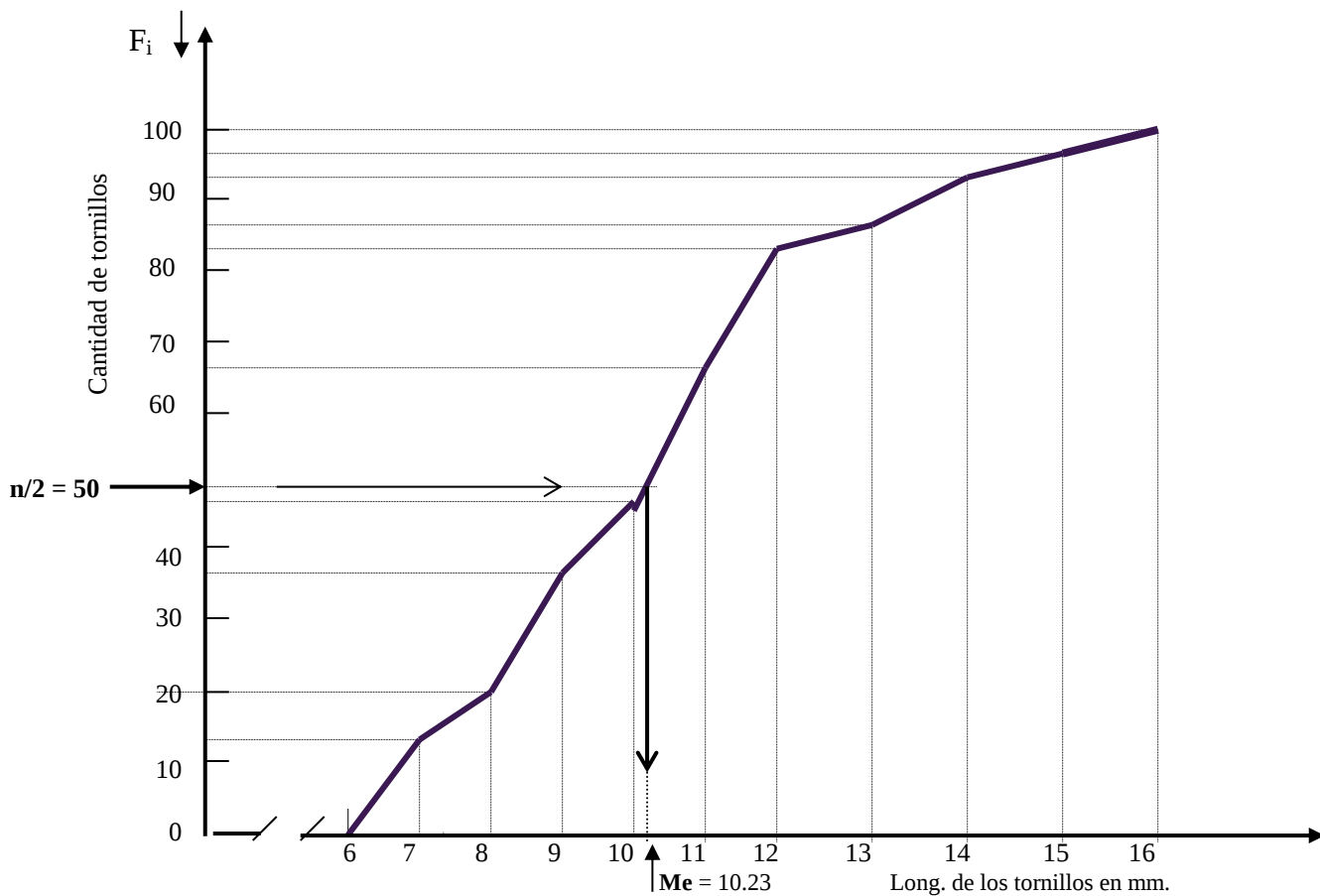


Gráfico 14: Ojiva de las frecuencias absolutas acumuladas de las longitudes de los tornillos producidos por una máquina.

c) Calcule el rango y la desviación estándar. Interprete las mismas.

Como los datos del problema están dados por el agrupamiento en 10 intervalos, el rango se calcula haciendo la diferencia entre la última marca de clase y la primera:

$$x'_{10} - x'_1 = 15.5\text{mm.} - 6.5\text{mm.} = 9\text{mm.}$$

Interpretación: La variabilidad total de la longitud de los tornillos fabricados es de 9mm.

Recordemos que la desviación estándar es la raíz cuadrada de la varianza, entonces necesitamos calcular primero la varianza.

Cálculo de la varianza

El cálculo de la varianza para datos agrupados en intervalos de clase, está dado por:

$$S^2 = \frac{\sum_{i=1}^m (x'_i - \bar{x})^2 f_i}{n - 1} \quad \text{en la que } x'_i \text{ es la marca de clase de los intervalos.}$$

Dispondremos los datos en una tabla, con el objeto de facilitar el cálculo. Recordemos que

$$\bar{x} = 10.144\text{mm}$$

Tabla XXI.- Longitud de tornillos producidos por una máquina

Intervalos de clase	Marca de clase x'_i	Frecuencia absoluta f_i	$x'_i - \bar{x}$	$(x'_i - \bar{x})^2$	$(x'_i - \bar{x})^2 f_i$
6-7	6.5	11	6.5 - 10.14 = -3.64	$(-3.64)^2 = 13.25$	$13.25 \times 11 = 145.75$
7-8	7.5	9	7.5 - 10.14 = -2.64	$(-2.64)^2 = 6.97$	$6.97 \times 9 = 62.73$
8-9	8.5	14	-1.64	2.69	37.65
9-10	9.5	11	-0.64	0.41	4.51
10-11	10.5	22	0.36	0.13	2.85
11-12	11.5	14	1.36	1.85	25.89
12-13	12.5	7	2.36	5.57	38.99
13-14	13.5	5	3.36	11.29	56.45
14-15	14.5	4	4.36	19.01	76.04
15-16	15.5	3	5.36	28.73	86.19
Total		100			537.05

$$S^2 = 537.05 / 99 = 5.4247 \text{ mm}^2$$

$$\text{La desviación estándar } S = \sqrt{5.4247 \text{ mm}^2} = 2.3291\text{mm}$$

Interpretación: La dispersión promedio de cómo varían las longitudes de los tornillos mayores que la longitud promedio y de las longitudes inferiores a la longitud promedio es de 2.3291mm.

Actividad: Calcule la varianza con la fórmula de trabajo.

d) Calcule el recorrido intercuartílico e interprételo.

El recorrido intercuartílico

Esta medida considera la extensión del 50% central de los datos y por lo tanto no sufre la influencia de los valores extremos. Se define como la diferencia entre el tercer cuartil y primer cuartil.

$$\text{El recorrido intercuartílico} \quad \text{RI} = Q_3 - Q_1$$

El procedimiento para el cálculo de los cuartiles es similar al seguido para el cálculo de la mediana.

Tabla XXII.-Longitud de los tornillos fabricados por una máquina

Intervalos De clase	Frecuencia f_i	Frec. absol. acumuladas F_i ↓
6-7	11	11
7-8	9	20
8-9	14	34
9-10	11	45
10-11	22	67
11-12	14	81
12-13	7	88
13-14	5	93
14-15	4	97
15-16	3	100
Total	100	

- Se calculan las frecuencias acumuladas “menos que” F_i ↓

Cálculo del primer cuartil

La fórmula para el primer cuartil está dada por:

$$Q_1 = L_{j-1} + \frac{\frac{n}{4} - F_{(j-1)}}{f_j} \cdot a_j$$

- Se calcula $\frac{n}{4} = \frac{100}{4} = 25$

- En la columna F_i , se ubica la primera frecuencia que supera a $n / 4$, en nuestro caso que supera a 25. Esa frecuencia es $F_3 = 34$.

- A 34 le corresponde el intervalo 8-9; 8 es el límite inferior del intervalo y $f_j = f_3 = 14$ la frecuencia absoluta del intervalo

- En nuestro problema todos los intervalos son iguales, entonces $a_3 = a = 1$, es la amplitud del intervalo.

- La $F_{(j-i)}$ es $F_2 = 20$

Reemplazando los valores obtenidos en la fórmula dada, tenemos que:

$$Q_1 = 8 + \frac{25 - 20}{14} \cdot 1 = 8 + \frac{5}{14} = 8.36\text{mm}$$

Cálculo del tercer cuartil

La fórmula para el tercer cuartil está dada por: $Q_3 = L_{j-1} + \frac{\frac{3n}{4} - F_{(j-1)}}{f_j} \cdot a_j$

- Se calcula $\frac{3n}{4} = \frac{300}{4} = 75$

- En la columna F_i , se ubica la primera frecuencia que supera a $3n / 4$, en nuestro caso que supera a 75. Esa frecuencia es $F_6 = 81$

- A 81 le corresponde el intervalo 11-12; 11 es el límite inferior del intervalo y $f_j = f_6 = 14$ la frecuencia absoluta del intervalo.

- En nuestro problema todos los intervalos son iguales, entonces $a_6 = a = 1$, es la amplitud del intervalo.

- La $F_{(j-i)}$ es $F_5 = 67$

Reemplazando los valores obtenidos en la fórmula dada, tenemos que:

$$Q_3 = 11 + \frac{75-67}{14} \cdot 1 = 11 + \frac{8}{14} = 11.57\text{mm.}$$

$$RI = 11.57\text{mm} - 8.36\text{mm} = 3.21\text{mm}$$

Interpretación: La variabilidad del 50% de las longitudes centrales de la variable, es de 3.21mm.

45

e) Calcule el coeficiente de asimetría y el de curtosis.

Asimetría

$$\begin{array}{ccccc} \bar{x} & < & \mathbf{Me} & < & \mathbf{Mo} \\ \downarrow & & \downarrow & & \downarrow \\ 10.14\text{mm} & & 10.23\text{ mm} & & 10.58\text{mm} \end{array}$$

Como la media es menor que la mediana y la moda está a la derecha, podemos decir que la distribución de los datos tiene asimetría negativa o que es sesgada hacia la izquierda.

Veamos el cálculo del coeficiente:

$$As = \frac{\bar{x} - Mo}{S} = \frac{10.144\text{mm} - 10.58\text{mm}}{2.3291\text{mm}} = -0.19\text{mm}$$

Podemos concluir diciendo que la distribución tiene asimetría negativa pequeña.

Curtosis

Para el cálculo necesitamos conocer los percentiles 10 y 90.

En forma análoga al procedimiento seguido para el cálculo de cuartiles, hallaremos los percentiles.

$$P_{10} = 6 + \frac{10-0}{11}1 = 6.91\text{mm} \quad \text{y} \quad P_{90} = 13 + \frac{90-88}{5}1 = 13 + \frac{2}{5} = 13.4\text{mm}$$

Reemplazando en la fórmula para el cálculo del coeficiente de curtosis, tenemos:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{11.57 - 8.36}{2(13.4\text{mm} - 6.91\text{mm})} = \frac{3.21\text{mm}}{12.98\text{mm}} = 0.25$$

La distribución es mesocúrtica. En este caso la moda es similar a la altura de la normal, aunque la distribución es asimétrica.

Problema 8

Las tres distribuciones siguientes muestran las ventas (en millones de dólares), de las 50 fábricas más grandes de un país en un año.

DISTRIBUCIÓN A		DISTRIBUCIÓN B		DISTRIBUCIÓN C	
Ventas	Cantidad de fábricas	Ventas	Cantidad de fábricas	Ventas	Cantidad de fábricas
0-10	14	0-10	14	0-5	8
10-20	18	10-20	18	5-15	11
20-30	6	20-30	6	15-30	10
30-40	2	30-40	2	30-35	5
40-50	1	40 y más	10	35-65	4
50-60	1			65-80	3
60-70	2			80-95	5
70-80	1			95-110	4
80-90	1				
90-100	2				
100-110	2				

- Defina la variable en estudio y clasifíquela.
- Calcule y grafique las frecuencias acumuladas de la forma “menor que” para la distribución B.
- Grafique las frecuencias absolutas para la distribución C.
- Con respecto a cada distribución, decida si es posible determinar la media y la mediana. En caso afirmativo, calcúlelas e interpréte las.
- Calcule el coeficiente de asimetría cuando sea posible.
- Compare la dispersión de las distribuciones A y C.
-

Solución

- La variable en estudio es X: Venta en millones de dólares de las 50 fábricas más grandes del país. La variable es numérica continua. La escala de medición es de razón.

Algunas veces al construir la distribución de frecuencias, puede ocurrir que los datos se agrupen en **intervalos con anchos desiguales**. También puede ocurrir que cuando se realicen observaciones extremas y por no seleccionar tamaños variables, se consideren clases o **intervalos con extremos abiertos**. Son los casos de la distribución C y la distribución B respectivamente. Estos casos presentan problemas en la realización de gráficas y de cálculos descriptivos, por lo que se deberá tener cautela en su tratamiento.

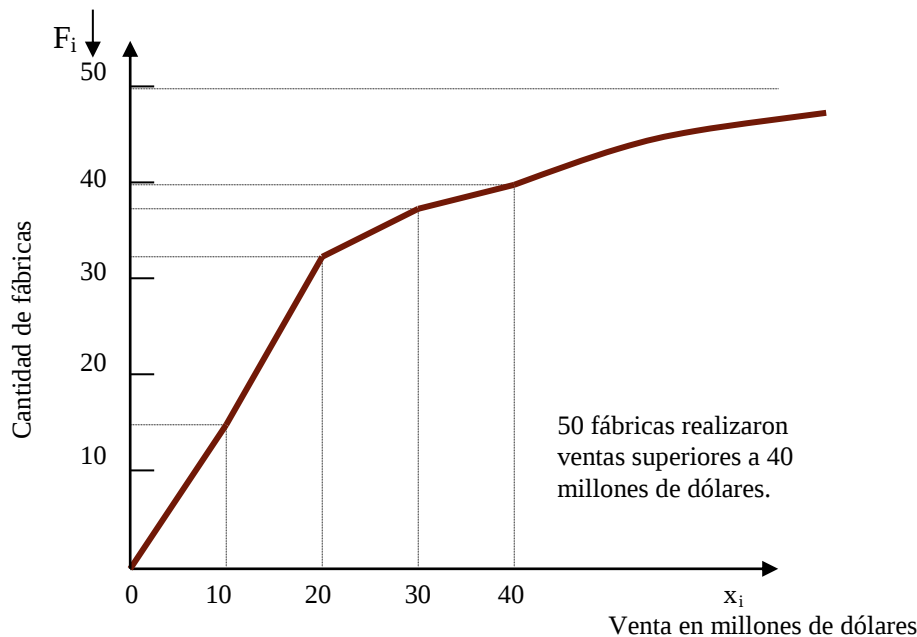
- Calcule y grafique las frecuencias acumuladas de la forma “menor que” para la distribución B.

Tabla XXII: Ventas en millones de dólares

DISTRIBUCIÓN B		
Ventas Intervalos	Cantidad de Fábricas f_i	Frecuencias Absolutas Acum. F_i ↓
0-10	14	14
10-20	18	32
20-30	6	38
30-40	2	40
40 y más	10	50

Como se trata de una variable continua, la gráfica correspondiente es una **ojiva**.

La distribución B se particulariza por poseer el último intervalo con un extremo abierto. Para la construcción de la ojiva en el último intervalo se hace referencia a la situación.

**Gráfico15:** Ojiva de las frecuencias absolutas acumuladas “menor que”.

c) Grafique las frecuencias absolutas para la distribución C.

Para el caso de esta distribución, que se caracteriza por poseer intervalos con amplitud distinta, se deberá compensar los anchos variables observando “que la altura de cada barra rectangular está representada sobre una base de **frecuencia por intervalo estándar** o una base de **porcentaje por intervalo estándar**, según la gráfica construida va a ser un histograma de frecuencia o un histograma de porcentaje”.⁹

Como los intervalos son desiguales, para realizar el histograma correspondiente, primero estandarizamos la amplitud, tomando como unidad : $a = 5$. Los cálculos se muestran en la tabla siguiente:

⁹ Berennson, M:L:-Levine, D.M. Estadística para Administración y Economía. Conceptos y Aplicaciones.

Tabla XXIII: Ventas en millones de dólares

DISTRIBUCIÓN C				
Ventas x_i	Cantidad de Fábricas f_i	Amplitud del Intervalo a_i	Amplitud estandarizada $a=5=\text{unidad}$ $a'_i = a_i / a$	Frecuencia por intervalo estandarizado $f'_i = f_i / a'_i$
0-5	8	5	1	8
5-15	11	10	2	5.5
15-30	10	15	3	3.33
30-35	5	5	1	5
35-65	4	30	6	0.66
65-80	3	15	3	1
80-95	5	15	3	1.66
95-110	4	15	3	1.33

Observación: La estandarización de la amplitud se realiza solamente a los efectos de realizar el gráfico correspondiente.

Para graficar se consideran en el eje de las abscisas los intervalos, y en el eje de las ordenadas las frecuencias por cada intervalo estandarizado.

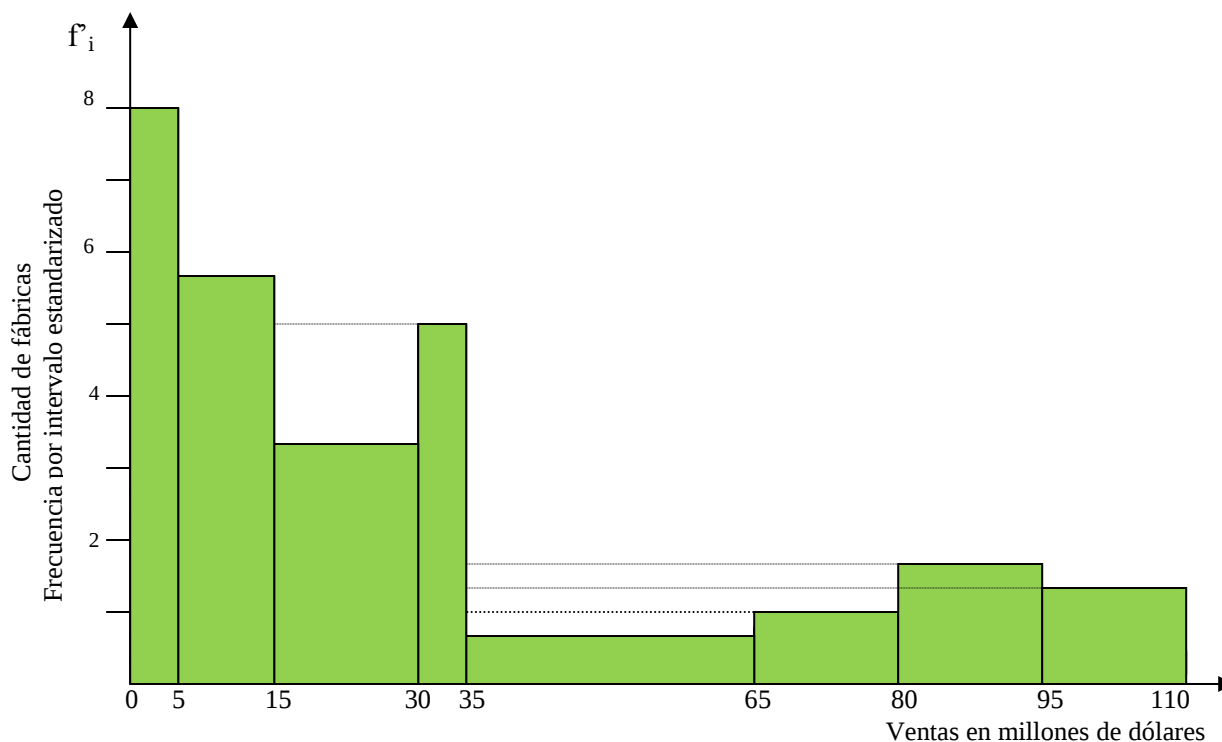


Gráfico 16: Histograma de las frecuencias absolutas estandarizadas de las ventas de las 50 fábricas más grandes del país.

d) Con respecto a cada distribución, decida si es posible determinar la media y la mediana. En caso afirmativo, calcúlelas e interprételas.

Distribución A: En esta distribución es posible el cálculo de las dos medidas y el cálculo es similar al visto en el problema 7.

DISTRIBUCIÓN A				
Ventas x_i	Marca de clase x'_i	Cantidad de Fábricas f_i	$x'_i \cdot f_i$	Frec. Absoluta acumulada F_i
0-10	5	14	70	14
10-20	15	18	270	32
20-30	25	6	150	38
30-40	35	2	70	40
40-50	45	1	45	41
50-60	55	1	55	42
60-70	65	2	130	44
70-80	75	1	75	45
80-90	85	1	85	46
90-100	95	2	190	48
100-110	105	2	210	50
Σ		50	1350	

Para la media aritmética tenemos: $\bar{x} = \frac{1350}{50} = 27$

Interpretación: La venta promedio de las 50 fábricas más grandes del país ascienden a 27.000.000 de dólares.

Observación: La distribución de las ventas es asimétrica, por lo que la media no es la medida de posición más adecuada para describir a estos datos.

Para la mediana tenemos:

$$n / 2 = 25$$

La menor frecuencia acumulada que supera a 25 es 32.

A 32 la corresponde el intervalo 10-20. La frecuencia absoluta de este intervalo es 18. La amplitud del intervalo es 10.

Reemplazando en la fórmula correspondiente obtenemos:

$$Me = 10 + \frac{25 - 14}{18} \cdot 10 = 16.11$$

Interpretación:

-El 50% de las fábricas produce ventas de u\$s 16.110.000 o menos.

-El 50% de las fábricas produce ventas de u\$s 16.110.000 o más.

Distribución B

Esta distribución se caracteriza por tener el último intervalo abierto, por este motivo no es posible calcular la media aritmética ya que no podemos contar con la marca de clase para este intervalo.

El cálculo de la mediana es posible y se lo realiza siguiendo el procedimiento conocido.

Realice el cálculo para hallar la mediana y verifique que es igual u\$s 16.110.000

Distribución C

Para esta distribución, que se caracteriza por tener intervalos con anchos desiguales, es posible calcular la media y la mediana siguiendo el procedimiento conocido y prestando atención al ancho del intervalo.

DISTRIBUCIÓN C					
Ventas x_i	Marca de clase x'_i	Cantidad de Fábricas f_i	$x_i f_i$	Amplitud del intervalo a_i	Frec. absolutas acumuladas $F_i \downarrow$
0-5	2.5	8	20	5	8
5-15	10	11	110	10	19
<u>15-30</u>	<u>22.5</u>	<u>10</u>	<u>225</u>	<u>15</u>	<u>29</u>
30-35	32.5	5	162.5	5	34
35-65	50	4	200	30	38
65-80	72.5	3	217.5	15	41
80-95	87.5	5	437.5	15	46
95-110	102.5	4	410	15	50

La media aritmética

$$\bar{x} = \frac{1782.5}{50} = 35.65$$

Interpretación: Las ventas de las 50 fábricas más grandes del país en promedio asciende a u\$s 35.650.000

La distribución C también es asimétrica, por lo que la media no es la medida de posición más representativa de los datos.

Cálculo de la mediana

$n/2 = 25$, la menor frecuencia absoluta acumulada que la supera es 29.

A $29 \rightarrow [15, 30)$. Este intervalo tiene frecuencia absoluta 10 y la amplitud es de 15.

Entonces la mediana es:

$$Me = 15 + \frac{25 - 19}{10} \cdot 15 = 24$$

$$Me = \text{u\$s } 24.000.000$$

e) Calcule el coeficiente de asimetría cuando sea posible.

$$As = \frac{\bar{x} - Mo}{S} \quad \text{Pero dijimos que no es posible calcular la media de la distribución B. Por lo tanto}$$

no es posible calcular el As para esta distribución. Sí es posible calcularlos para las distribuciones A y C.

El procedimiento es el visto en otros problemas.

51

f) Comparare la dispersión de las distribuciones A y C.

Para comparar la dispersión de las distribuciones A y C, debemos calcular el coeficiente de variación para cada distribución.

Otra medida de variabilidad

El **coeficiente de variación** o de variabilidad es una medida de dispersión relativa. Se expresa como porcentaje. El CV, mide la dispersión en términos de la media.

$$\text{En símbolos: } CV = \frac{S}{\bar{x}} 100$$

“El coeficiente de variación se utiliza para comparar la homogeneidad de dos series de datos aun cuando estén expresadas en distintas unidades de medida”.¹⁰

A medida que el coeficiente de variación disminuye, se observa una mayor homogeneidad en los datos, o lo que es lo mismo, los datos están más concentrados alrededor del promedio.

En el cálculo del coeficiente de variación interviene la desviación estándar que, por lo expresado con respecto a la media, tampoco resulta ser la medida de dispersión más adecuada para describir la variabilidad de las distribuciones A y C, pero a modo de ejemplo utilizaremos este coeficiente para comparar la dispersión de las dos distribuciones.

Cálculo de CV para la distribución A

Vamos a calcular S siguiendo el procedimiento conocido. La media aritmética para esta distribución es 27 millones de dólares. Para el cálculo de la desviación estándar, organizaremos los datos en la siguiente tabla:

¹⁰ ESTADÍSTICA aplicada a la investigación. U.N.de C.

DISTRIBUCIÓN A				
Ventas x_i	Marca de clase x'_i	Cantidad de Fábricas f_i	$(x'_i - \bar{x})^2$	$(x'_i - \bar{x})^2 f_i$
0-10	5	14	$(5-27)^2 = 484$	6776
10-20	15	18	$(15-27)^2 = 144$	2592
20-30	25	6	$(25-27)^2 = 4$	24
30-40	35	2	$(35-27)^2 = 64$	128
40-50	45	1	$(45-27)^2 = 324$	324
50-60	55	1	784	784
60-70	65	2	1444	2888
70-80	75	1	2304	2304
80-90	85	1	3364	3364
90-100	95	2	4624	9248
100-110	105	2	6084	12168
Σ		50		40600

$$S = \sqrt{\frac{40600}{49}} = 28.78$$

$$CV = \frac{201.50}{27} \cdot 100 = 7.46 \%$$

Cálculo de CV para la distribución C

La media aritmética para esta distribución es 35.65 millones de dólares.

Dispondremos los datos en la siguiente tabla:

DISTRIBUCIÓN C			
Ventas x_i	Marca de clase x'_i	Cantidad de Fábricas f_i	$(x'_i - \bar{x})^2 f_i$
0-5	2.5	8	8791.38
5-15	10	11	7237.1475
15-30	22.5	10	1729.225
30-35	32.5	5	49.6125
35-65	50	4	823.69
65-80	72.5	3	4073.7675
80-95	87.5	5	13442.1125
95-110	102.5	4	4468.9225
Σ		50	40615.8575

$$S = \sqrt{\frac{40615.8575}{49}} = 28.79$$

$$CV = \frac{28.79}{35.65} \cdot 100 = 80.76 \%$$

El CV de la distribución A = 106.59% y el CV de la distribución C = 80.76%. Esto nos dice que la distribución C es más homogénea. También se dice que la distribución A presenta mayor variabilidad.

Realiza la siguiente actividad:

- 1.- Grafique las frecuencias absolutas para las distribuciones A y B y realice comentarios.
- 2.- Complete el siguiente cuadro:

Medidas descriptivas	Distribución A	Distribución B	Distribución C
Media			
Mediana			
Moda			

Problema 9

El siguiente lote de 20 datos corresponde a lecturas de humedad (en porcentos) reducidos al entero más próximo:

Lote de datos

29 44 12 53 21
 34 39 25 48 23
 17 24 27 32 34
 15 42 21 28 37

Representar mediante un diagrama de Tallo y Hojas.

Solución

Serie simple: es un conjunto de pocos datos (generalmente $n < 30$).

Dado un lote de pocos datos cuantitativos (serie simple), además de los métodos exploratorios vistos (tabla de frecuencias), existe otra técnica sencilla de gran utilidad llamada “Diagrama de Tallo y Hojas” desarrollada por Tukey.

Para realizar el diagrama en primer lugar se ordenan los datos:

Lote de datos ordenados

12 15 17 21 21
 23 24 25 27 28
 29 32 34 34 37
 39 42 44 48 53

Luego se construye el diagrama colocando en una columna todos los números del lote de datos eliminando la última cifra (unidades) ordenados de menor a mayor. Esta columna constituye el tallo. A la derecha de cada número del tallo, se agrega la última cifra (unidad) de cada dato, ordenados también de menor a mayor, estas cifras constituyen las hojas.

Diagrama de Tallo y Hoja para el estudio de humedad

Tallo	Hojas							
1	2	5	7					
2	1	1	3	4	5	7	8	9
3	2	4	4	7	9			
4	2	4	8					
5	3							

La interpretación de este gráfico es la siguiente:

Por ejemplo tomando la fila 2 puede observarse que hay dos registros de lecturas de humedad de 21, una de 23, una de 24 etc.

De la misma manera se interpretan las demás filas.

Observación:

- a) En la **hoja** debe presentarse un solo dígito.
- b) En el **tallo** pueden presentarse uno o más dígitos.

Ejemplo: Para representar el número 219 se divide en 21 | 9

21 es el tallo
9 es la hoja

Problema 10

El largo de 8 hojas expresadas en mm de una cierta especie vegetal arroja los siguientes valores:

71; 65; 64; 64; 66; 65; 69; 70

Calcule medidas de posición.

Solución:

Recordemos que siempre debemos definir la variable de estudio, clasificarla e indicar la escala de medición.

X: Largo de hojas expresadas en mm.

La variable es cuantitativa continua y la escala de medición es de razón.

Los 10 datos constituyen una **serie simple**

Cálculo de la media aritmética

La expresión para el cálculo de la media aritmética en una **serie simple** (datos no agrupados) es-

$$\text{tá dada por: } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Reemplazando por los datos del problema tenemos:

$$\bar{x} = \frac{71 + 65 + 64 + 64 + 66 + 65 + 69 + 70}{8} = 66.75\text{mm}$$

En promedio las hojas tienen una longitud de 66.75mm

Cálculo de la mediana

Para el cálculo de la mediana conviene ordenar los datos de menor a mayor:

$$64; \quad 64; \quad 65; \quad 65; \quad \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \quad 66; \quad 69; \quad 70; \quad 71$$

Me

Como la serie tiene un número par de datos la mediana está dado por el promedio de los dos valores centrales n_4 y n_5 . En nuestro problema $Me = \frac{65 + 66}{2} = 65.5\text{mm}$

- 4 de las hojas tienen un largo de 65.5 mm o menos.

- 4 de las hojas tienen un largo de 65.5 mm o más.

Para el cálculo de los cuartiles se procede en forma similar al caso del cálculo de la mediana. Los cuartiles dividen a la serie simple en cuatro partes iguales:

$$64; \quad 64; \quad \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \quad 65; \quad 65; \quad \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \quad 66; \quad 69; \quad \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \quad 70; \quad 71$$

Q₁ Q₂ Q₃

$$Q_1 = \frac{64 + 65}{2} = 64.5\text{mm}; \quad Q_2 = Me = 65.5\text{mm}; \quad Q_3 = \frac{69 + 70}{2} = 69.5\text{mm}$$

Observación: Si el número de datos de la serie es impar, entonces la mediana está dada por el valor central.

Cálculo de la moda

Longitud en mm	64	65	66	69	70	71
Cantidad	2	2	1	1	1	1

Esta serie simple de datos tiene dos valores modales: 64mm y 65mm. Se dice entonces que es bimodal.

Problema 11

El registro de los pesos (expresados en Kg.) de 5 niños al nacer es el siguiente:

2.800 3.600 3.700 2.500 3.900

Calcule e interprete medidas de dispersión.

Solución

La variable de estudio es X: Peso de 5 niños al nacer.

Es una variable cuantitativa continua y la escala de medición es de razón.

En este problema tenemos también una **serie simple** o datos no agrupados.

Recordemos que la medida de dispersión a utilizar depende de la medida de posición elegida. En nuestro problema y a modo de ejemplificar calcularemos la desviación media, la desviación mediana, la varianza y la desviación estándar.

Cálculo de la desviación media

La desviación media se define como: $DM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$

Para facilitar el cálculo se ordenan los datos en la siguiente tabla:

Tabla XXIV: Desviación media del peso de los niños al nacer

Pesos en Kg.	Desviación $x_i - \bar{x}$	Desviación absoluta $ x_i - \bar{x} $
2.800	-0.500	0.500
3.600	0.300	0.300
3.700	0.400	0.400
2.500	-0.800	0.800
3.900	0.600	0.600
$\Sigma = 16.500$		$\Sigma = 2.600$

La media aritmética es: $\bar{x} = 3.300\text{Kg}$

$$DM = \frac{2.6}{5} = 0.52\text{Kg}$$

Interpretación: En promedio, los desvíos de los pesos con respecto a la media es de 0.52Kg.

Cálculo de la desviación mediana

Para obtener la desviación mediana reemplazaremos en la expresión de la desviación media, la media

aritmética por la mediana: $DMe = \frac{\sum_{i=1}^n |x_i - Me|}{n}$

En primer lugar se calcula la mediana y para ello conviene ordenar los datos de menor a mayor y como $n = 5$, el valor de la mediana está dado por el valor central.

2.500 2.800 3.600 3.700 3.900
Me

La mediana $Me = 3.600$ Kg. y la tabla de las desviaciones es:

57

Tabla XXV: Desviación mediana del peso de los niños al nacer

Pesos en Kg.	Desviación $x_i - Me$	Desviación absoluta $ x_i - Me $
2.800	-0.800	0.800
3.600	0.000	0.000
3.700	0.100	0.100
2.500	-1.100	1.100
3.900	0.300	0.300
		$\Sigma = 2.300$

La desviación mediana $DMe = 2.300 / 5 = 0.460$ Kg.

Cálculo de la varianza

La varianza para datos sin agrupar está dada por: $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$

La desviación estándar está dada por: S

La fórmula de trabajo o fórmula abreviada para el cálculo de la varianza es la siguiente:

$$S^2 = \frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n-1)}$$

Verifique que la S^2 es 0.375 Kg^2 y la S es 0.612 Kg .

Observación:

Puede encontrar estas medidas en forma directa, utilizando una calculadora científica con comandos que realicen cálculos estadísticos. Para ello debe cargar los datos y luego presionar las teclas de la media, de la varianza de la desviación estándar según el cálculo que desea realizar (lea el manual de uso de la calculadora).