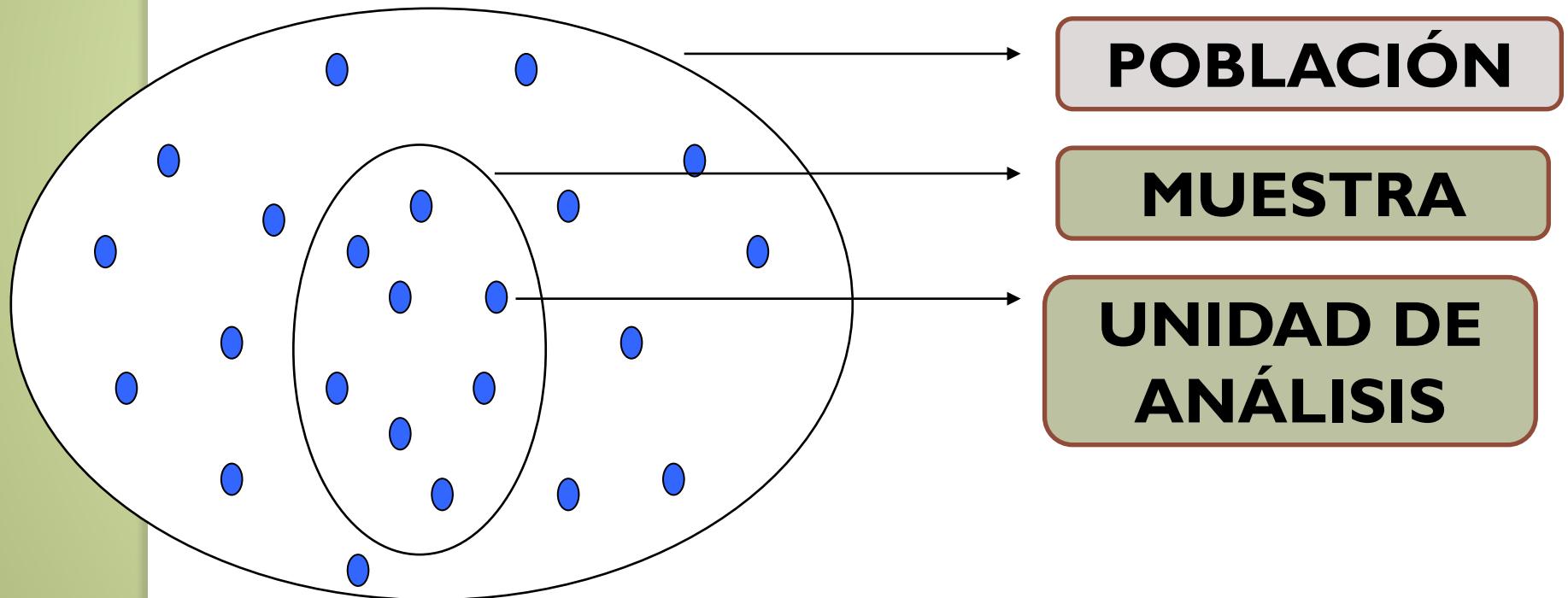


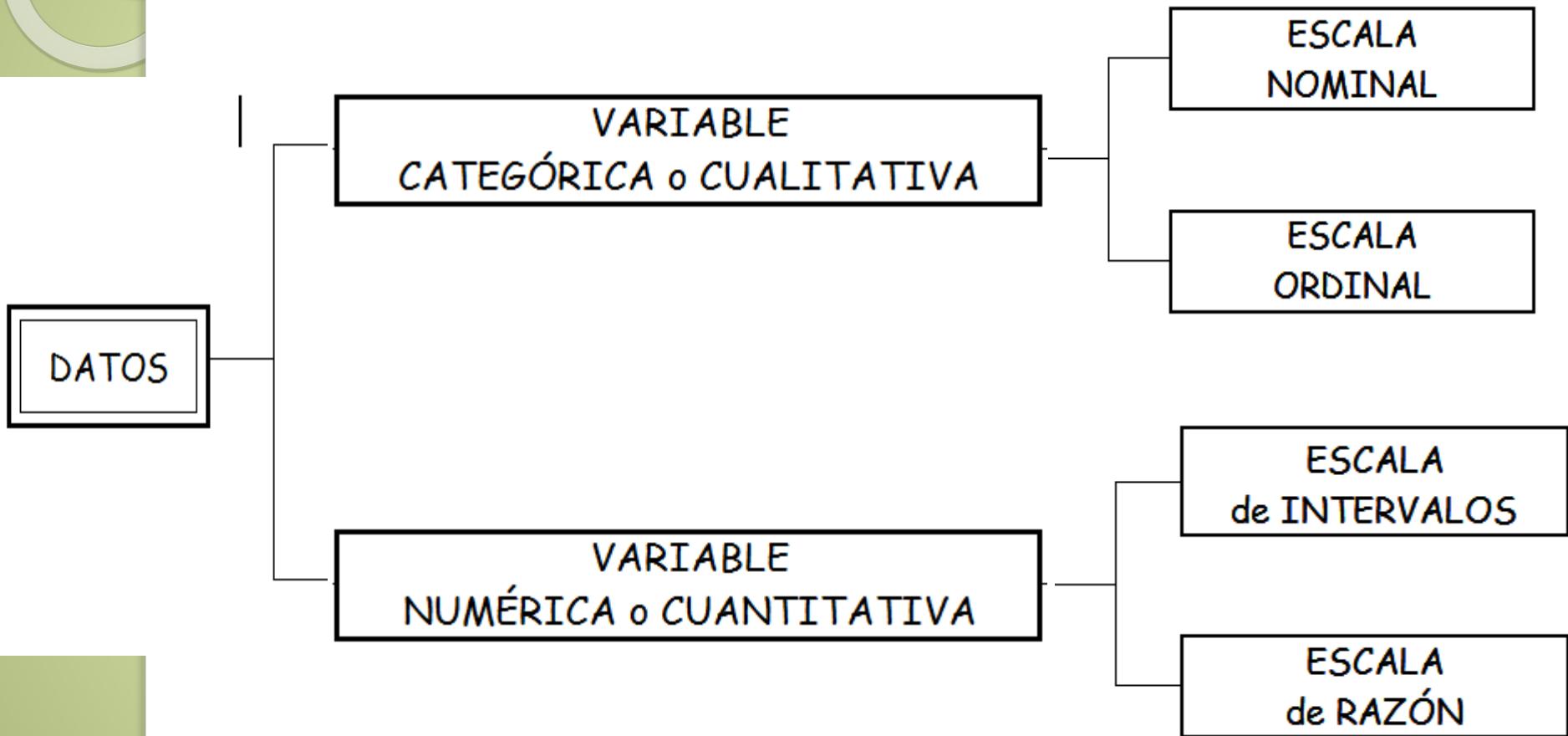


- **ESTADÍSTICA DESCRIPTIVA**

Algunos conceptos imprescindibles



Tipos de datos y escala de medición



Estadística Descriptiva y Análisis de Datos

Presentación de Datos

***TEXTO**

***TABLAS**

***GRÁFICOS**



- DATOS SIN AGRUPAR

Variable cuantitativa con datos sin agrupar

- Sea X : “**Número de cuadras caminadas por 14 alumnos de una escuela rural, para llegar cada mañana**”.

5	5	5	6	8	4	4	2	1	8	6	6	4	5
---	---	---	---	---	---	---	---	---	---	---	---	---	---

- **Primeramente ordenamos los datos**

1	2	4	4	4	5	5	5	5	6	6	6	8	8
---	---	---	---	---	---	---	---	---	---	---	---	---	---

Frecuencia absoluta- relativa

- **Frecuencia absoluta:**
- Es el número de veces que se presenta cada valor de la variable.

$$\sum_{i=1}^m f_i = n$$

- **Frecuencia relativa:**

Es el cociente entre la frecuencia absoluta f_i y el número total de elementos n de la muestra.

$$f_{ri} = \frac{f_i}{n}$$

$$0 \leq f_{ri} \leq 1$$

TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

X	f	F
1	1	1
2	1	2
4	3	5
5	4	9
6	3	12
8	2	14
Total	$\sum f_i = 14$	

Gráfico de bastones para una variable cuantitativa con datos sin agrupar

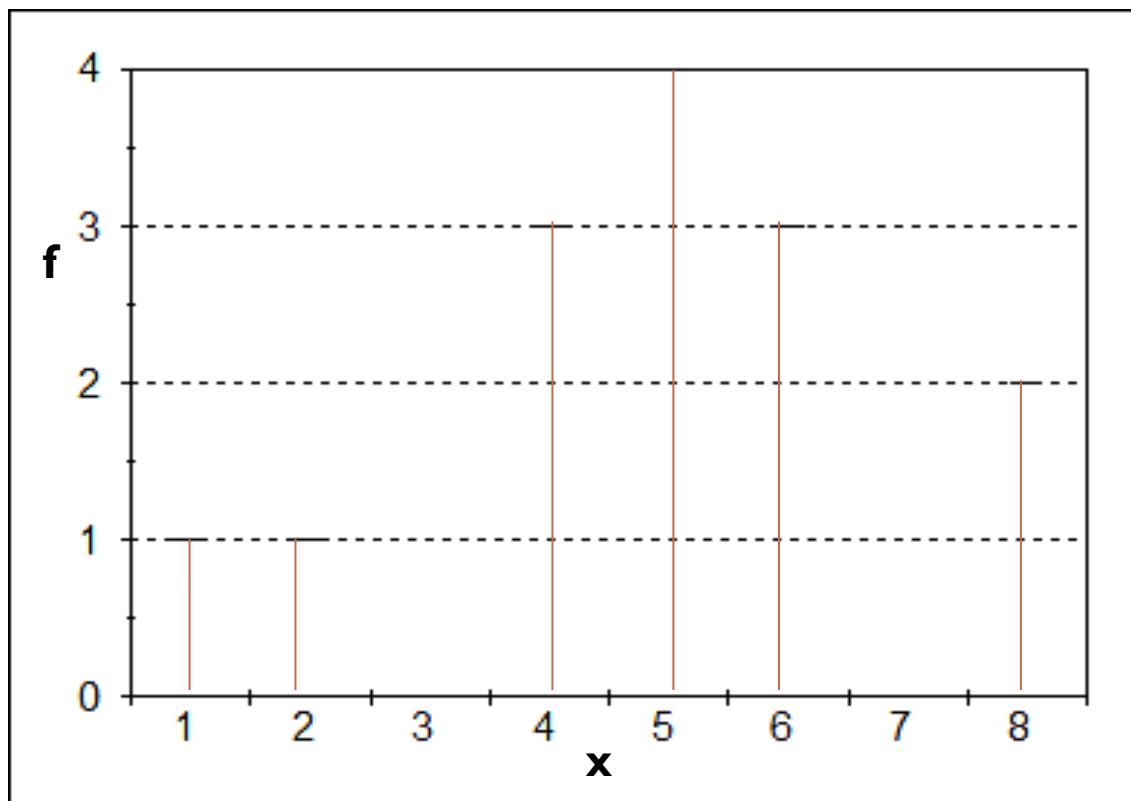


Gráfico de escalera para las F.A.

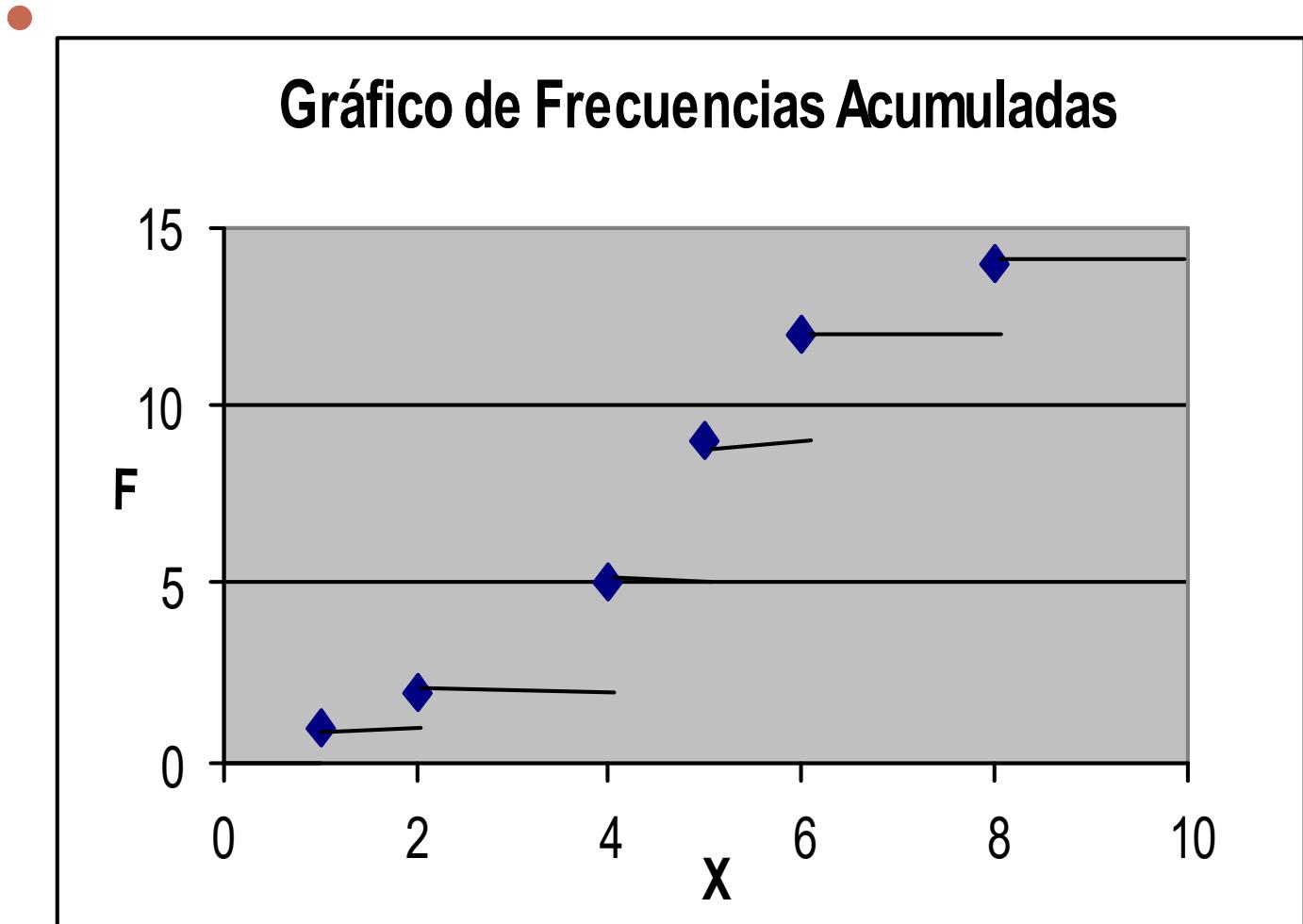




Gráfico de tronco y hojas o de Tallo y hojas

Se desea analizar cuánto demora un procesador X en guardar un archivo de cierto tamaño medido en segundos.

0,2	0,4	0,5	0,5	0,7	0,7	0,8	0,9	0,9	1,2
1,2	1,2	1,4	1,4	1,5	1,6	1,9	2,1	2,2	2,4
2,6	2,6	3,7	3,8	3,9					

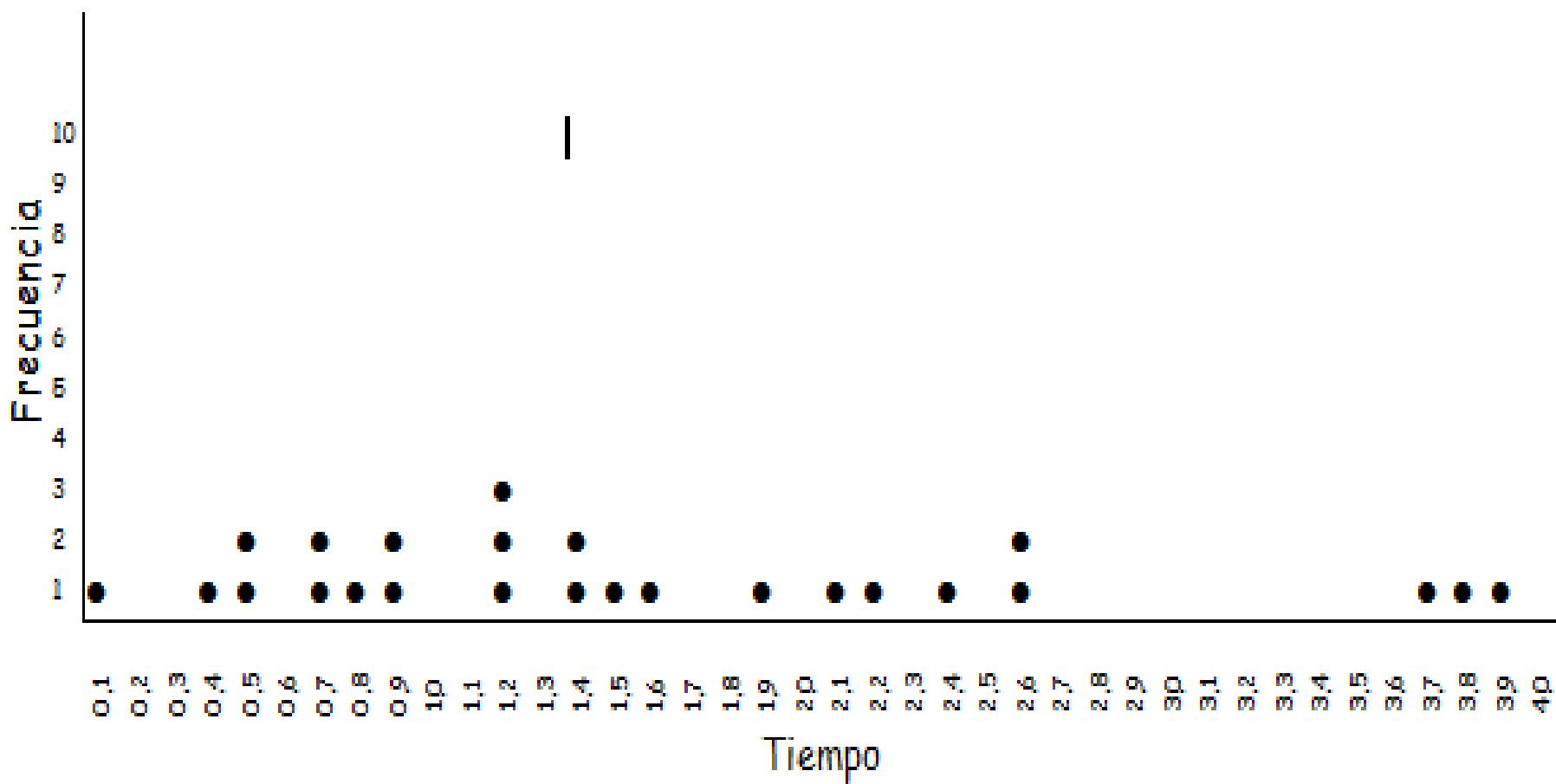
Troncos	Hojas	Frecuencia	Frecuencia relativa
0	2 4 5 5 7 7 8 9 9	9	0,36
1	2 2 2 4 4 5 6 9	8	0,32
2	1 2 4 6 6	5	0,20
3	7 8 9	3	0,12
		n = 25	1,00

Stem-and-Leaf Display for Tiempo: unit = 0,1 1|2 represents 1,2

2	0 24
9	0 5577899
(5)	1 22244
11	1 569
8	2 124
5	2 66
3	3 1
3	3 789

DIAGRAMA DE PUNTOS

Tiempo de guardado de determinados archivos por un procesador X



Fuente: Datos hipotéticos



Ejemplo Variable cualitativa

En un estudio realizado por el Instituto del hierro y el acero de Estados Unidos durante el año 1992, se analizó las cantidades (en miles de toneladas) de importaciones de acero, en distintos países:

Principales fuentes de importaciones de acero en Estados Unidos durante 1992

Países x_i	Frecuencia simple absoluta f_i	Frecuencia simple relativa f_{ri}	Frecuencia simple relativa porcentual $f_{ri} \%$
Bélgica y Luxemburgo	1247	0,3041	30,41 %
Japón	1072	0,2615	26,15 %
Alemania	460	0,1122	11,22 %
Canadá	367	0,0895	8,95 %
Francia	299	0,0729	7,29 %
Reino Unido	250	0,0610	6,10 %
Otros	405	0,0988	9,88 %
	$n = 4100$	1,0000	100,00 %

Fuente: U.S. Department of Commerce. Datos preparados por el American Iron and Steel Institute, publicados en Charting Steel's Progress in 1992.

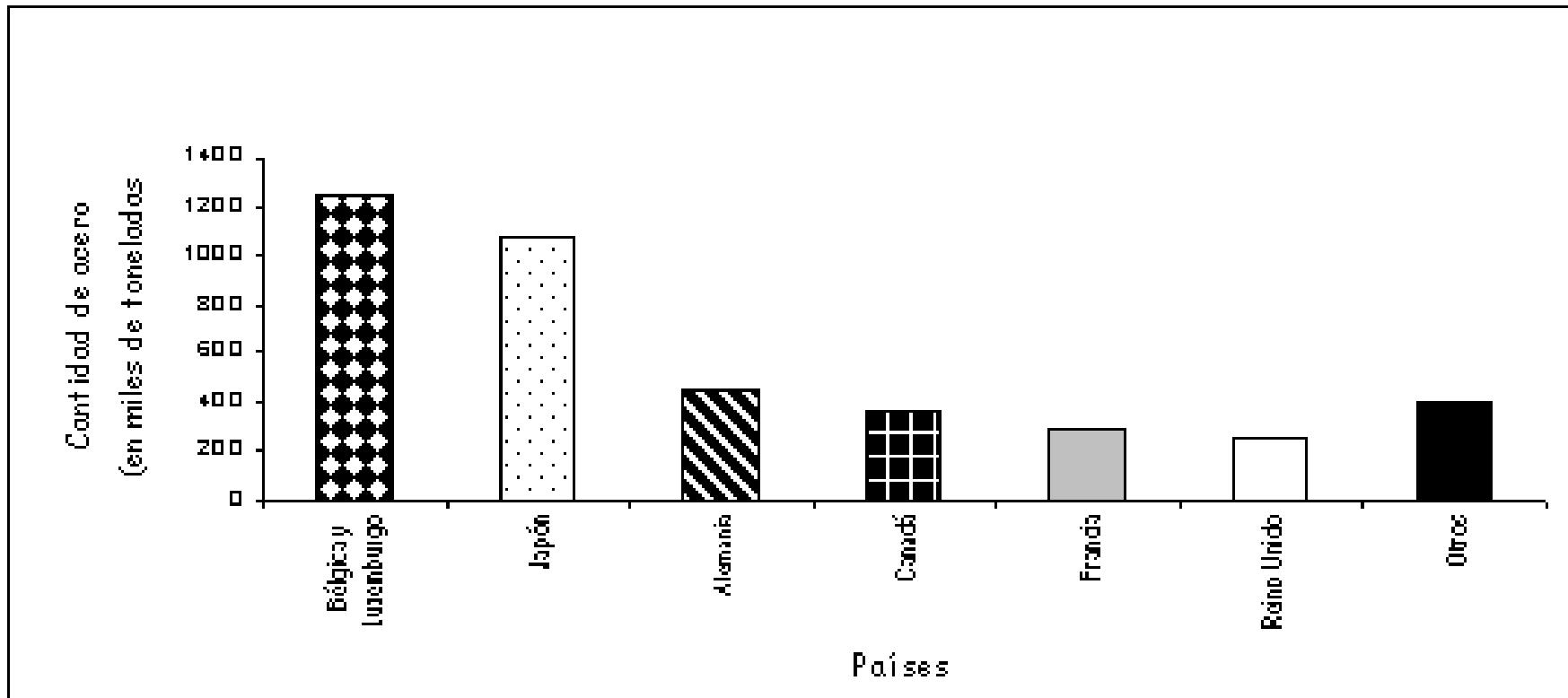
Nota: Para poder operar con los datos de la tabla o referirnos a ella, podemos representar la característica a observar (países) mediante la variable X y a la modalidad i -ésima de dicha variable con la notación x_i .

GRÁFICOS



Gráfico de barras verticales

Principales fuentes de importaciones de acero en Estados Unidos durante 1992

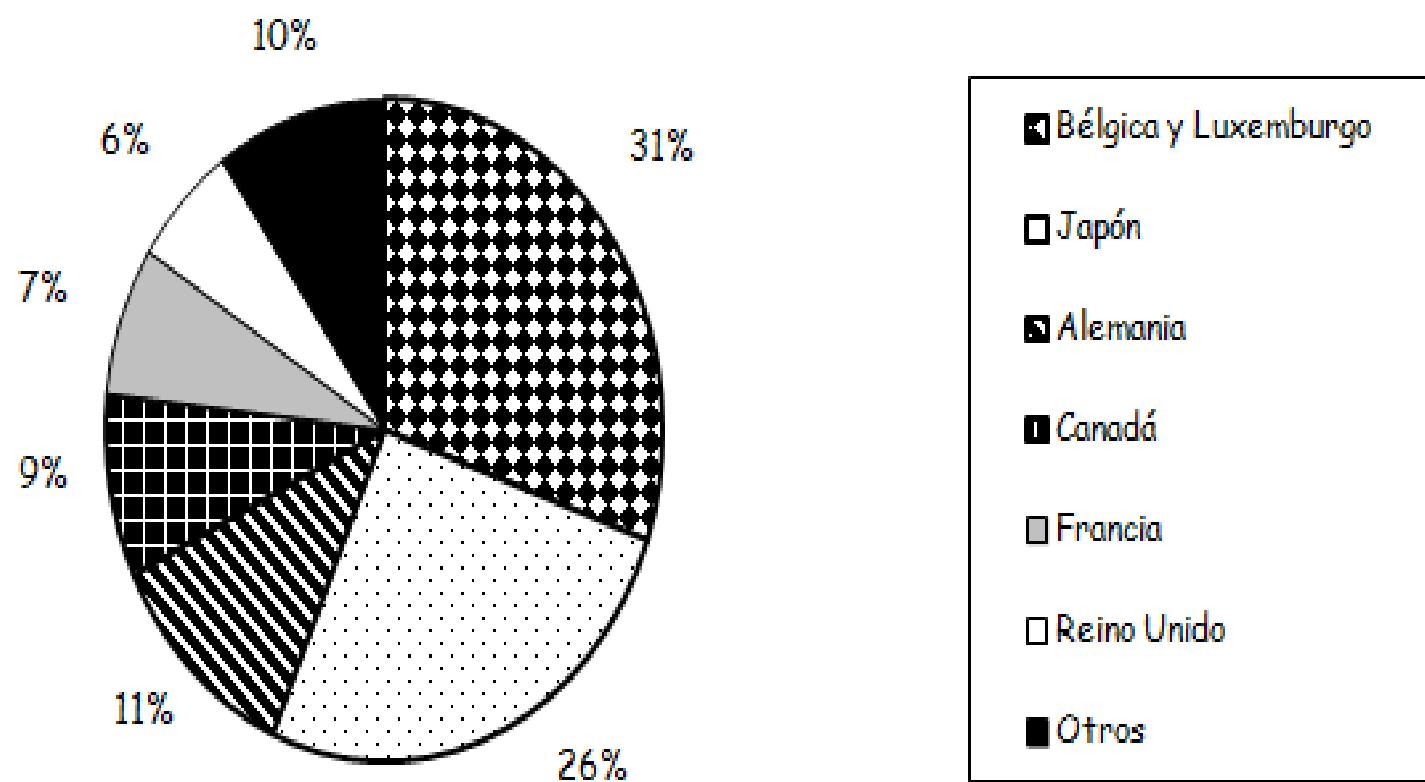


Fuente: U.S. Department of Commerce. Datos preparados por el American Iron and Steel Institute, publicados en *Charting Steel's Progress* in 1992.



Gráfico de sectores

Principales fuentes de importaciones de acero en Estados Unidos durante 1992



Fuente: U.S. Department of Commerce. Datos preparados por el American Iron and Steel Institute, publicados en Charting Steel's Progress in 1992.



Ejemplo:

Las siguientes son las alturas, en centímetros, de sesenta alumnos universitarios:

150	160	161	160	160	172	162	160	172	151
.
163	168	171	178	179	164	176	163	182	162

Estatura de sesenta estudiantes universitarios de Mendoza en 2004

Valores observados x_i	Frecuencia simple absoluta f_i	Frecuencia simple relativa $f_{ri} = f_i / n$	Frecuencia simple relativa porcentual $f_{ri}\%$	Frecuencia acumulada absoluta F_i	Frecuencia acumulada relativa $F_{ri} = F_i/n$	Frecuencia acumulada relativa porcentual $F_{ri}\%$
149	1	0,0167	1,67 %	1	0,0167	1,67%
150	1	0,0167	1,67 %	2	0,0333	3,33%
151	1	0,0167	1,67 %	3	0,0500	5,00%
184	1	0,0167	1,67 %	60	1,0000	100,00%
	$n = 60$					

Fuente: Datos hipotéticos



- DATOS AGRUPADOS

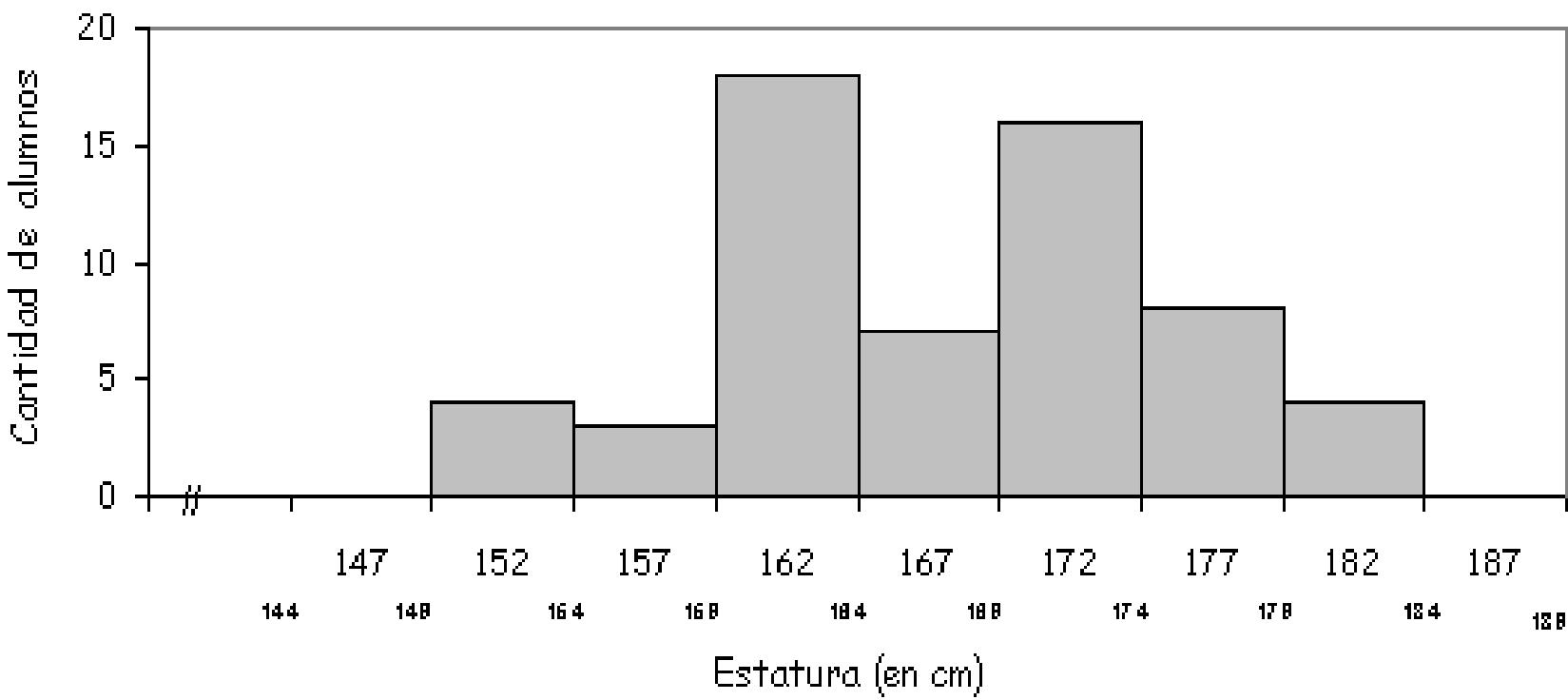
Estatura de sesenta estudiantes universitarios de Mendoza en 2004

Intervalos o clases	Punto medio	Frecuencia simple absoluta	Frecuencia simple relativa	Frecuencia simple relativa porcentual	Frecuencia acumulada absoluta	Frecuencia acumulada relativa	Frecuencia acumulada relativa porcentual
	x_i	f_i	f_{ri}	$f_{ri}\%$	F_i	F_{ri}	$F_{ri}\%$
[149 , 154)	151,5	4	0,0667	6,67%	4	0,0667	6,67%
[154 , 159)	156,5	3	0,0500	5,00%	7	0,1167	11,67%
[159 , 164)	161,5	18	0,3000	30,00%	25	0,4167	41,67%
[164 , 169)	166,5	7	0,1166	11,66%	32	0,5333	53,33%
[169 , 174)	171,5	16	0,2667	26,67%	48	0,8000	80,00%
[174 , 179)	176,5	8	0,1333	13,33%	56	0,9333	93,33%
[179 , 184]	181,5	4	0,0667	6,67%	60	1,0000	100,00%
		$n = 60$	1,0000	100 %			

Fuente: Datos hipotéticos

HISTOGRAMA

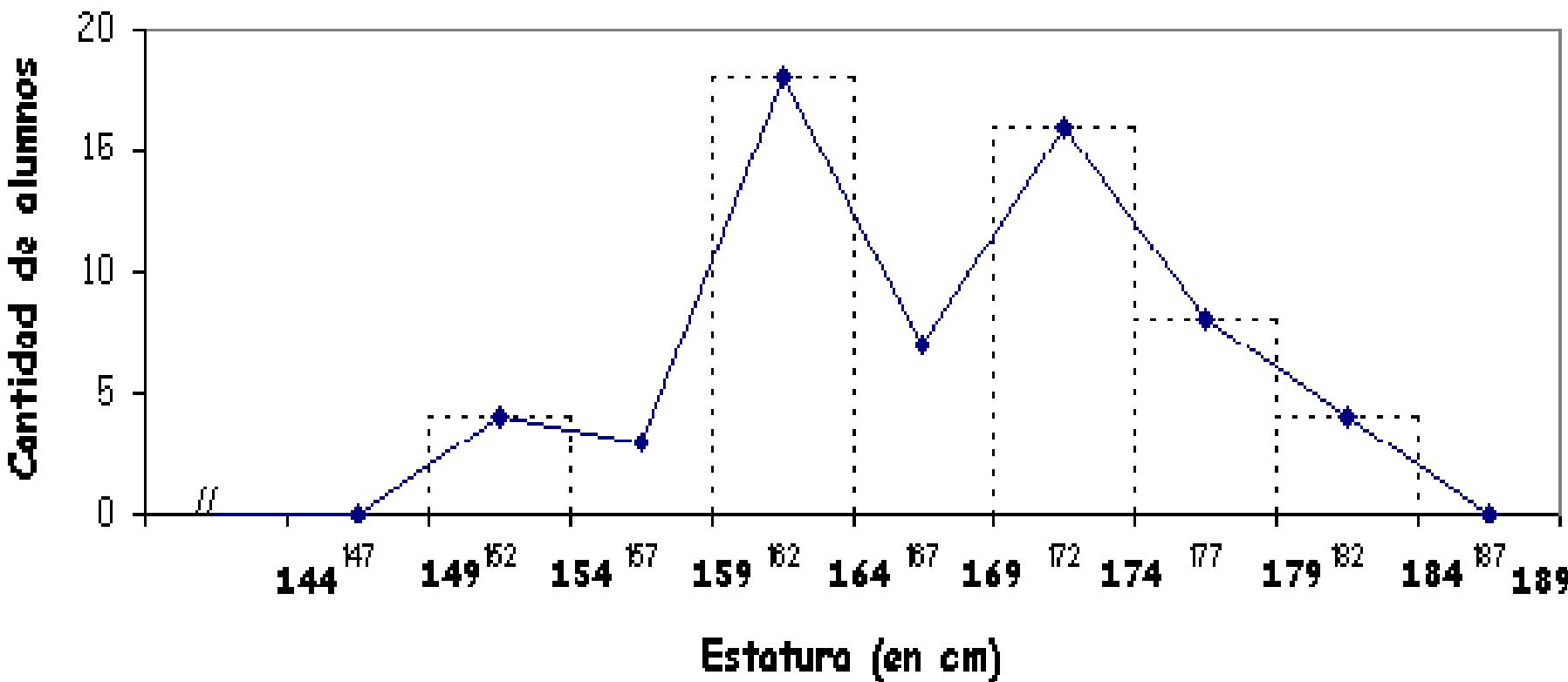
Estatura de un grupo de estudiantes universitarios



Fuente: Datos hipotéticos

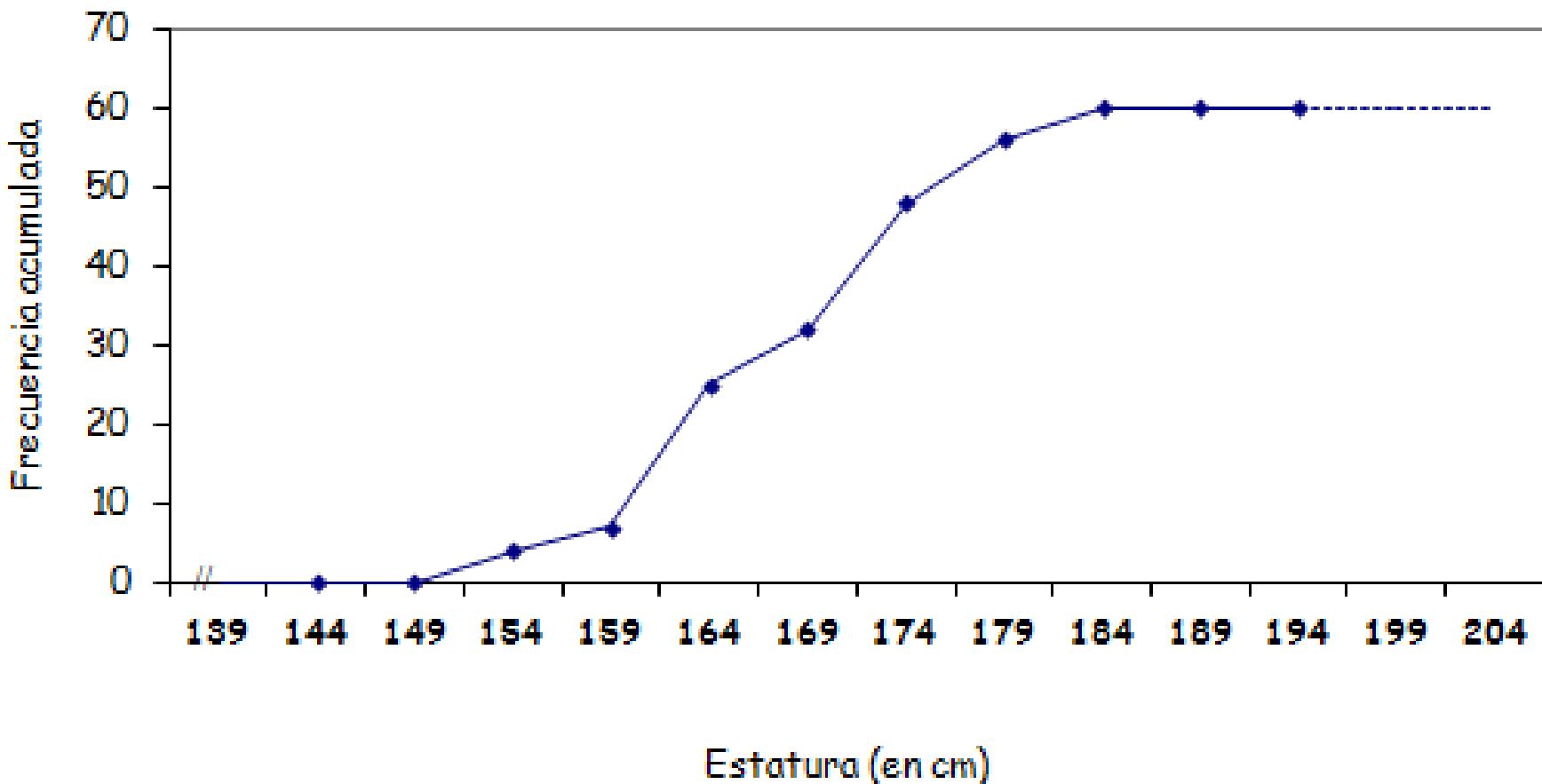
POLIGONO DE FRECUENCIAS

Estatura de un grupo de estudiantes universitarios



OJIVA

Estatura de un grupo de estudiantes universitarios



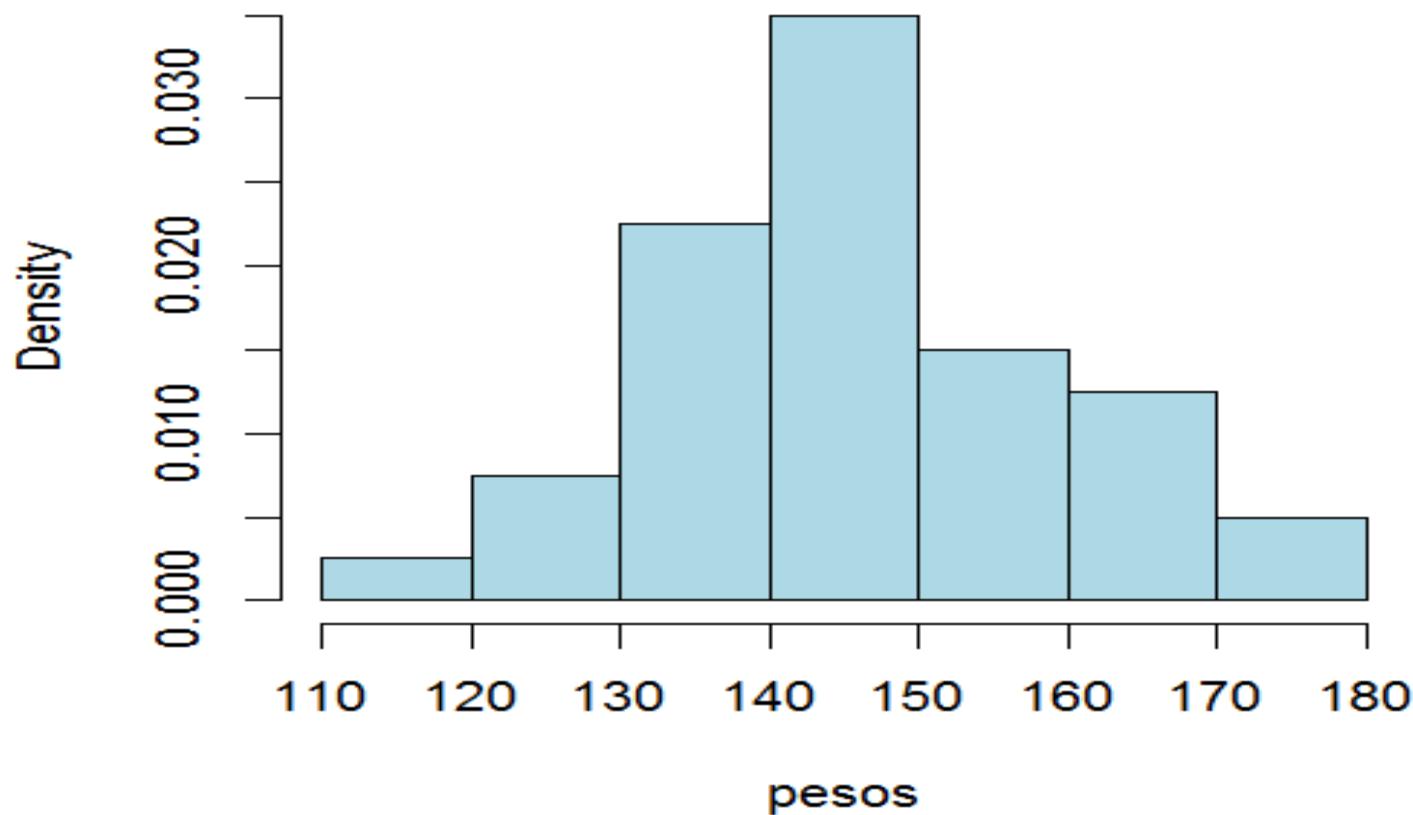
Ejemplo con R

- Sea X: El peso de 40 estudiantes (libras)

119	125	126	128	132	135	135	135
136	138	138	140	140	142	142	144
144	145	145	146	146	147	147	148
149	150	150	152	153	154	156	157
158	161	163	164	165	168	173	176

- >pesos=c(119,...,176)
- Si los datos no estan ordenados
- > sort(pesos)
- >hist(pesos)
- >hist(pesos, col="lightblue", probability=T)
- hist(pesos, col="lightblue", probability=T, ylab="frecuencia relativa", main="Pesos de los estudiantes")
- >stem(pesos)

Histogram of pesos



Medidas numéricas descriptivas

MEDIDAS DE TENDENCIA CENTRAL

MEDIDAS DE DISPERSIÓN

MEDIDAS DE POSICIÓN

MEDIDAS DE FORMA

Medidas numéricas descriptivas

- Medidas de tendencia central
 - Media
 - Mediana
 - Moda
- Medidas de dispersión
 - Rango
 - Varianza
 - Desviación estándar
 - Coeficiente de Variación
- Medidas de posición
 - Cuartiles
 - Deciles
 - Percentiles

Medidas de tendencia central

- Media Es el promedio aritmético de los datos.
- Mediana El **valor de la variable** que ocupa la **posición central**, en un conjunto ordenado de datos.
- Moda Es el **valor de la variable** que se presenta con **mayor frecuencia**



- **MEDIDAS PARA DATOS SIN AGRUPAR**

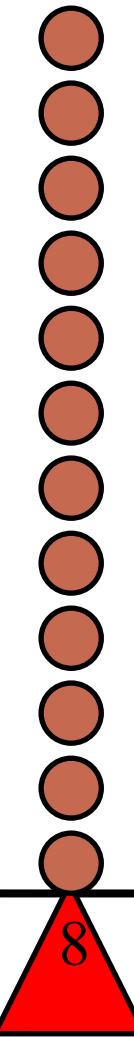
Seguimos con el ejemplo de los talleres FIX

Los talleres de trasmisión Fix-An están analizando el tiempo que les toma a los mecánicos retirar, reparar y volver a colocar una trasmisión.

A continuación se analizará el tiempo en horas que se tardó en reparar doce transmisiones en tres sucursales distintas de la empresa.

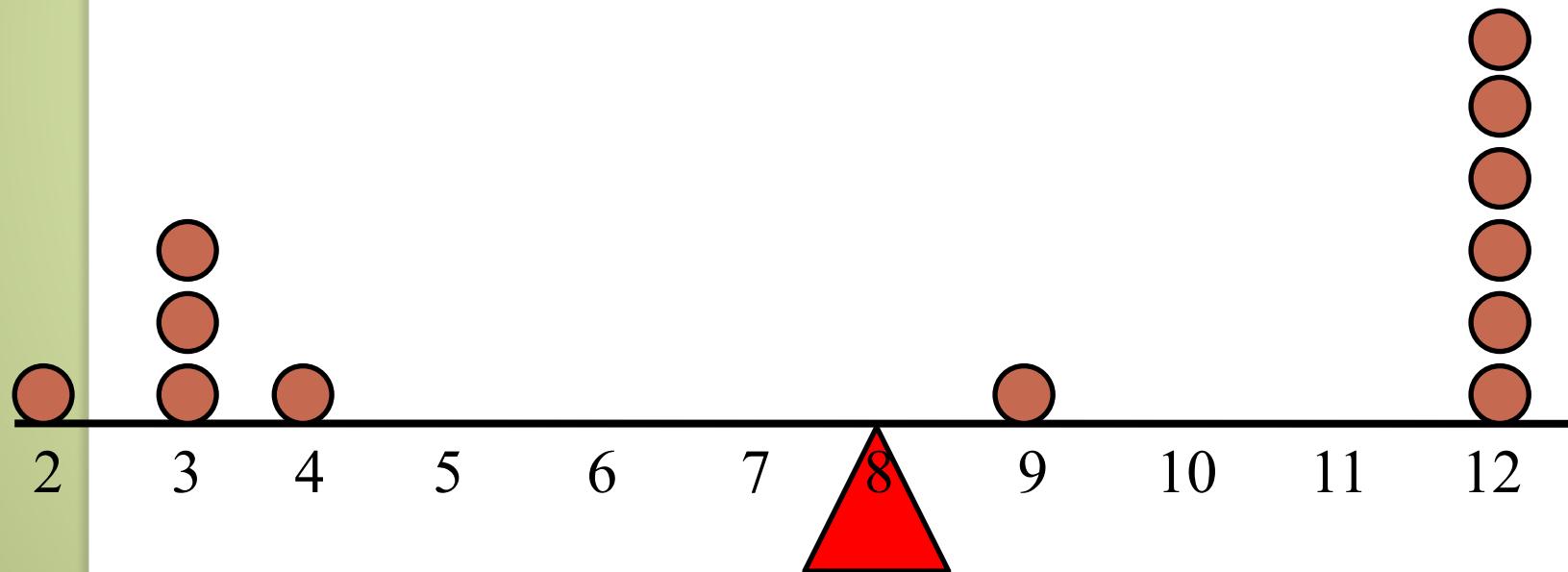
Primer sucursal....

¿Cuanto demoraron en promedio en armar un transmisión?



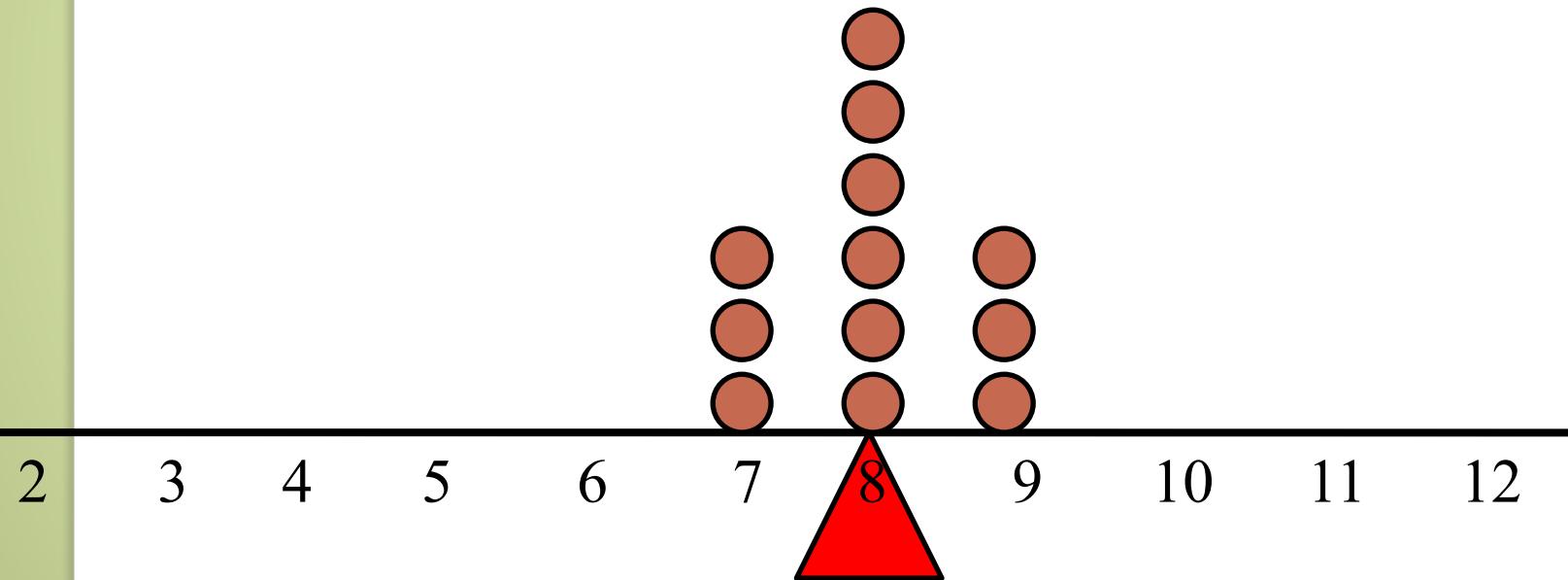
Segunda sucursal...

¿Cuanto demoraron en promedio en armar una transmisión?

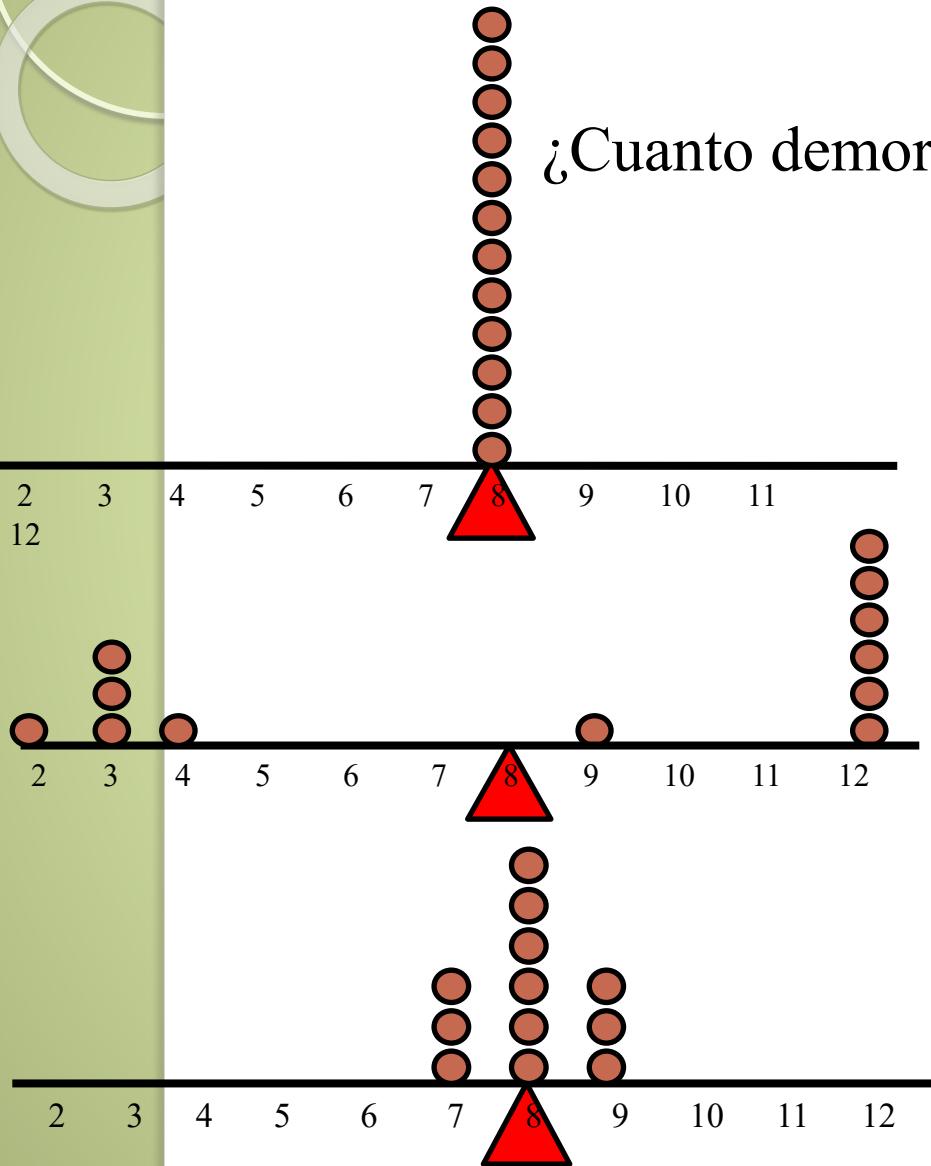


Tercera sucursal....

¿Cuanto de moraron en promedio en armar una transmisión?



Las tres sucursales....



¿Cuanto demoraron en promedio cada sucursal?

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{n}$$
$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{12} = 8$$

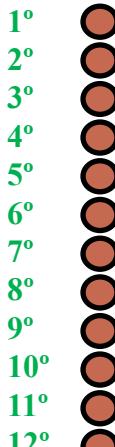
Determinación de Mediana

- Si el conjunto de datos es **impar** y están ordenados en forma creciente o decreciente, el valor de la mediana es el **valor central**.
- Si el conjunto de datos es **par** y están ordenados en forma creciente o decreciente, el valor de la mediana se calcula como el promedio aritmético de las dos observaciones centrales.

Determinación de la mediana

- Si $n=impar$
- Ejemplo
- 2-4-6-8-9
- 1-Ubicamos el lugar central $L=(n+1)/2=3$
- 2- Observamos el valor que se encuentra en el lugar central
- $X_{me}=6$

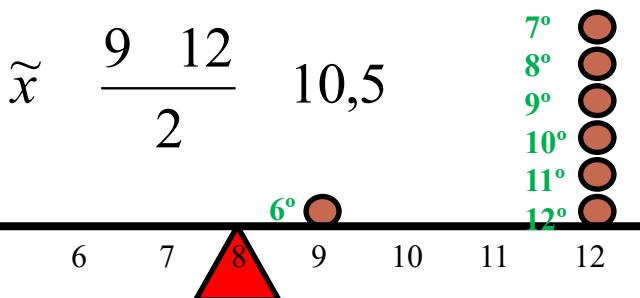
Si $n=\text{par.}$ Las tres sucursales....



8



¿Cuál es el valor de la **Mediana** en cada sucursal?



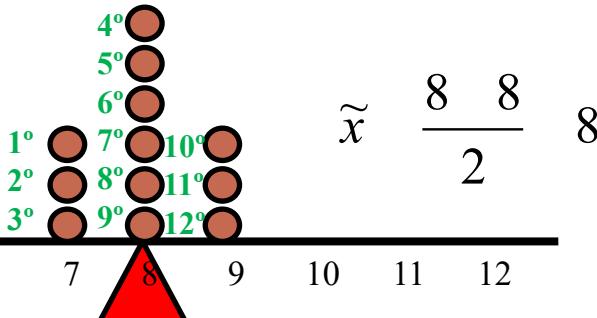
$$\tilde{x} \quad \frac{9 + 12}{2} \quad 10,5$$

Datos pares

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

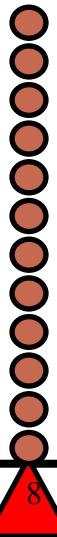
Orden de la mediana entre el 6° y 7°

Mediana: promedio de los valores centrales



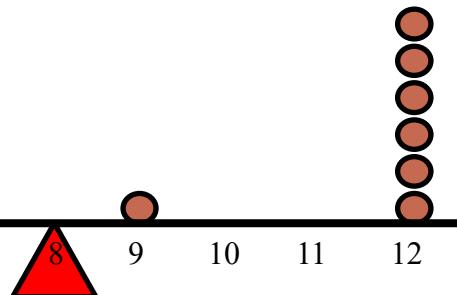
$$\tilde{x} = \frac{8 + 8}{2} = 8$$

Las tres sucursales....

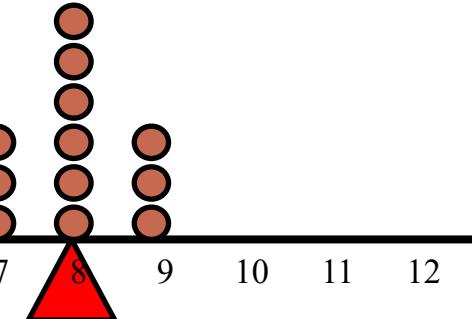


¿Cuál es el valor de la **moda** en cada sucursal?

Moda:8



Moda:3,12

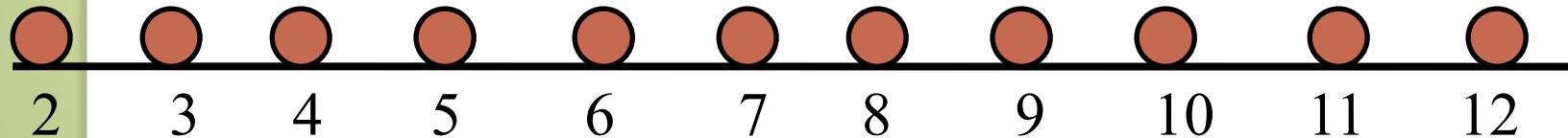


Moda:8

Moda

¿Cuál es el valor de la **moda** en esta nueva sucursal?

NO HAY MODA



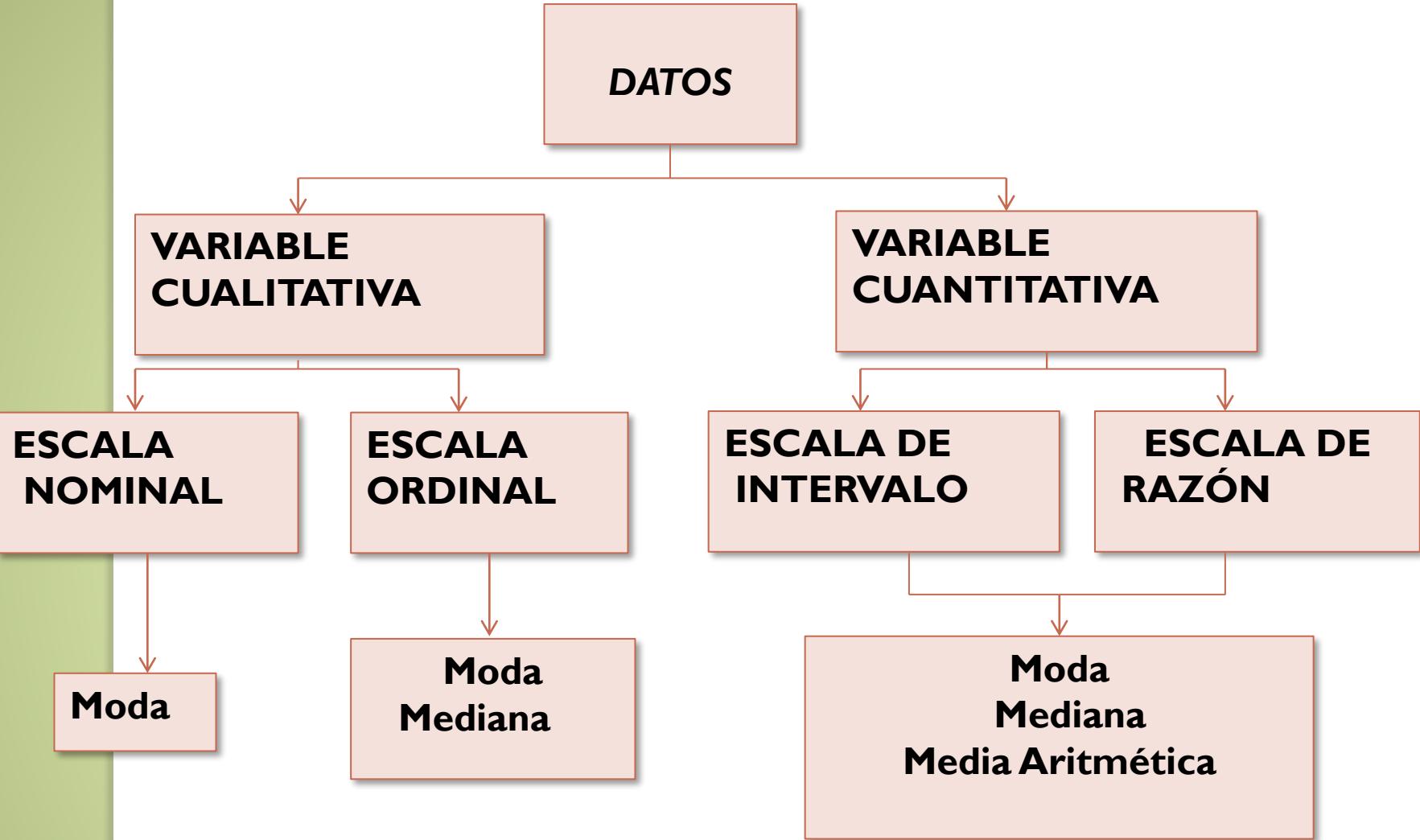
- Podría ver más de una moda?

MEDIDAS DE TENDENCIA CENTRAL

MEDIA

MEDIANA

MODA
VENTAJAS Y DESVENTAJAS



Con R

- `>xbarra=mean(pesos)`
- `>xbarra`

$$\bar{x} \quad 146,8$$

- `>scuadrado=var(pesos)`

MEDIDA DE DISPERSIÓN

- **RANGOS**
- **VARIANZA**
- **DESVIACIÓN ESTÁNDAR**
- **COEFICIENTE DE VARIACIÓN**

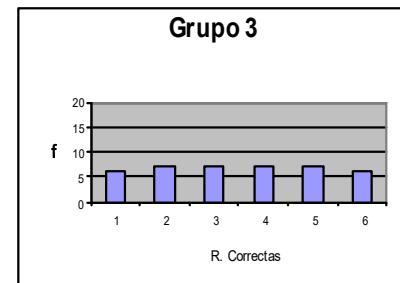
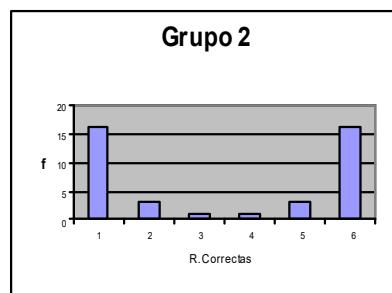
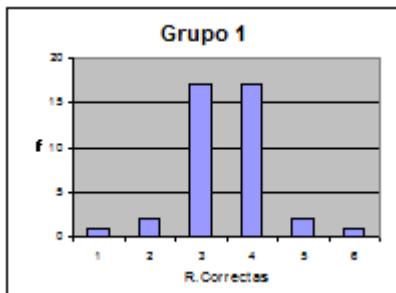


MEDIDAS DE DISPERSIÓN

- Las medidas de dispersión nos proporcionan una medida del mayor o menor agrupamiento de los datos respecto a los valores de tendencia central.
- Son positivas (mayores o iguales a 0)
- Un valor cero indica ausencia de dispersión

Medidas de TC- Medidas de dispersión

- Un promedio puede ser engañoso a menos que vaya acompañado de otra información que nos diga la amplitud o sus desviaciones con relación al promedio.



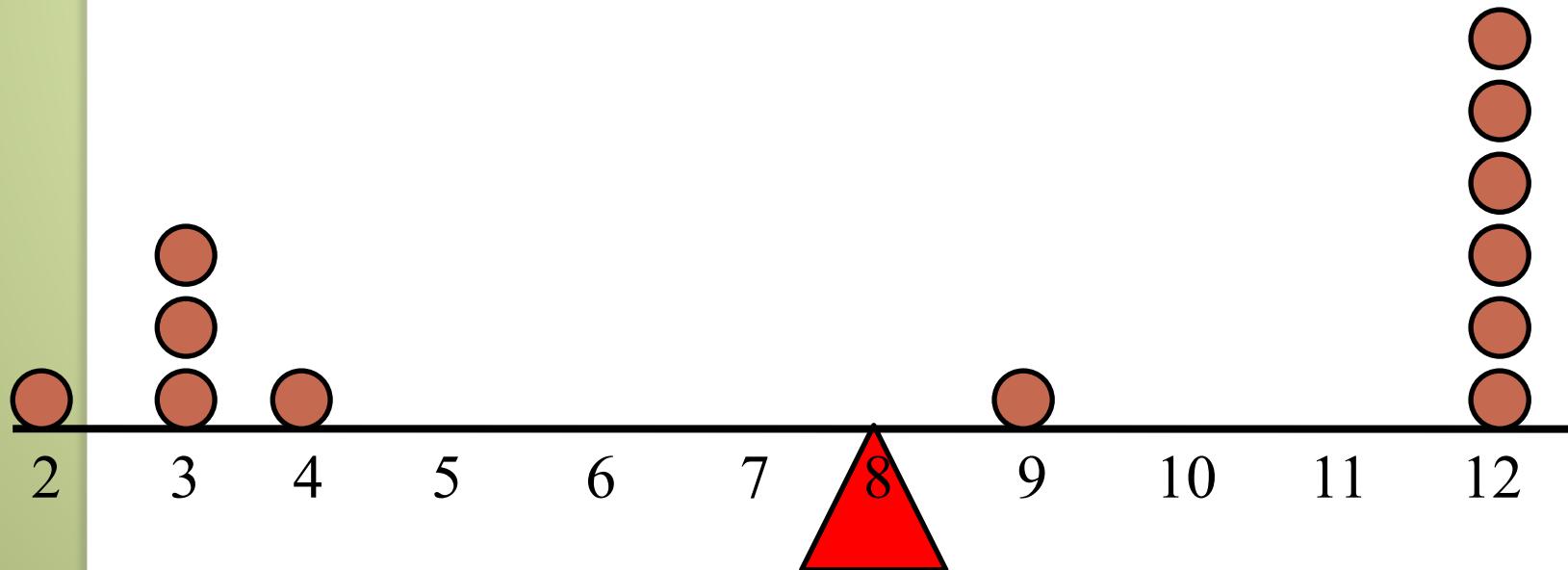
- Tienen la misma media aritmética, 2,5 puntos ¿pero podemos afirmar que hay homogeneidad entre los grupos?. Gráficamente vemos que el valor de la media aritmética no es suficiente para describir cada una de las situaciones.



Medidas de dispersión- Rango

RANGO

$$R \quad x_{\max} \quad x_{\min} \quad 12 \quad 2 \quad 10$$

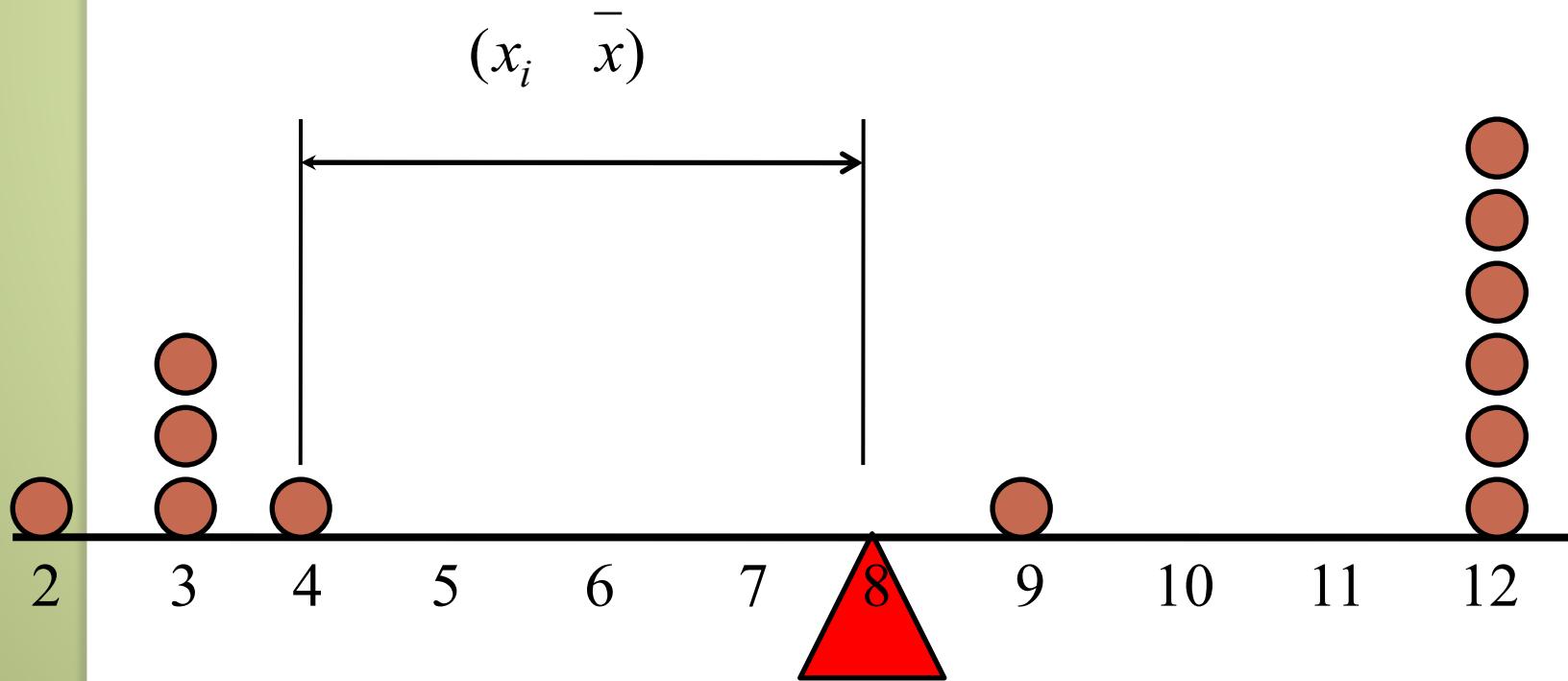


Rango- Rango intercuartil- R. Interdecil

- $R = \text{xmax} - \text{xmin}$
- El rango proporciona una rápida indicación de la variabilidad existente entre las observaciones de un conjunto de datos.
- La diferencia entre los percentiles 75avo y 25avo recibe el nombre de **recorrido intercuartil**, sólo incluye el 50% central de la distribución.

Medidas de dispersión

$$\text{medidas de dispersión} = \sum_{i=1}^n (x_i - \bar{x})^2$$



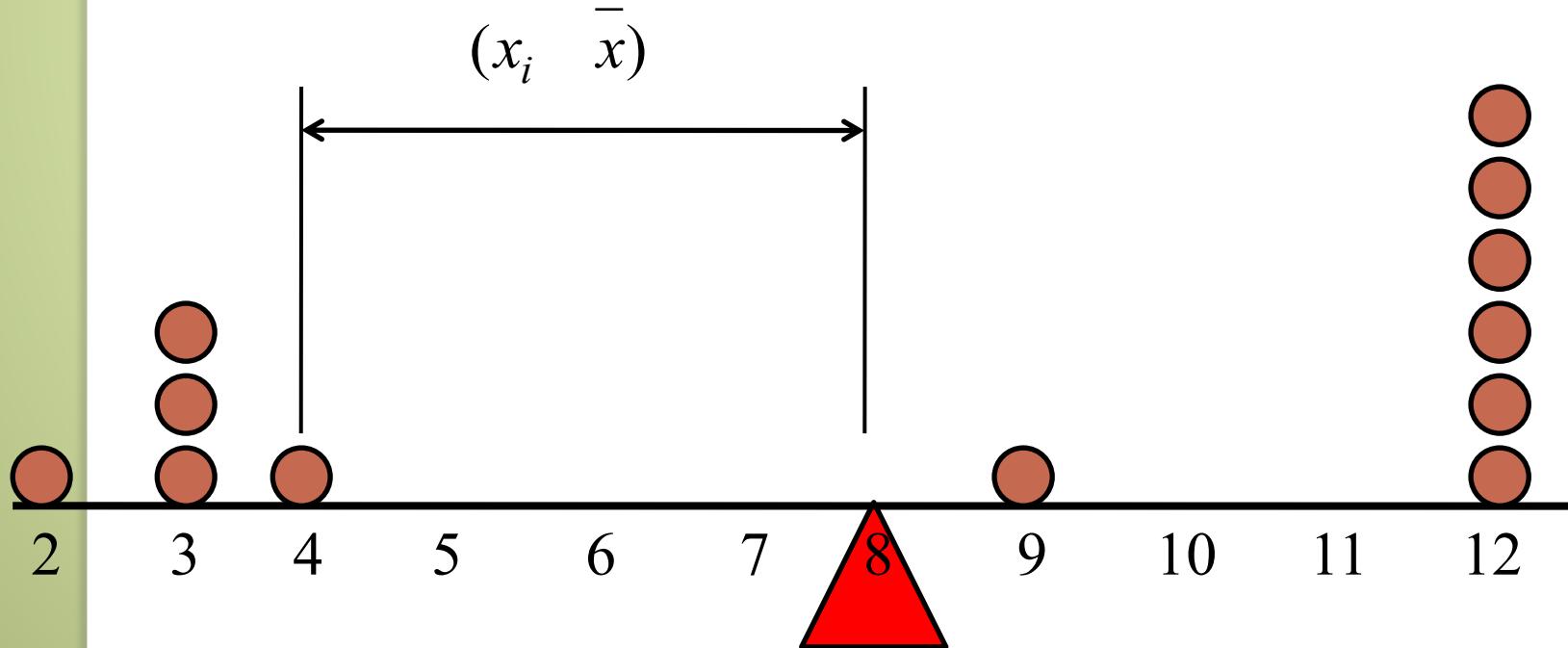
Medidas de dispersión

VARIANZA
MUESTRAL

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

DESVIACIÓN
ESTÁNDAR
MUESTRAL

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$



Varianza-Desviación Estándar

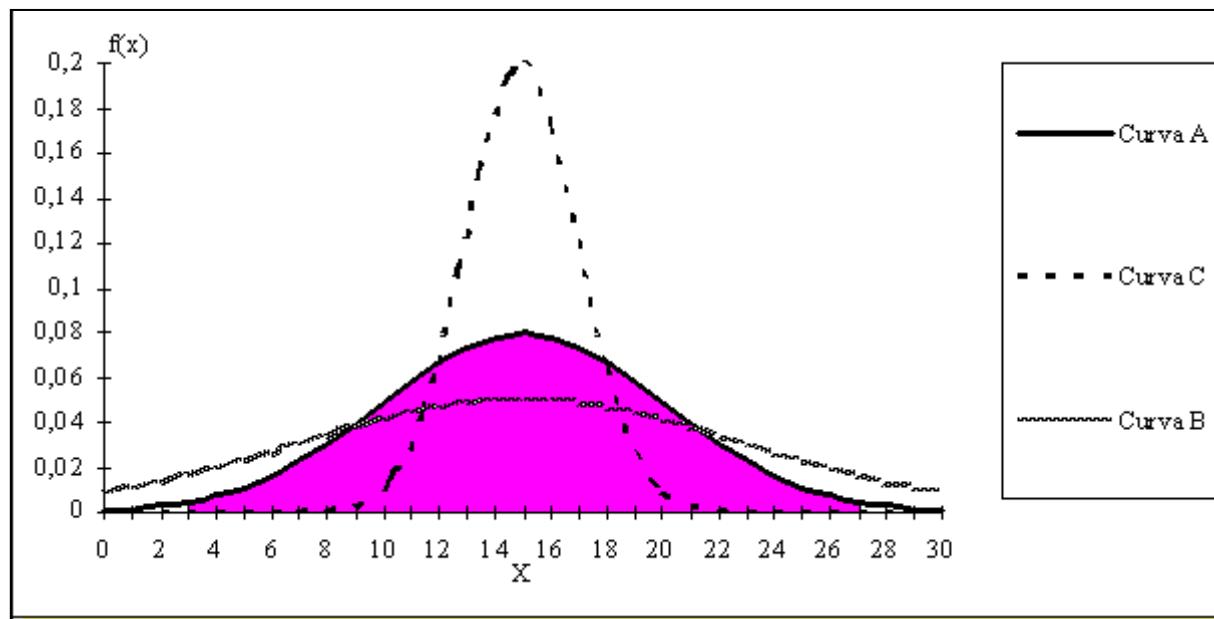
- **La varianza** de las observaciones x_1, x_2, \dots, x_n es el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones.

$$s^2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

• **desviación estándar**

$$s = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Desviaciones estándar



Coeficiente de variación


$$\bullet C.V. = \frac{s}{\bar{x}}$$

- Esta medida es adimensional.
- Sirve para comparar distintas distribuciones
- Ejemplo: Nos preguntamos quién tiene más variabilidad “Las alturas de los elefantes” o “Las alturas de las hormigas”

Coeficiente de variación

$$CV = \frac{S_x}{\bar{X}} \quad Poblacional$$

$$CV = \frac{S_x}{\bar{X}} \quad Muestral$$

- Es adimensional
- Permite efectuar **comparaciones** de distribuciones de distintas poblaciones.
- Ejemplo: Nos permite comparar quién tiene mayor variabilidad ; “Las alturas de los elefantes (m)” o “Las alturas de las hormigas (mm)”
- Nos representa que proporción de la media representa la desviación estándar.

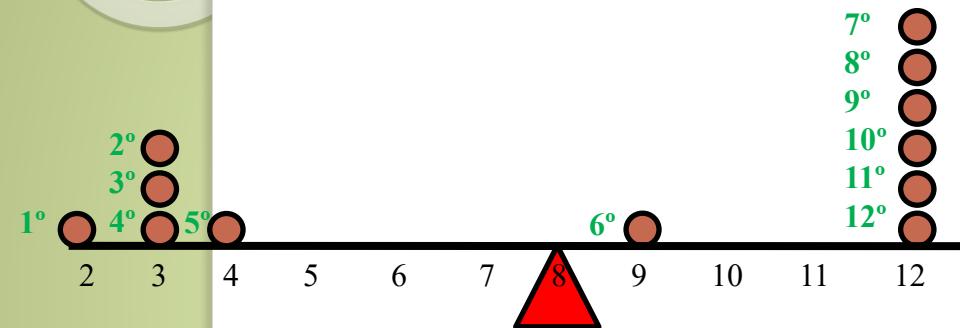
MEDIDAS DE POSICIÓN NO CENTRADAS

 **CUARTILES**

 **DECILES**

 **PERCENTILES**

Cuartiles, Deciles y Percentiles



Orden de las medidas de posición

$$Q_1^{\circ} \quad \frac{12}{4} 1 \quad 3,25 \quad Q_1 \quad 3$$

$$D_6^{\circ} \quad \frac{12}{10} 6 \quad 7,8 \quad D_6 \quad 12$$

$$P_{70}^{\circ} \quad \frac{12}{100} 70 \quad 9,1 \quad P_{70} \quad 12$$

$$Q_k^{\circ} \quad \frac{n-1}{4} k$$

$$D_k^{\circ} \quad \frac{n-1}{10} k$$

$$P_k^{\circ} \quad \frac{n-1}{100} k$$