

Organización de datos y representación gráfica

Tema 3 Análisis de Datos I en Psicología

Distribución de frecuencias

- Ejemplo: estudio de la población inmigrante según la zona de procedencia a partir.
- Población y muestra

Variable X			
Sudamérica	Modalidades		
Norteáfrica			
Centroáfrica			
Este-Europa			
China			
Surasia			

X = Procedencia población de inmigrantes (nominal)

Variable X	N_i		
Sudamérica	24	 Frecuencia absoluta = nº de observaciones en cada categoría	
Norteáfrica			
Centroáfrica			
Este-Europa			
China			
Surasia			

X = Procedencia población de inmigrantes (nominal)

Variable X	N_i		
Sudamérica	24		
Norteáfrica	60	 Frecuencia absoluta = nº de observaciones en cada categoría	
Centroáfrica			
Este-Europa			
China			
Surasia			

X = Procedencia población de inmigrantes (nominal)

Variable X	N_i		
Sudamérica	24	Frecuencia absoluta = n^o de observaciones en cada categoría	
Norteáfrica	60		
Centroáfrica	12		
Este-Europa	18		
China	21		
Surasia	15		
	Total=150		

X = Procedencia población de inmigrantes (nominal)

$$\sum_{n=1}^6 n_i = 150 = n$$

Variable X	N_i	p_i	
Sudamérica	24	$24/150=0.16$	
Norteáfrica	60		
Centroáfrica	12		
Este-Europa	18		
China	21		
Surasia	15		
	Total=150		

Frecuencia
relativa o
proporción



X = Procedencia población de inmigrantes (nominal)

Variable X	N_i	p_i	
Sudamérica	24	$24/150=0.16$	
Norteáfrica	60	$60/150=0.40$	
Centroáfrica	12	$12/150=0.08$	
Este-Europa	18	$18/150=0.12$	
China	21	$21/150=0.14$	
Surasia	15	$15/150=0.10$	
	Total=150	1,00	

X = Procedencia población de inmigrantes (nominal)

Variable X	N_i	p_i	P_i
Sudamérica	24	$24/150=0.16$	16
Norteáfrica	60	$60/150=0.40$	
Centroamérica			.08
Este-Europa	18	$18/150=0.12$	
China	21	$21/150=0.14$	
Surasia	15	$15/150=0.10$	
	Total=150	1,00	

Porcentaje

X = Procedencia población de inmigrantes (nominal)

Variable X	N_i	p_i	P_i
Sudamérica	24	$24/150=0.16$	16
Norteáfrica	60	$60/150=0.40$	40
Centroáfrica	12	$12/150=0.08$	8
Este-Europa	18	$18/150=0.12$	12
China	21	$21/150=0.14$	14
Surasia	15	$15/150=0.10$	10
	Total=150	1,00	100

X = Procedencia población de inmigrantes (nominal)

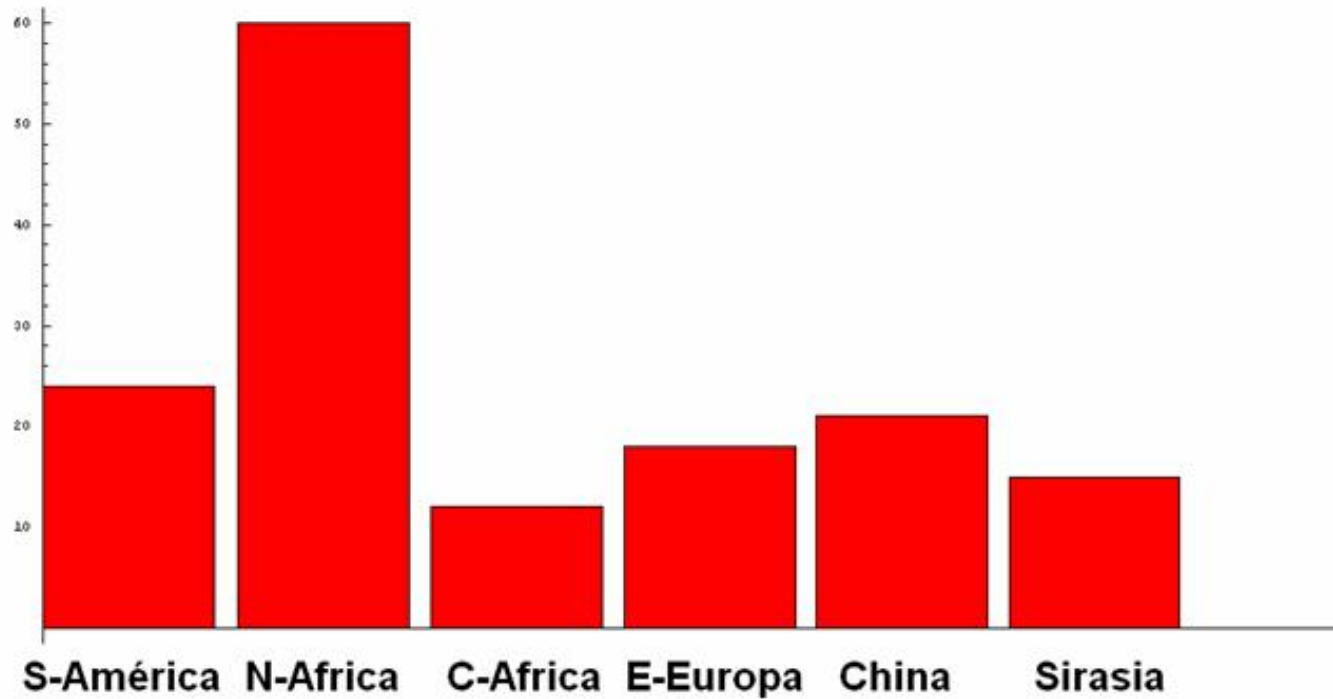
Frecuencia

Diagrama de rectángulos en un orden

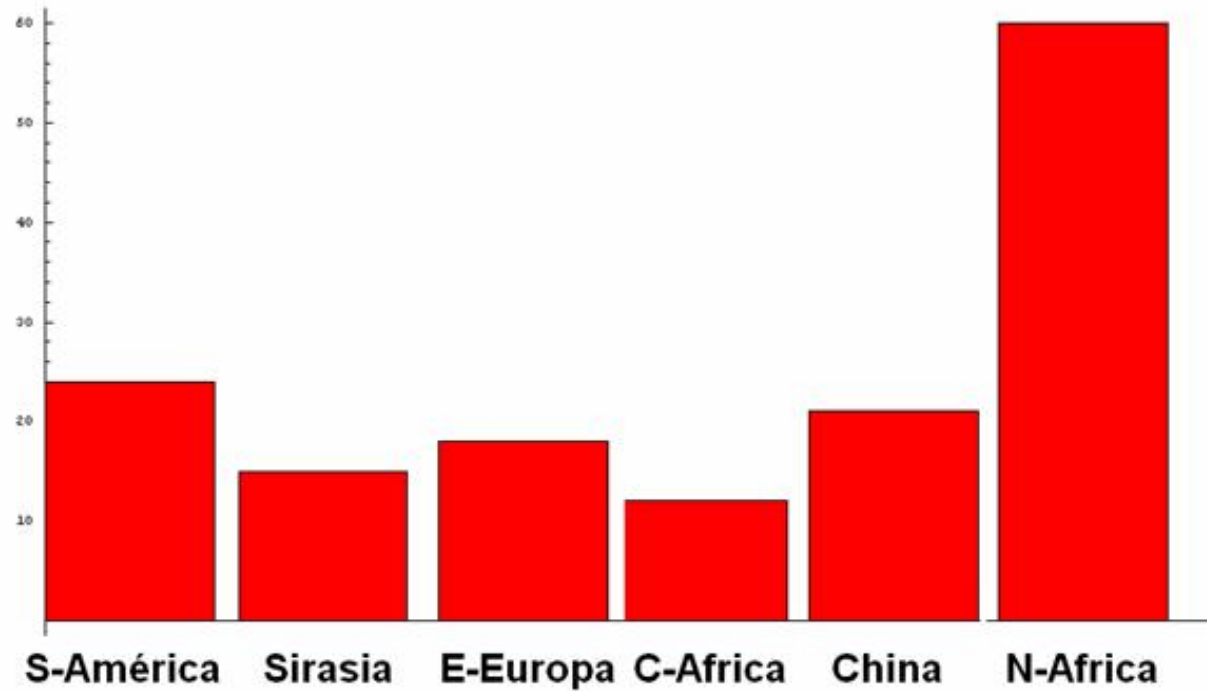
Frecuencia

Diagrama de rectángulos en otro orden

Distribución de frecuencias

- Con las variables ordinales se procede de la misma manera aunque las modalidades situadas en la tabla en un determinado orden.

X	n_i	p_i		
Máxima	12			
Moderada	48			
Mínima	42			
Nula	18			
	$n=120$			

X = Preocupación por las cuestiones políticas (ordinal)

X	n_i	p_i		
Máxima	12	$12/120=0.10$		
Moderada	48			
Mínima	42			
Nula	18			
	$n=120$			

X	n_i	p_i		
Máxima	12	$12/120=0.10$		
Moderada	48	$48/120=0.40$		
Mínima	42	$42/120=0.35$		
Nula	18	$18/120=0.15$		
	$n=120$	1.0		

X	n_i	p_i	n_a	
Máxima	12	$12/120=0.10$		
Moderada	48	$48/120=0.40$		
Mínima	42	$42/120=0.35$		
Nula	18	$18/120=0.15$	18	
	$n=120$	1.0		

X	n_i	p_i	n_a	
Máxima	12	$12/120=0.10$		
Moderada	48	$48/120=0.40$		
Mínima	42	$42/120=0.35$	$18+42=60$	
Nula	18	$18/120=0.15$	18	
	$n=120$	1.0		

X	n_i	p_i	n_a	
Máxima	12	$12/120=0.10$		
Moderada	48	$48/120=0.40$	$18+42+48=108$	
Mínima	42	$42/120=0.35$	$18+42=60$	
Nula	18	$18/120=0.15$	18	
	$n=120$	1.0		

X	n_i	p_i	n_a	
Máxima	12	$12/120=0.10$	$18+42+48+12=120$	
Moderada	48	$48/120=0.40$	$18+42+48=108$	
Mínima	42	$42/120=0.35$	$18+42=60$	
Nula	18	$18/120=0.15$	18	
	$n=120$	1.0		

X	n_i	p_i	n_a	p_a
Máxima	12	$12/120=0.10$	120	
Moderada	48	$48/120=0.40$	108	
Mínima	42	$42/120=0.35$	60	
Nula	18	$18/120=0.15$	18	
	$n=120$	1.0		

X	n_i	p_i	n_a	p_a
Máxima	12	$12/120=0.10$	120	
Moderada	48	$48/120=0.40$	108	
Mínima	42	$42/120=0.35$	60	
Nula	18	$18/120=0.15$	18	$18/120=0.15$
	$n=120$	1.0		

X	n_i	p_i	n_a	p_a
Máxima	12	$12/120=0.10$	120	
Moderada	48	$48/120=0.40$	108	
Mínima	42	$42/120=0.35$	60	$60/120=0.50$
Nula	18	$18/120=0.15$	18	$18/120=0.15$
	$n=120$	1.0		

X	n_i	p_i	n_a	p_a
Máxima	12	$12/120=0.10$	120	
Moderada	48	$48/120=0.40$	108	$108/120=0.90$
Mínima	42	$42/120=0.35$	60	$60/120=0.50$
Nula	18	$18/120=0.15$	18	$18/120=0.15$
	$n=120$	1.0		

X	n_i	p_i	n_a	p_a
Máxima	12	$12/120=0.10$	120	$120/120=1$
Moderada	48	$48/120=0.40$	108	$108/120=0.90$
Mínima	42	$42/120=0.35$	60	$60/120=0.50$
Nula	18	$18/120=0.15$	18	$18/120=0.15$
	$n=120$	1.0		

Frecuencia absoluta

- nº de veces que se repite cada uno de los valores de la variable (n_i)
- La suma de todas las frecuencias absolutas representa el total de la muestra (n)

Frecuencia relativa o proporción

- Cociente entre la frecuencia absoluta de cada clase y el número total de observaciones

$$p_i = \frac{n_i}{n}$$

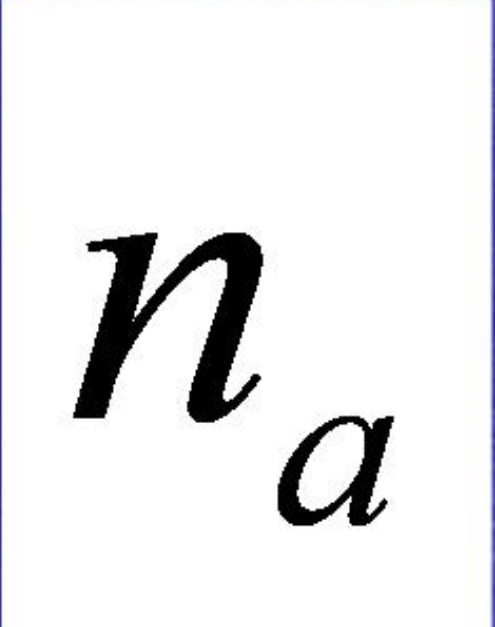
Porcentaje

- Valor de la frecuencia relativa multiplicado por 100

$$P_i = p_i \cdot 100$$

Frecuencia absoluta acumulada

- nº de veces que se repite cada modalidad o cualquiera de las modalidades anteriores


$$n_a$$

Proporción acumulada

- Cociente entre la frecuencia acumulada de cada clase y el total de observaciones.

$$p_a = \frac{n_a}{n}$$

Variable cuantitativa

X_i	n_i	p_i	n_a	p_a
63- 70	2	0,02	2	0,02
71- 78	8	0,08	10	0,10
79- 86	8	0,08	18	0,18
87- 94	19	0,19	37	0,37
95-102	26	0,26	63	0,63
102-110	17	0,17	80	0,80
111-118	9	0,09	89	0,89
119-126	7	0,07	96	0,96
127-134	4	0,04	100	1,00
	100	1,00		

Número de intervalos

- Glass y Stanley recomiendan 12-15 intervalos
- Botella, León y San Martín aconsejan un máximo de 9 intervalos

Número de intervalos

- Criterios cuantitativos
 - no mayor que la raíz cuadrada de n
 - Fórmula de Sturges:
Parte Entera ($1,5 + 3,3 \log[n]$)

- **Valor informado:** valor que se lee en el instrumento de medida, que tiene una determinada precisión.
- **Límites exactos de medida:** los valores que acotan el intervalo en el que se encuentra el valor real.

- La distribución de frecuencias de variables cuantitativas debe realizarse teniendo en cuenta los límites exactos de medida, no los valores informados.

Límites exactos = Valor informado $\pm (0,5 \times I)$ siendo I la unidad del instrumento de medida (precisión).

- **Rango** de la distribución
 $\text{Rango} = \text{Valor máximo} - \text{Valor mínimo} + 1$
 $\text{Rango} = \text{número de intervalos} \times I$
- **Intervalo**: sinónimo del concepto de modalidad, es cada uno de los grupos de valores que ocupan una fila en una distribución de frecuencias.

- Límites aparentes (virtuales o informados) son los valores mayor y menor de cada intervalo, teniendo en cuenta el nivel de precisión del instrumento de medida.

- **Límites reales o exactos**: son los valores máximo y mínimo que tendría cada intervalo si el instrumento de medida tuviera una precisión perfecta.
- **Punto medio del intervalo**: es la semisuma de los límites exactos o de los límites aparentes.

- **Amplitud del intervalo:** es la diferencia entre el límite exacto superior y el límite exacto inferior, o entre el límite aparente superior y el límite aparente inferior más la unidad de medida.

Distribución de valores dentro de un intervalo

Supuestos (excluyentes):

- La distribución de valores en un intervalo es homogénea.

Distribución de valores dentro de un intervalo

Supuestos (excluyentes):

- Los valores se concentran en el punto medio del intervalo y este punto medio es la única información que vamos a tener de dicho intervalo. En este caso se puede producir el error de agrupamiento al perderse la información de cada puntuación concreta.

Representaciones gráficas

- El tipo de representación gráfica depende de la variable objeto de estudio.
- **Variables cualitativas**: la información que se registra es la frecuencia de observaciones en cada categoría o modalidad. Se utilizan el diagrama de rectángulos y el pictograma (diagrama de sectores).

Diagrama de rectángulos

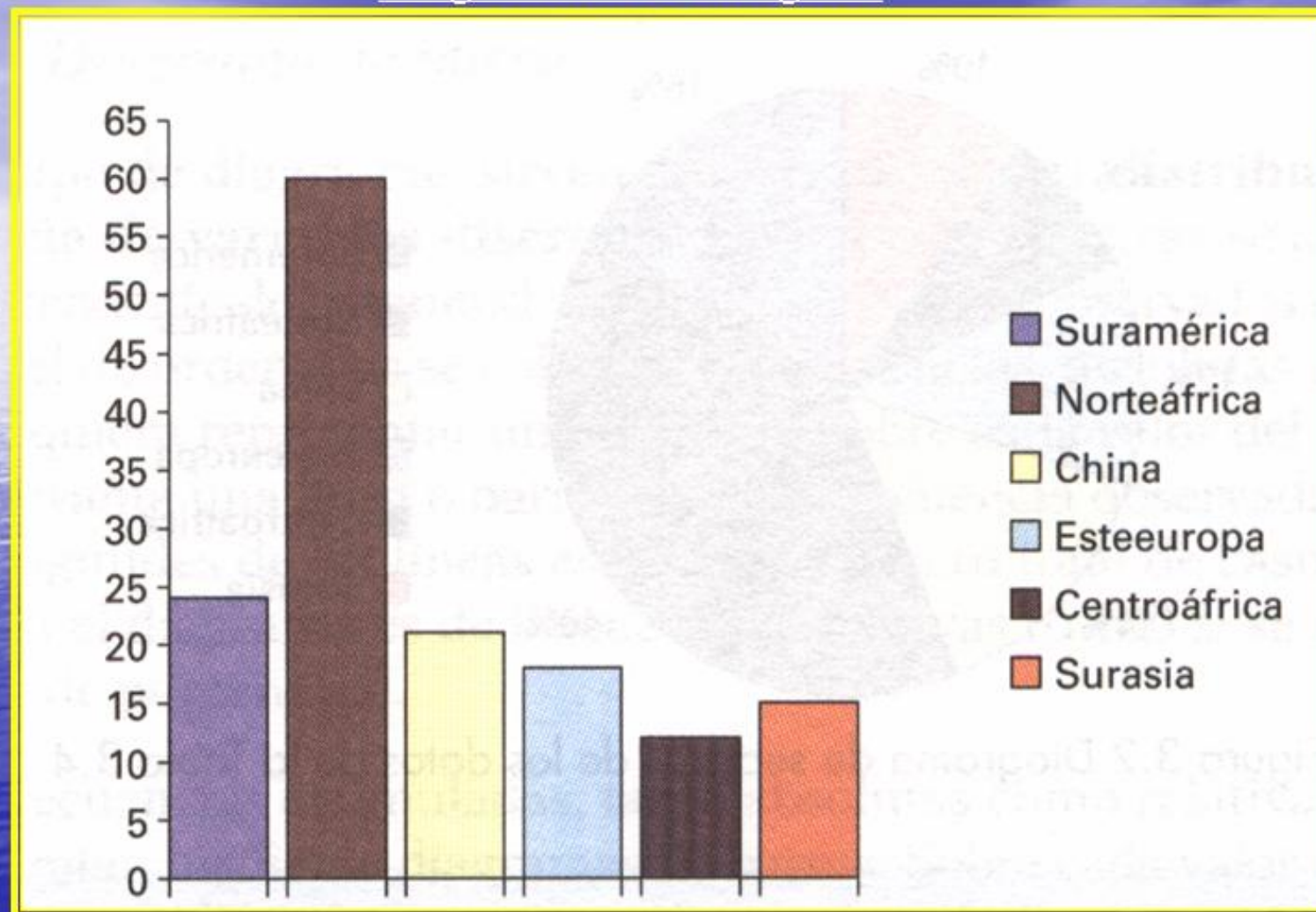
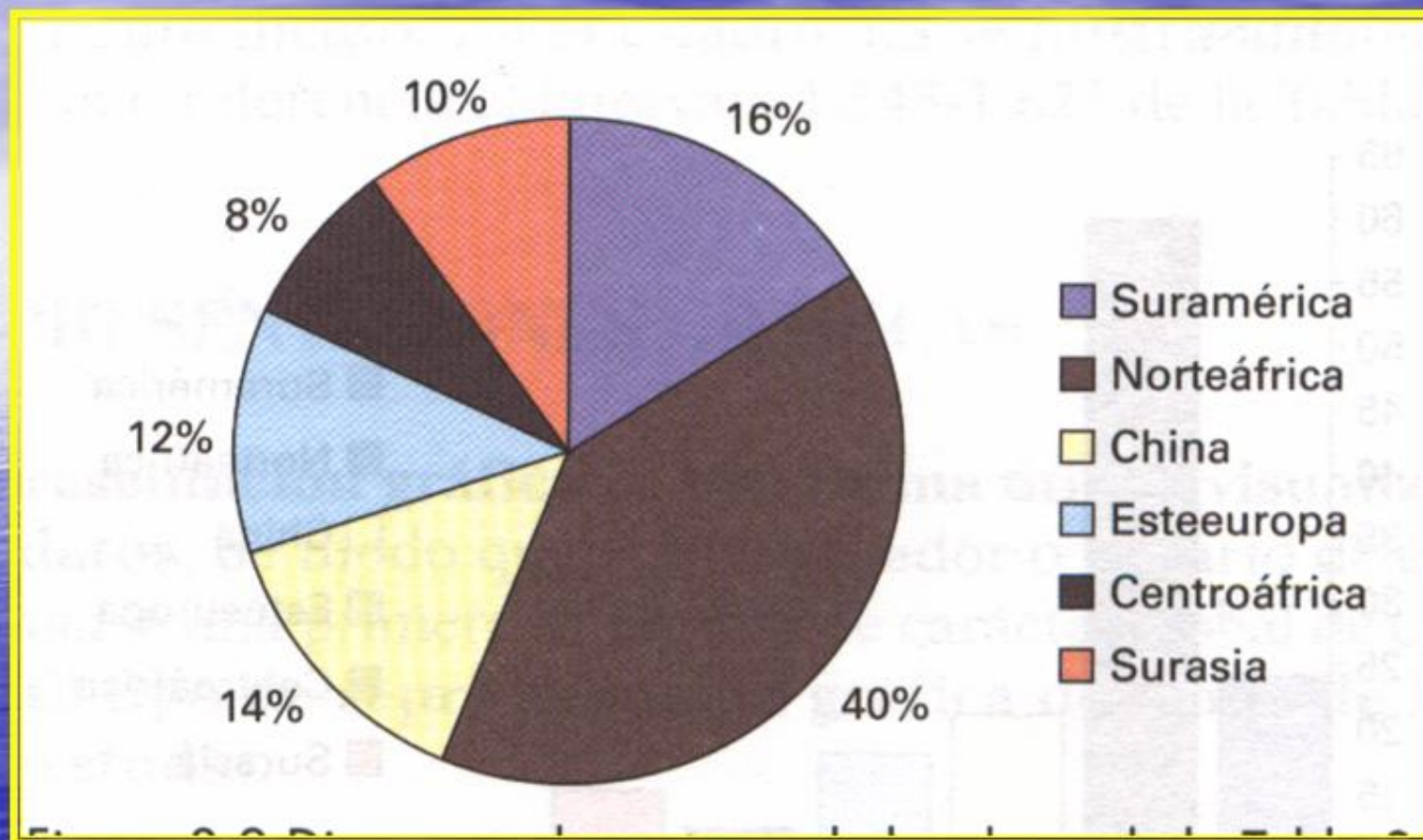
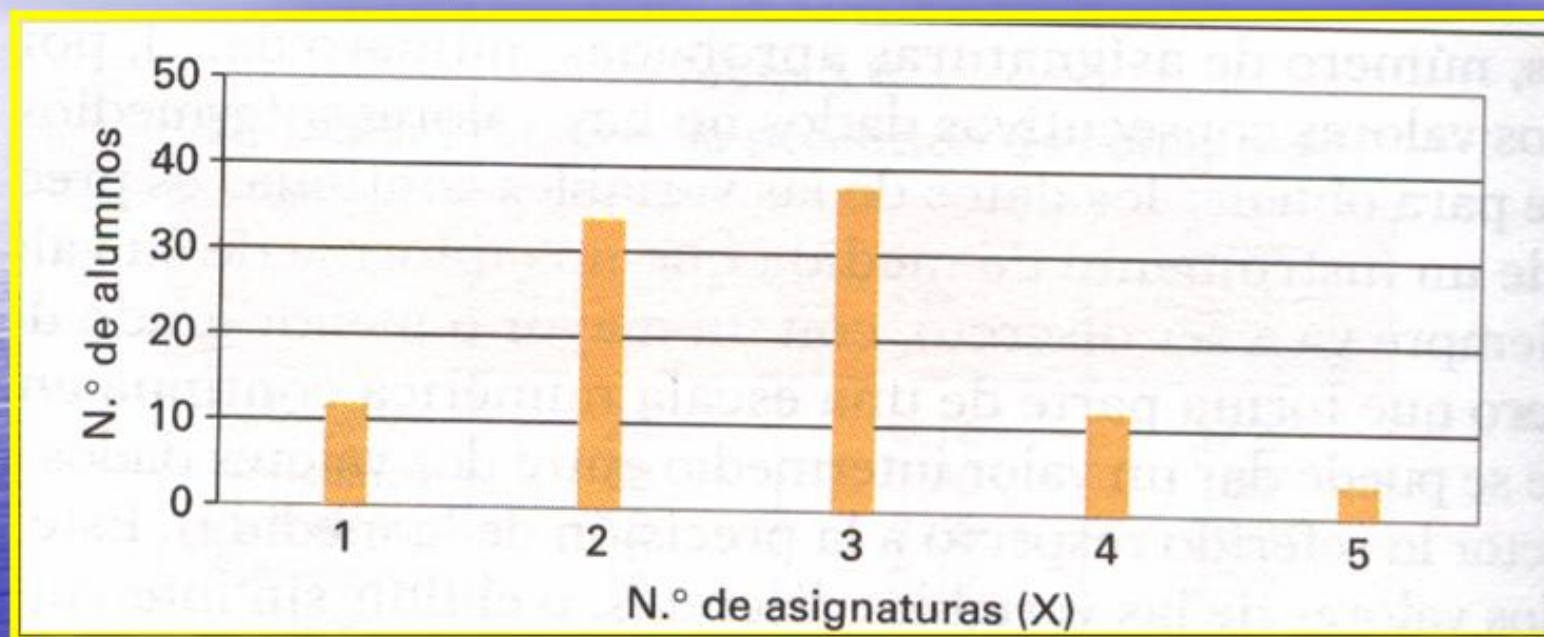


Diagrama de sectores



Variables cuantitativas

Diagrama de barras (variables discretas -ordinal)



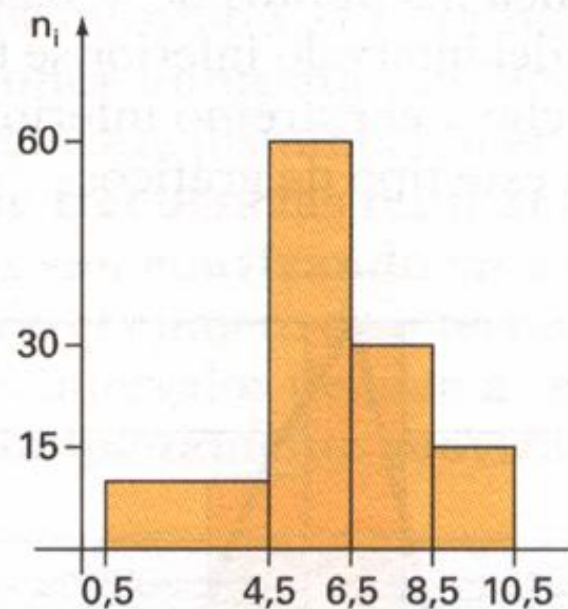
N.º de asignaturas superadas

Variables cuantitativas

Histograma
(variables continuas agrupadas en intervalos)

Tabla 3.6.

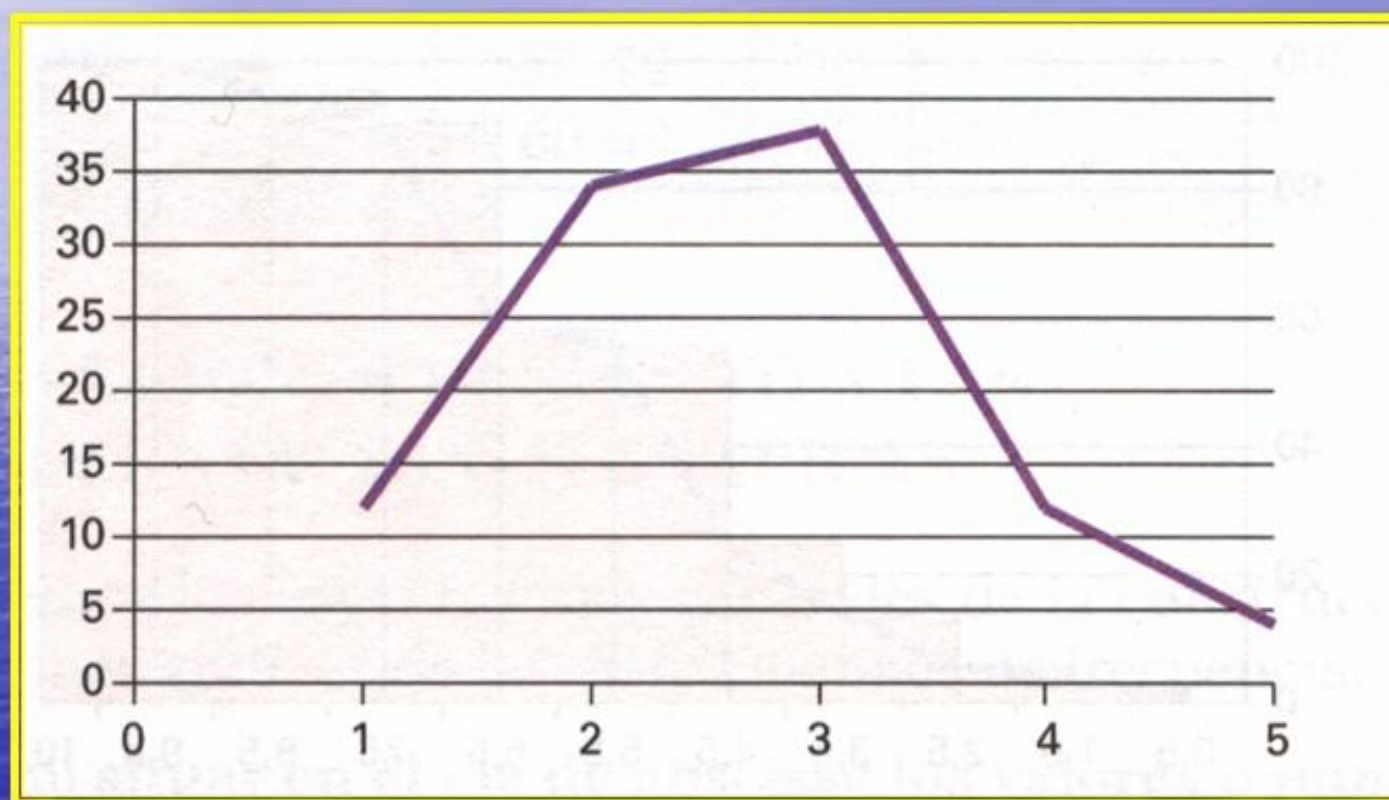
X	n_i
9-10	12
7-8	30
5-6	60
1-4	8
	110



Variables cuantitativas

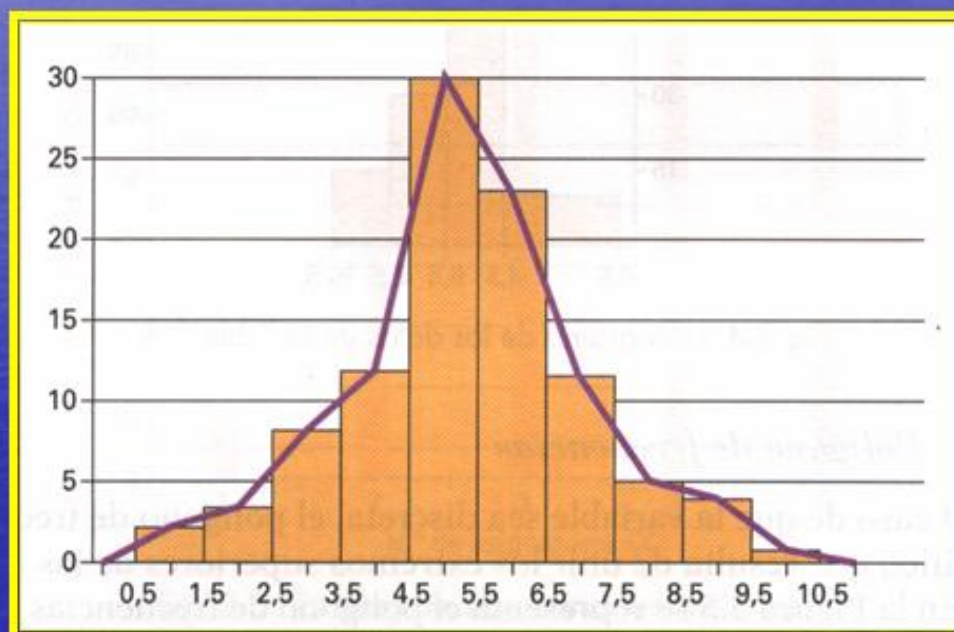
Polígono de frecuencias

(variables discreta: se unen los extremos superiores de las barras o líneas)



Variables cuantitativas

Polígono de frecuencias
(variables continua: se unen los puntos medios de la base superior de los rectángulos del histograma)



Convenciones respecto a las representaciones gráficas

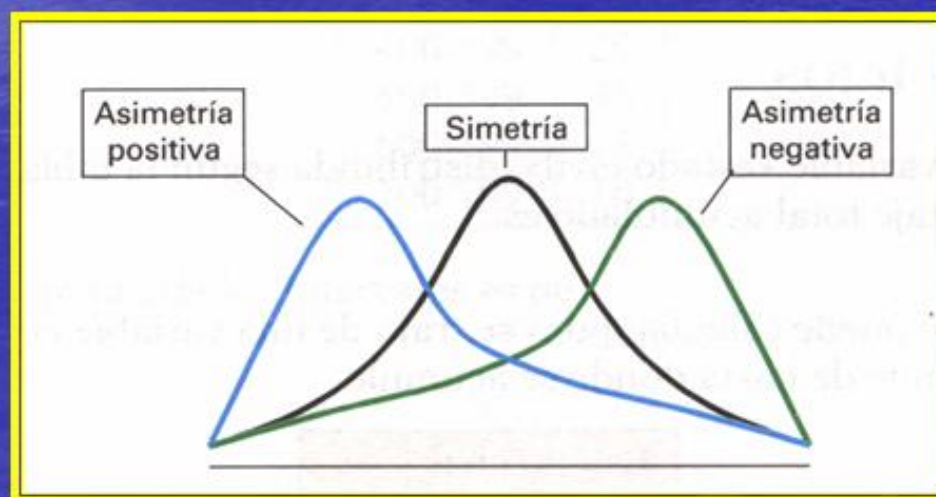
- Es usual situar en el eje de abscisas los valores o puntuaciones de la variable y en el de ordenadas las frecuencias.
- Los valores de la variable se situarán en orden ascendente de magnitud, de izquierda a derecha

Convenciones respecto a las representaciones gráficas

- Si los valores mínimos son muy altos, es usual hacer unos cortes en los ejes.
- Es necesario incorporar al gráfico la información necesaria para su correcta comprensión.

Propiedades de las distribuciones de frecuencias

- Tendencia central
- Variabilidad
- Sesgo o asimetría



Propiedades de las distribuciones de frecuencias

- Curtosis (apuntamiento)

