

TEMA 1. ESTADÍSTICA DESCRIPTIVA

- 1.1 Introducción: conceptos básicos
- 1.2 Tablas estadísticas y representaciones gráficas
- 1.3 Características de variables estadísticas unidimensionales
 - 1.3.1 Características de posición
 - 1.3.2 Características de dispersión
 - 1.3.3 Características de forma
- 1.4 Concepto de v.e. bidimensional
- 1.5 Distribuciones marginales y condicionadas
- 1.6 Covarianza
- 1.7 Dependencia e independencia estadística
- 1.8 Regresión y correlación. Introducción
- 1.9 Rectas de regresión
- 1.10 Coeficiente de determinación y coeficiente de correlación lineal
- 1.11 Otros tipos de ajuste

❖ 1.1. Introducción : conceptos básicos

➤ **ESTADÍSTICA:** “Estudio de los métodos de recogida y descripción de datos, así como del análisis de esta información”

❖ Etapas de un estudio estadístico

- 1 Recogida de datos
- 2 Ordenación, tabulación y gráficos*
- 3 Descripción de características*
- 4 Análisis formal

* Estadística descriptiva: parte de la estadística que se ocupa de las etapas 2 y 3

❖ Individuo, Población, Muestra

- **Población:** “Conjunto de elementos a los que se les estudia una característica”
- **Individuo:** “Cada uno de los elementos de la población”
- **Muestra:** “Subconjunto representativo de la población”

❖ Variables estadísticas. Modalidades

➤ **Variable estadística (v.e.):** "Característica propia del individuo objeto del estudio estadístico"

Ejemplos:

- Estatura
- Peso
- Color del pelo
- Nivel de colesterol
- N° de hijos de una familia

➤ **Modalidad:** "Cada una de las posibilidades o estados diferentes de una variable estadística"

➤ Exhaustivas e incompatibles

Ejemplo:

color del pelo:

- castaño
- rubio
- negro

❖ Tipos de variables estadísticas

➤ **Cualitativas:** Las características no son cuantificables

Ejemplos:

Profesión

Color del pelo

➤ **Cuantitativas:** Características cuantificables o numéricas

✓ **Discretas:** Numéricas numerables

Ejemplos:

Nº de hijos

Nº de viviendas

✓ **Continuas:** Numéricas no numerables

Ejemplos:

Talla

Peso

Nivel de colesterol

❖ 1.2. Tablas estadísticas y representaciones gráficas

➤ Variables discretas

✓ Frecuencias

- ◆ Absolutas, n_i (nº individuos modalidad i)
- ◆ Absolutas acumuladas, $N_i = n_1 + n_2 + \dots + n_i$
- ◆ Relativas, f_i (proporción indiv. modalidad i)
- ◆ Relativas acumuladas, $F_i = f_1 + f_2 + \dots + f_i$

x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
...
x_i	n_i	N_i	f_i	F_i
...
x_k	n_k	N_k	f_k	F_k
	n		1	

Absolutas, n_i

Absolutas acumuladas, N_i

Relativas
 $f_i = n_i / n$

Relativas acumuladas
 $F_i = N_i / n$

➤ **Variables continuas: Intervalos**

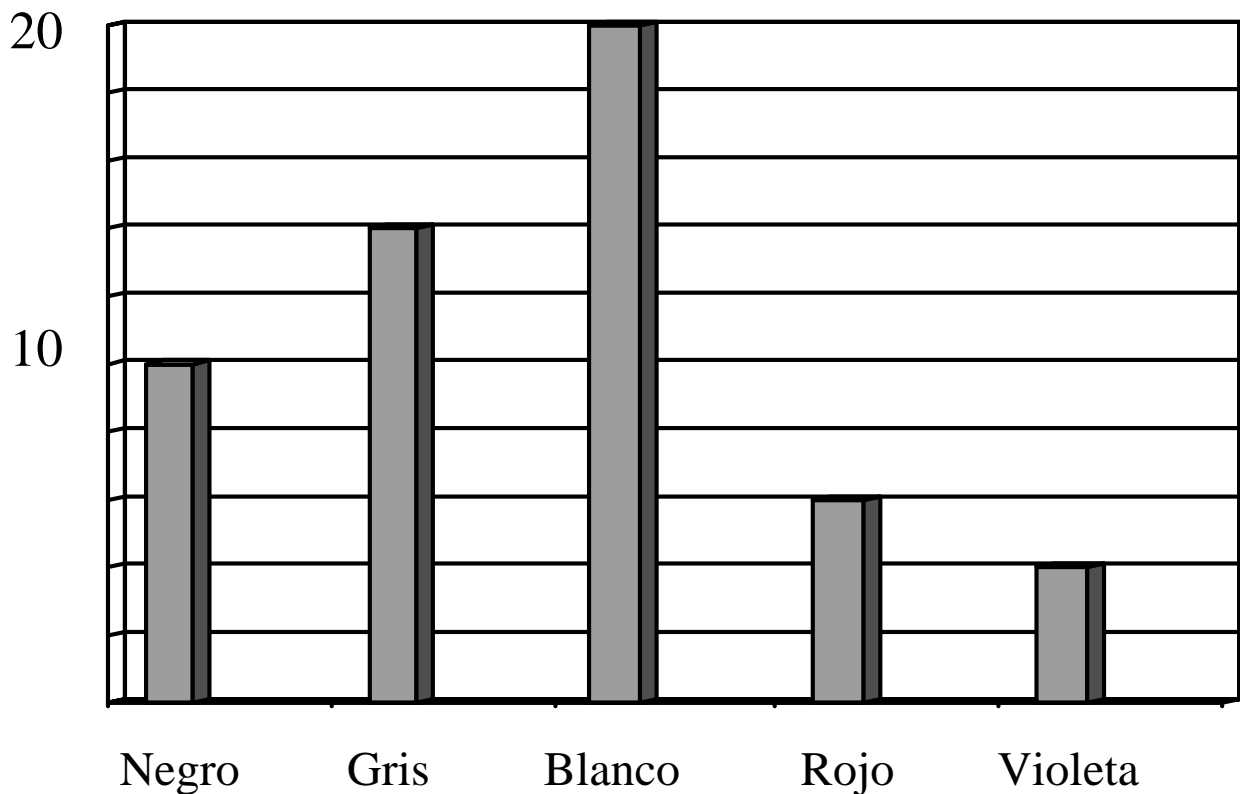
<i>Intervalo I_i</i>	x_i	n_i	N_i	f_i	F_i
$e_0 - e_1$	x_1	n_1	N_1	f_1	F_1
...
$e_{i-1} - e_i$	x_i	n_i	N_i	f_i	F_i
...
$e_{k-1} - e_k$	x_k	n_k	N_k	f_k	F_k
		n		1	

- Marca de clase x_i (punto medio de cada intervalo)
- Amplitud a_i (distancia entre los extremos)
- Extremos [...)

❖ Gráficos estadísticos

➤ V. e. Cualitativas: Gráfico rectangular

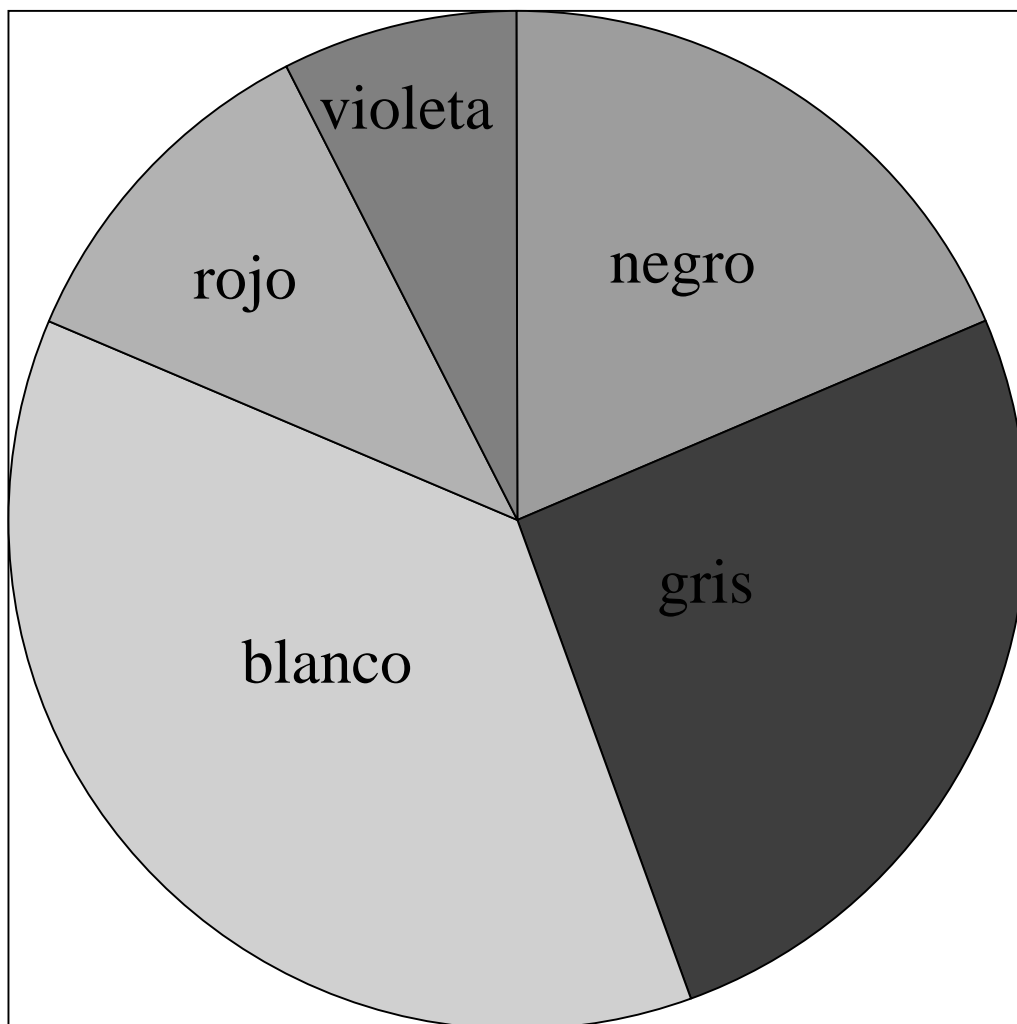
Color Plumaje	Nº de Aves (n_i)
Negro	10
Gris	14
Blanco	20
Rojo	6
Violeta	4



➤ **V. e. Cualitativas: Gráfico de sectores**

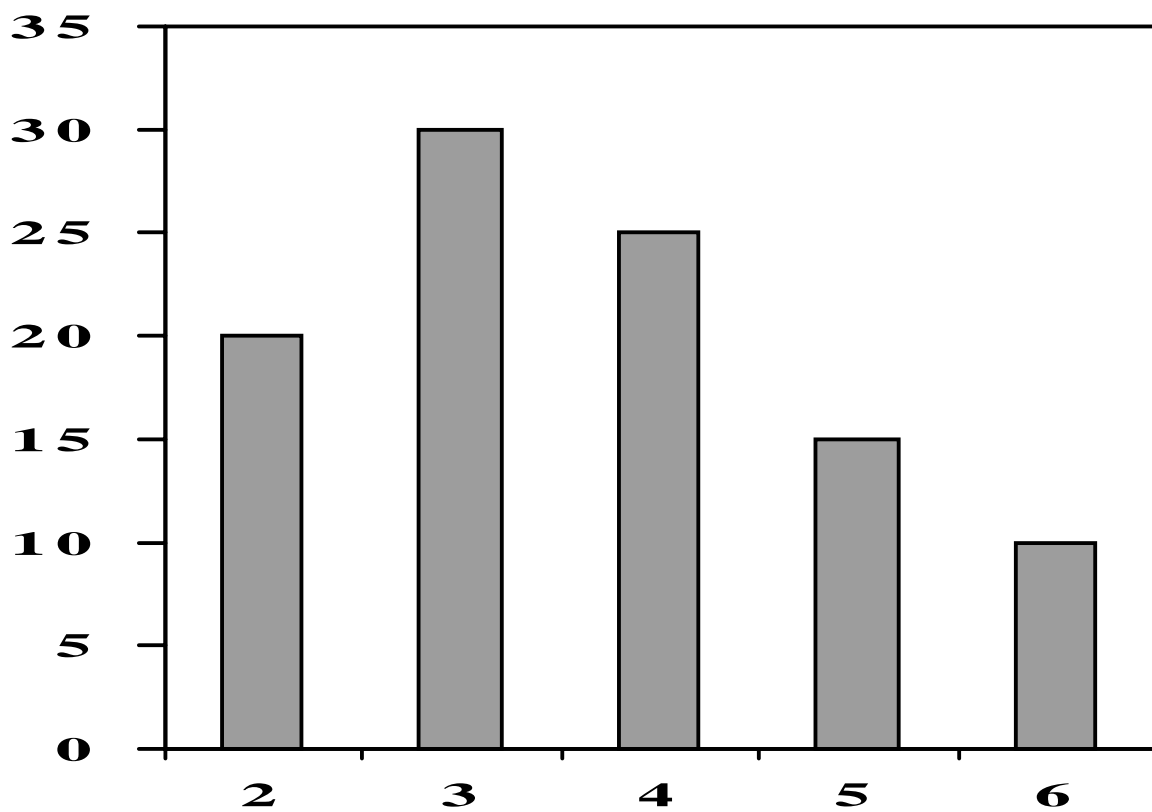
Color Plumaje	Nº de Aves (n_i)
Negro	10
Gris	14
Blanco	20
Rojo	6
Violeta	4

Grados
de cada
sector = $360^\circ f_i$



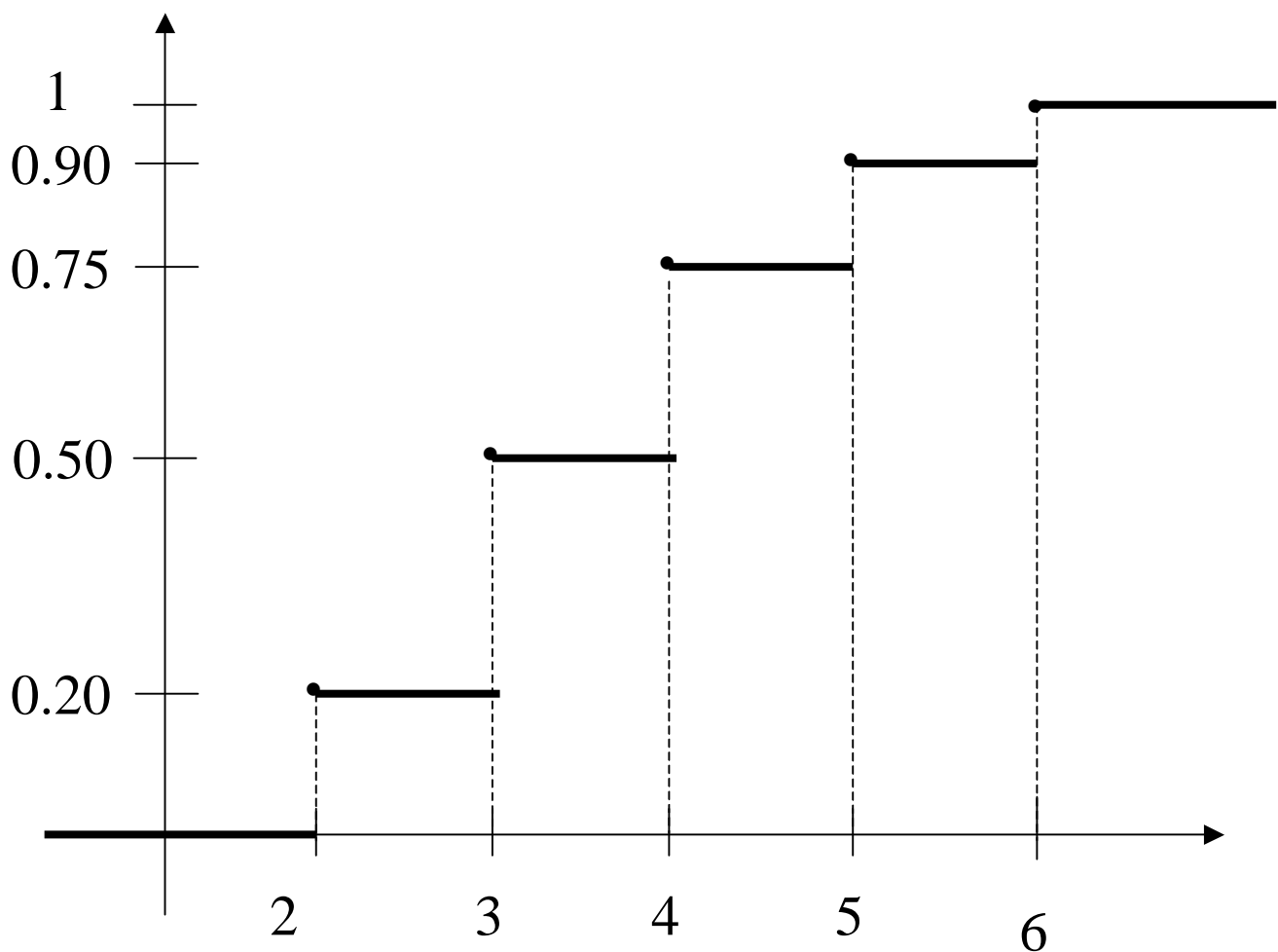
➤ **V. e. Discretas: Gráfico de barras**

Nº de crías	Nº animales: n_i	f_i	F_i
2	20	0.20	0.20
3	30	0.30	0.50
4	25	0.25	0.75
5	15	0.15	0.90
6	10	0.10	1
	$n = 100$		



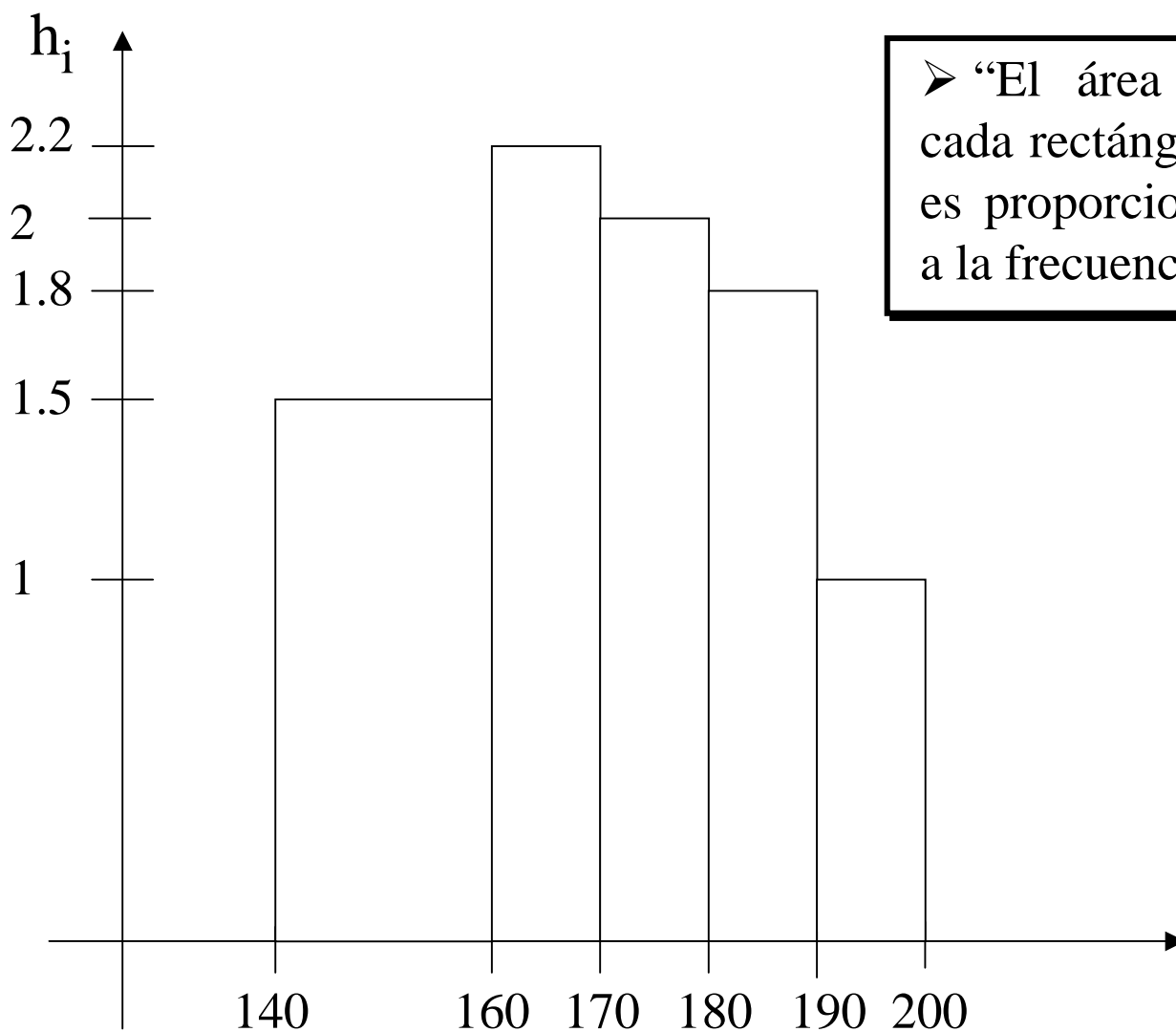
➤ **V. e. Discretas: Curva acumulativa de distribución**

Nº de crías	Nº animales: n_i	f_i	F_i
2	20	0.20	0.20
3	30	0.30	0.50
4	25	0.25	0.75
5	15	0.15	0.90
6	10	0.10	1
	$n = 100$		



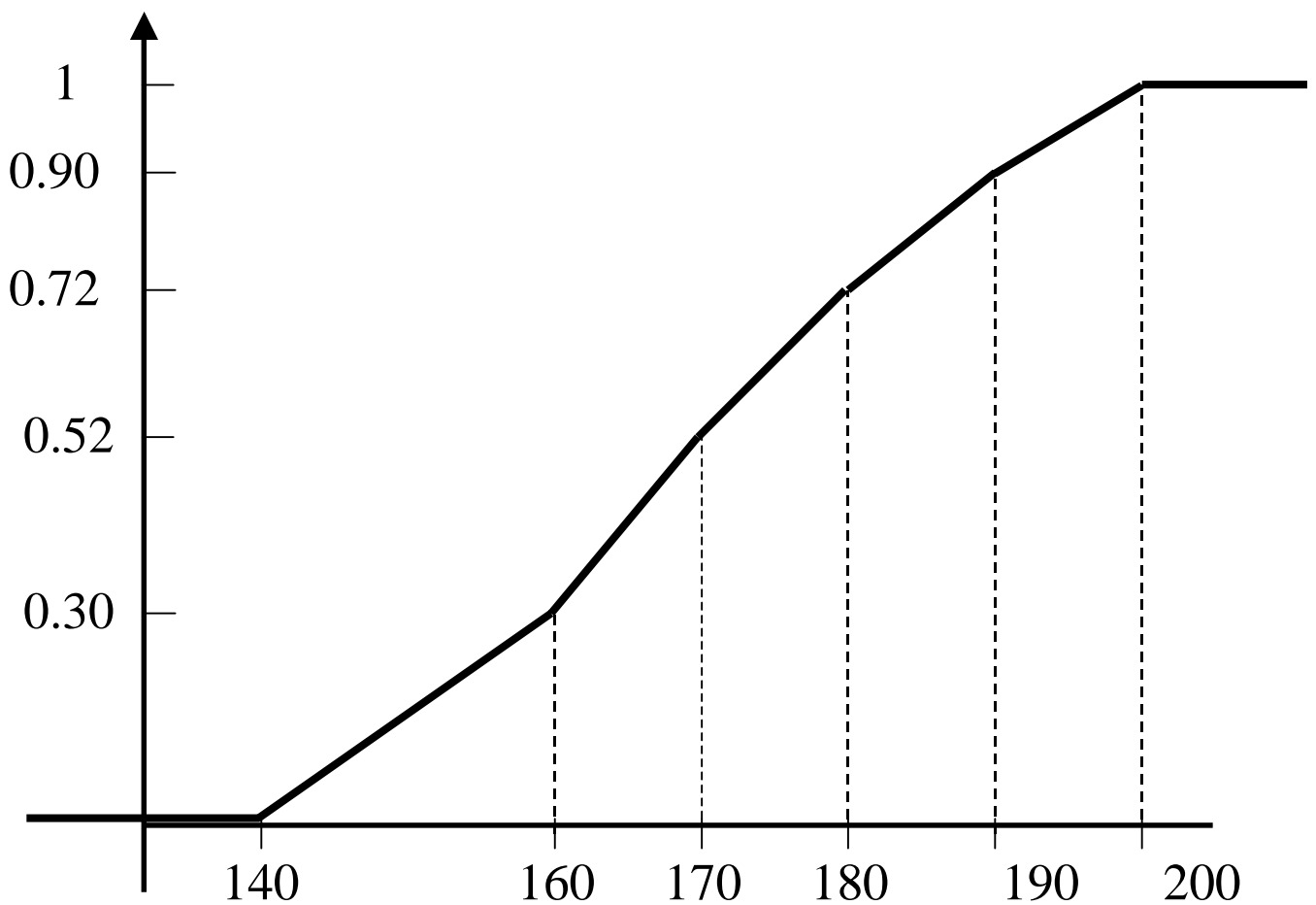
➤ **V. e. Continuas: Histograma**

Estatura	n_i	$h_i = n_i / a_i$
140-160	30	1.5
160-170	22	2.2
170-180	20	2
180-190	18	1.8
190-200	10	1
	100	



➤ **V. e. Continuas: Curva
acumulativa de distribución**

Talla	n_i	f_i	F_i
140-160	30	0.30	0.30
160-170	22	0.22	0.52
170-180	20	0.20	0.72
180-190	18	0.18	0.90
190-200	10	0.10	1
	100		



❖ **1.3. Características de variables estadísticas unidimensionales**

◆ **1.3.1 Características de Posición**

❖ **Media aritmética**

$$\bar{x} = \sum_{i=1}^k f_i x_i = \frac{\sum_{i=1}^k n_i x_i}{n}$$

Estatura	Nº Personas n_i	M. Clase x_i	$n_i x_i$
140-150	20	145	2900
150-160	100	155	15500
160-180	80	170	13600
180-200	10	190	1900
	$n = 210$		33900

$$\text{Media: } \bar{x} = \frac{\sum_{i=1}^k n_i x_i}{n} = \frac{33900}{210} = 161.42$$

❖ Moda

- ❑ Valor de la variable más frecuente
- ✓ Puede haber más de una moda → Plurimodal

➤ Variables discretas

- Datos en serie

2, 2, 3, 3, 3, 3, 5, 6, 7 $Mo = 3$

- Datos en tabla

◆ Ejemplo



x_i	n_i
1	34
2	36
3	45
4	22
5	17

$Mo = 3$

➤ Variables continuas

$$Mo = e_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} a_i$$

◆ Ejemplo



x_i	n_i	$h_i = n_i / a_i$
140-160	30	1.5
160-170	22	2.2
170-180	20	2
180-190	18	1.8
190-200	10	1
	100	

$$Mo = 160 + \frac{(2.2 - 1.5)}{(2.2 - 1.5) + (2.2 - 2)} \times 10 = 167.777$$

➤ Observaciones:

1. Puede utilizarse la frecuencia relativa
2. Si las amplitudes son iguales se puede proceder directamente con las frecuencias

❖ Mediana

□ Valor de la variable que ocupa el lugar central en una serie de datos ordenados.

▪ El 50% de los elementos de la población tienen un valor de la variable menor de la mediana. El 50% de los elementos de la población tienen un valor de la variable mayor.

➤ Variables discretas

▪ Datos en serie

◆ Ejemplos

▪ N° impar de observaciones:

$$2, 2, 2, 3, \textcircled{5}, 6, 7, 7, 8 \rightarrow \text{Me} = 5$$

▪ N° par de observaciones:

$$4, 6, 6, 6, | 7, 8, 8, 9 \rightarrow \text{Me} = 6 - 7$$

Indeterminado entre 6 y 7

➤ **Variables discretas**

▪ **Datos en tabla**

◆ **Ejemplo**

x_i	n_i	N_i	f_i	F_i
0	4	4	0.142	0.142
1	6	10	0.214	0.357
2	10	20	0.357	0.714
3	5	25	0.178	0.892
4	3	28	0.107	1
	28		1	

$n/2 = 14$
 $F_i = 1/2$



$Me = 2$

➤ **Observación:** Si $n/2$ coincide con un N_i



la mediana está indeterminada entre x_i y x_{i+1}

➤ **Variables continuas**

$$Me = e_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} (e_i - e_{i-1}) + \frac{\frac{n}{2} - N_{i-1}}{n_i} (e_i - e_{i-1})$$

◆ **Ejemplo**

Tallas	n_i	N_i	f_i	F_i
140-150	15	15	0.15	0.15
150-160	30	45	0.30	0.45
→ 160-170	25	70	0.25	0.70
170-180	20	90	0.20	0.90
180-200	10	100	0.10	1
	100			

$n/2 = 50$
 $F_i = 1/2$

$$Me = 160 + \frac{0.5 - 0.45}{0.25} \times 10 = 160 + 2 = 162$$

➤ **Observación:** Si $n/2$ coincide con un N_i ↩

la mediana es el extremo superior del intervalo que le corresponde

❖ Percentiles

□ Definición: P_k , $k:1,2,\dots,99$, “percentil k ”, valor de la variable que deja por debajo, el $k\%$ de los valores de la variable

$$\begin{aligned} Q_1 &= P_{25} \rightarrow \text{Cuartil } 1^\circ \\ Q_2 &= P_{50} \rightarrow \text{Cuartil } 2^\circ = Me \\ Q_3 &= P_{75} \rightarrow \text{Cuartil } 3^\circ \end{aligned}$$

Cuantiles, aún más general

$$\begin{aligned} D_1 &= P_{10} \rightarrow \text{Decil } 1^\circ \\ D_2 &= P_{20} \rightarrow \text{Decil } 2^\circ \\ &\dots \\ D_9 &= P_{90} \rightarrow \text{Decil } 9^\circ \end{aligned}$$

■ Cálculo para v.e. discretas:

Igual que la mediana, cambiando $n/2$ por $nk/100$

■ Cálculo para v.e. continuas:

$$P_k = e_{i-1} + \frac{\frac{nk}{100} - N_{i-1}}{n_i} a_i = e_{i-1} + \frac{\frac{k}{100} - F_{i-1}}{f_i} a_i$$

◆ Ejemplos percentiles v.e. discreta

x_i	n_i	N_i
2	20	20
3	30	50
4	44	94
5	20	114
6	10	124
	124	

$nk/100 = 124 \times 40 / 100 = 49.6$

$nk/100 = 124 \times 95 / 100 = 117.8$

Percentil 40, $P_{40} = 3$

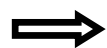
Percentil 95, $P_{95} = 6$

$$nk/100 = 124 \times 25 / 100 = 31$$



Percentil 25, $P_{25} = 3 = Q_1$

$$nk/100 = 124 \times 50 / 100 = 62$$



Percentil 50, $P_{50} = 4 = Me = Q_2$

$$nk/100 = 124 \times 75 / 100 = 93$$



Percentil 75, $P_{75} = 4 = Q_3$

◆ Ejemplos percentiles v.e. continua

Tallas	n_i	N_i	f_i	F_i
140-150	15	15	0.15	0.15
150-160	30	45	0.30	0.45
160-170	25	70	0.25	0.70
170-180	20	90	0.20	0.90
180-200	10	100	0.10	1
	100			

→ } ← P_{40}
 → } ← P_{75}

$$P_k = e_{i-1} + \frac{\frac{nk}{100} - N_{i-1}}{n_i} a_i = e_{i-1} + \frac{\frac{k}{100} - F_{i-1}}{f_i} a_i$$

$$P_{40} = 150 + \frac{40 - 15}{30} \times 10 = 150 + \frac{0.4 - 0.15}{0.30} \times 10 = 158.33$$

$$P_{75} = 170 + \frac{75 - 70}{20} \times 10 = 170 + \frac{0.75 - 0.70}{0.20} \times 10 = 172.5 = Q_3$$

◆ 1.3.2. Características de Dispersión

✓ “Miden la Homogeneidad de las observaciones”

❖ Rango o recorrido

➤ Valor máximo menos valor mínimo de la variable

❖ Recorrido intercuartílico

➤ $Q_3 - Q_1$

❖ Varianza

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

❖ Desviación típica

$$\sigma = \sqrt{\sigma^2}$$

❖ Coeficiente de variación

$$C.V. = \frac{\sigma}{\bar{x}}$$

◆ Ejemplo

x_i	n_i	$n_i x_i$	$n_i x_i^2$
4	20	80	320
6	40	240	1440
8	44	352	2816
10	36	360	3600
12	22	264	3168
	162	1296	11344

$$\sigma^2 = \text{Var}[X] = \frac{\sum_{i=1}^k n_i x_i^2}{n} - \bar{x}^2 = \frac{11344}{162} - \left(\frac{1296}{162}\right)^2 = 6.02$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{6.02} = 2.4535$$

❖ **Momentos no centrales (Respecto al origen)**

$$m_r = \frac{\sum_{i=1}^k n_i x_i^r}{n} = \frac{\sum_{i=1}^k n_i x_i^r}{\sum_{i=1}^k n_i}$$

$$r=1 \rightarrow m_1 = \sum_{i=1}^k f_i x_i = \frac{\sum_{i=1}^k n_i x_i}{n} = \bar{x}$$

$$r=2 \rightarrow m_2 = \sum_{i=1}^k f_i x_i^2 = \frac{\sum_{i=1}^k n_i x_i^2}{n}$$

$$s^2 = \frac{\sum_{i=1}^k n_i x_i^2}{n} - \bar{x}^2 = m_2 - (\bar{x})^2$$

❖ Momentos centrales (Respecto a la media)

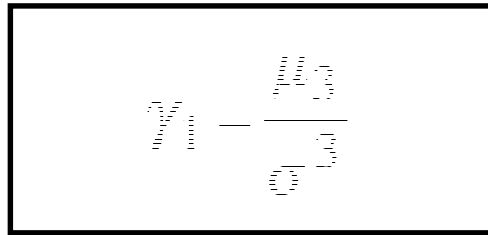
$$\mu_r = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^r}{n}$$

$$r = 1 \rightarrow \mu_1 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})}{n} = 0$$

$$r = 2 \rightarrow \mu_2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{n} = \sigma^2$$

◆ 1.3.3 Características de forma

❖ Coeficiente de Sesgo (Asimetría)



► Si $\gamma_1 = 0 \Rightarrow$ Distribución simétrica

► Si $\gamma_1 > 0 \Rightarrow$ Distribución sesgada a la derecha

► Si $\gamma_1 < 0 \Rightarrow$ Distribución sesgada a la izquierda

❖ Coeficiente de Curtosis (Aplastamiento)

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

► Si $\gamma_2 = 0 \Rightarrow$ Distribución igual de aplastada que la distribución Normal

► Si $\gamma_2 > 0 \Rightarrow$ Distribución menos aplastada que la distribución Normal

► Si $\gamma_2 < 0 \Rightarrow$ Distribución más aplastada que la distribución Normal

❖ 1.4 Concepto de variable estadística bidimensional

◆ Ejemplo . X : “Peso”, Y : “Estatura”

$X \backslash Y$	140-160	160-180	180-200	>200	Marginal X
40-60	10	6	2	0	18
60-80	8	12	6	2	28
80-100	1	8	10	6	25
Marginal Y	19	26	18	8	71

✓ Frecuencias Marginales
 Frecuencias Marginales de X
 Frecuencias Marginales de Y

✓ Frecuencias Condicionadas
 Frecuencias Condicionadas de X
 Frecuencias Condicionadas de Y

❖ 1.5 Distribuciones marginales y condicionadas

➤ Distribución marginal de X

◆ Distribución de la variable X : “Peso”

$X \setminus Y$	140-160	160-180	180-200	>200	Marginal X
40-60	10	6	2	0	18
60-80	8	12	6	2	28
80-100	1	8	10	6	25
Marginal Y	19	26	18	8	71

➤ **Distribución marginal de X**

◆ **Distribución de la variable X : “Peso”**

X	Frecuencias Marginales
40-60	18
60-80	28
80-100	25
	71

✓ Media Marginal de X

✓ Mediana Marginal de X

✓ Moda Marginal de X

✓ Varianza Marginal de X

➤ **Distribución marginal de Y**

◆ **Distribución de la variable Y : “Estatura”**

$X \setminus Y$	140-160	160-180	180-200	>200	Marginal X
40-60	10	6	2	0	18
60-80	8	12	6	2	28
80-100	1	8	10	6	25
Marginal Y	19	26	18	8	71

➤ **Distribución marginal de Y**

◆ **Distribución de la variable Y : “Estatura”**

Y	Frecuencias Marginales
140-160	19
160-180	26
180-200	18
>200	8
	71

✓ Media Marginal de Y

✓ Mediana Marginal de Y

✓ Moda Marginal de Y

✓ Varianza Marginal de Y

➤ **Distribuciones de X
condicionadas a valores de Y**

◆ **Ejemplo . Distribución de X
condicionada a $160 < Y < 180$**

$X \backslash Y$	140-160	160-180	180-200	>200	Marginal X
40-60	10	6	2	0	18
60-80	8	12	6	2	28
80-100	1	8	10	6	25
Marginal Y	19	26	18	8	71

**♦ Ejemplo . Distribución de X
condicionada a $160 < Y < 180$**

X	Frecuencias condicionadas
40-60	6
60-80	12
80-100	8
	26

✓ Medias condicionadas de X

✓ Varianzas condicionadas de X

➤ **Distribuciones de Y
condicionadas a valores de X**

◆ **Ejemplo . Distribución de Y
condicionada a $60 < X < 80$**

$X \backslash Y$	140-160	160-180	180-200	>200	Marginal X
40-60	10	6	2	0	18
60-80	8	12	6	2	28
80-100	1	8	10	6	25
Marginal Y	19	26	18	8	71

◆ **Ejemplo . Distribución de Y condicionada a $60 < X < 80$**

Y	Frecuencias condicionadas
140-160	8
160-180	12
180-200	6
>200	2
	28

✓ Medias condicionadas de Y

✓ Varianzas condicionadas de Y

❖ 1.6 Covarianza

$$\text{Cov}[X, Y] = \sigma_{xy} = \frac{\sum_i \sum_j n_{ij} (x_i - \bar{x})(y_j - \bar{y})}{n} =$$

$$= \frac{\sum_i \sum_j n_{ij} x_i y_j}{n} - \bar{x} \bar{y}$$

❖ 1.7 Dependencia e independencia estadística

➤ Independencia estadística

- No hay relación entre las variables

$$\text{Si } n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \quad \forall i, j$$

➤ Dependencia estadística

- ≡ Hay relación entre las variables

El grado de relación se mide mediante un coeficiente de asociación

◆ Ejemplo. Variables X e Y independientes

$X \setminus Y$	Y_1	Y_2	Y_3	Y_4	$n_{i \cdot}$
X_1	n_{11} = 2	n_{12} = 6	n_{13} = 4	n_{14} = 8	$n_{1 \cdot}$ = 20
X_2	n_{21} = 3	n_{22} = 9	n_{23} = 6	n_{24} = 12	$n_{2 \cdot}$ = 30
X_3	n_{31} = 1	n_{32} = 3	n_{33} = 2	n_{34} = 4	$n_{3 \cdot}$ = 10
$n \cdot j$	$n \cdot 1$ = 6	$n \cdot 2$ = 18	$n \cdot 3$ = 12	$n \cdot 4$ = 24	n = 60

Independencia estadística

Si $n_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n} \quad \forall i, j$

$$n_{23} = \frac{n_{2 \cdot} \cdot n_{\cdot 3}}{n} = \frac{30 \times 12}{60} = 6$$

$$n_{31} = \frac{n_{3 \cdot} \cdot n_{\cdot 1}}{n} = \frac{10 \times 6}{60} = 1$$

◆ Ejemplo. Variables X e Y no independientes

$X \setminus Y$	Y_1	Y_2	Y_3	Y_4	$n_{i \cdot}$
X_1	n_{11} $= 3$	n_{12} $= 6$	n_{13} $= 4$	n_{14} $= 8$	$n_{1 \cdot}$ $= 21$
X_2	n_{21} $= 3$	n_{22} $= 10$	n_{23} $= 6$	n_{24} $= 12$	$n_{2 \cdot}$ $= 31$
X_3	n_{31} $= 1$	n_{32} $= 3$	n_{33} $= 2$	n_{34} $= 4$	$n_{3 \cdot}$ $= 10$
$n \cdot j$	$n \cdot 1$ $= 7$	$n \cdot 2$ $= 19$	$n \cdot 3$ $= 12$	$n \cdot 4$ $= 24$	n $= 62$

Independencia estadística

$$\text{Si } n_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n} \quad \forall i, j$$

$$n_{23} = \frac{n_{2 \cdot} \cdot n_{\cdot 3}}{n} = \frac{31 \times 12}{62} = 6$$

$$n_{31} \neq \frac{n_{3 \cdot} \cdot n_{\cdot 1}}{n} = \frac{10 \times 7}{62} = 1.129 \neq 1$$

♦ Ejemplo. Dependencia Funcional

❖ .- Dadas las siguientes distribuciones bidimensionales:

1. ¿Son independientes las variables X e Y ?
2. ¿Dependen funcionalmente las variables X e Y ?

a.

$X \setminus Y$	10	15	20
1	0	3	0
2	1	0	0
3	0	0	5
4	0	1	0

b.

$X \setminus Y$	10	15	20	25
1	0	3	0	4
2	0	0	1	0
3	2	0	0	0

c.

$X \setminus Y$	10	15	20
1	0	5	0
2	3	0	0
3	0	0	2

d.

$X \setminus Y$	10	15	20
1	3	2	0
2	1	0	2
3	0	1	1

1. ¿Son independientes las variables X e Y ?

a.

$X \setminus Y$	10	15	20	Marginal X
1	0	3	0	3
2	1	0	0	1
3	0	0	5	5
4	0	1	0	1
Marginal Y	1	4	5	10

$$n_{12} \neq \frac{n_{1.} \cdot n_{.2}}{n} = \frac{3 \times 4}{10} = 1.2 \neq 3 \quad \Rightarrow$$

Las variables X e Y no son independientes

b.

$X \setminus Y$	10	15	20	25	Marginal X
1	0	3	0	4	7
2	0	0	1	0	1
3	2	0	0	0	2
Marginal Y	2	3	1	4	10

$$n_{23} \neq \frac{n_{2.} \cdot n_{.3}}{n} = \frac{1 \times 1}{10} = 0.1 \neq 1 \quad \Rightarrow$$

Las variables X e Y no son independientes

1. ¿Son independientes las variables X e Y ?

c.

$X \setminus Y$	10	15	20	Marginal X
1	0	5	0	5
2	3	0	0	3
3	0	0	2	2
Marginal Y	3	5	2	10

$$n_{11} \neq \frac{n_{1.} \cdot n_{.1}}{n} = \frac{5 \times 3}{10} = 1.5 \neq 0 \quad \Rightarrow$$

Las variables X e Y no son independientes

d.

$X \setminus Y$	10	15	20	Marginal X
1	3	2	0	5
2	1	0	2	3
3	0	1	1	2
Marginal Y	4	3	3	10

$$n_{21} \neq \frac{n_{2.} \cdot n_{.1}}{n} = \frac{3 \times 4}{10} = 1.2 \neq 1 \quad \Rightarrow$$

Las variables X e Y no son independientes

2. ¿Dependen funcionalmente las variables X e Y ?

a.

$X \setminus Y$	10	15	20
1	0	3	0
2	1	0	0
3	0	0	5
4	0	1	0

Y Depende funcionalmente de X

X No Depende funcionalmente de Y

b.

$X \setminus Y$	10	15	20	25
1	0	3	0	4
2	0	0	1	0
3	2	0	0	0

Y No Depende funcionalmente de X

X Depende funcionalmente de Y

2. ¿Dependen funcionalmente las variables X e Y ?

c.

$X \setminus Y$	10	15	20
1	0	5	0
2	3	0	0
3	0	0	2

X Depende funcionalmente de Y

Y Depende funcionalmente de X

d.

$X \setminus Y$	10	15	20
1	3	2	0
2	1	0	2
3	0	1	1

X No Depende funcionalmente de Y

Y No Depende funcionalmente de X

❖ 1.8 Regresión y correlación. Introducción

❖ Regresión

- Búsqueda de una función que relacione ambas variables y sirva para predecir una variable a partir de la otra

$$y = f(x)$$

❖ Correlación

- Estudio del nivel de relación entre las variables

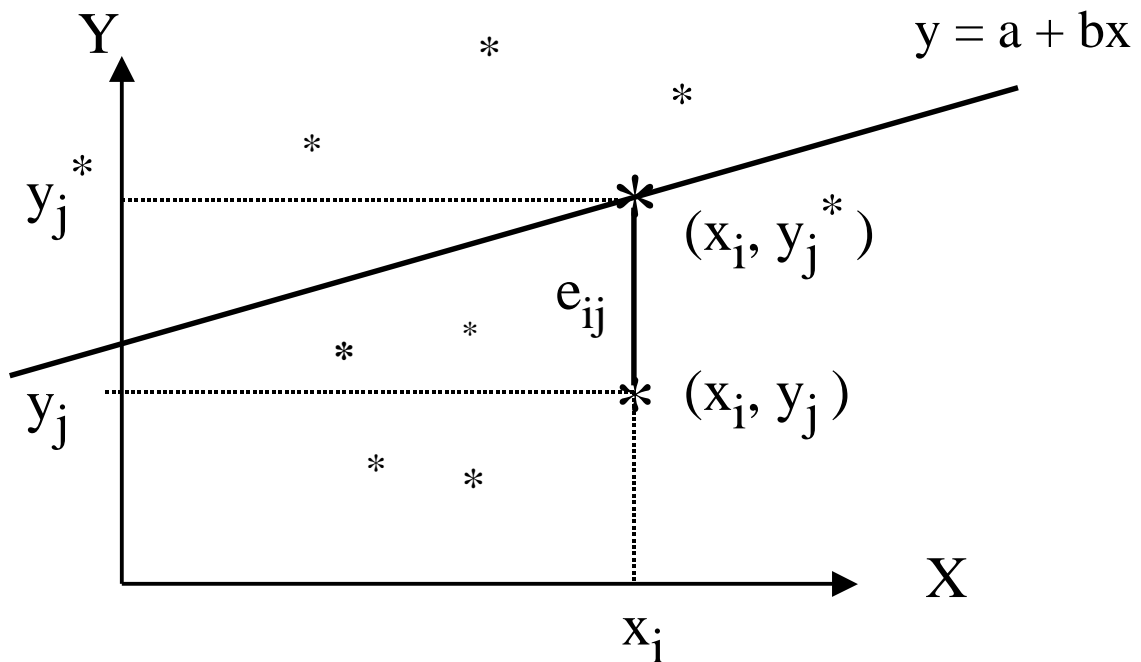
✓ Nube de puntos (diagrama de dispersión): gráfico de las observaciones (datos bidimensionales)

✓ Línea o función de regresión: tipo de función que mejor se ajuste a la nube de puntos:

◆ Lineal ; Cuadrática; Exponencial...

❖ 1.9 Rectas de regresión

❖ Recta de mínimos cuadrados de Y / X



$$\text{Residuos} = e_{ij} = y_j - (a + bx_i)$$

$$\begin{aligned} \min \sum_i \sum_j e_{ij}^2 &= \min \sum_i \sum_j (y_j - y_j^*)^2 = \\ &= \min \sum_i \sum_j (y_j - (a + bx_i))^2 \quad \Rightarrow \end{aligned}$$

Ecuaciones normales

❖ Recta de mínimos cuadrados de Y/X

$$y = f(x) = a + bx$$

$$b = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\frac{\sum n_i x_i y_i}{n} - \bar{x} \bar{y}}{\frac{\sum n_i x_i^2}{n} - \bar{x}^2}$$

$$a = \bar{y} - b\bar{x}$$

$$y - \bar{y} = b(x - \bar{x})$$

b = coeficiente de regresión de Y/X
“Variación de Y si X aumenta en una unidad”

❖ Recta de mínimos cuadrados de X / Y

$$x = f(y) = c + dy$$

$$d = \frac{\text{Cov}[X, Y]}{\text{Var}[Y]} = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{\frac{\sum n_i x_i y_i}{n} - \bar{x} \bar{y}}{\frac{\sum n_i y_i^2}{n} - \bar{y}^2}$$

$$c = \bar{x} - d \bar{y}$$

$$x - \bar{x} = d(y - \bar{y})$$

d = coeficiente de regresión de X / Y

“Variación de X si Y aumenta en una unidad”

❖ 1.10 Coeficiente de determinación y coeficiente de correlación lineal

❖ Coeficiente de determinación

➤ "Proporción de la varianza explicada por la regresión"

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} ; \quad 0 \leq r^2 \leq 1$$

❖ Coeficiente de correlación lineal de Pearson

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} ; \quad -1 \leq r \leq 1$$

$r = 0 \Leftrightarrow$ *Independencia*

$r > 0 \Leftrightarrow$ *Dependencia directa*

$r < 0 \Leftrightarrow$ *Dependencia inversa*

$r = \pm 1 \Leftrightarrow$ *Dependencia funcional lineal*

◆ Ejemplo. $X=$ “Estatura”, $Y=$ “Peso”

x_i	y_i	$x_i y_i$	x^2_i	y^2_i
160	52	8320	25600	2704
172	64	11008	29584	4096
174	65	11310	30276	4225
176	72	12672	30976	5184
180	78	14040	32400	6084
$\Sigma=862$	$\Sigma= 331$	$\Sigma= 57350$	$\Sigma= 148836$	$\Sigma= 22293$

$$\bar{x} = \frac{862}{5} = 172.4 ; \quad \bar{y} = \frac{331}{5} = 66.2$$

$$\sigma_{xy} = \frac{\sum n_i x_i y_i}{n} - \bar{x} \bar{y} = \frac{57350}{5} - 172.4 \times 66.2 = 57.12$$

$$\sigma_x^2 = \frac{\sum n_i x_i^2}{n} - \bar{x}^2 = \frac{148836}{5} - 172.4^2 = 45.44$$

$$\sigma_y^2 = \frac{\sum n_i y_i^2}{n} - \bar{y}^2 = \frac{22293}{5} - 66.2^2 = 76.16$$

$$y = a + bx$$

$$b = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{57.12}{45.44} = 1.257$$

$$a = \bar{y} - b\bar{x} = 66.2 - 1.257 \times 172.4 = -150.5068$$

$$y = a + bx = -150.5068 + 1.257x$$

Para $x = 170 \Rightarrow$

$$y = a + bx = -150.5068 + 1.257 \times 170 = 63.1832$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{57.12}{\sqrt{45.44} \sqrt{76.16}} = 0.9708$$

❖ 1.11 Otros tipos de ajuste

➤ Parabólico

$$y = ax^2 + bx + c$$

➤ Exponencial

$$y = ab^x$$

➤ Potencial

$$y = ax^b$$

➤ Hiperbólico

$$y = \frac{a}{x}$$