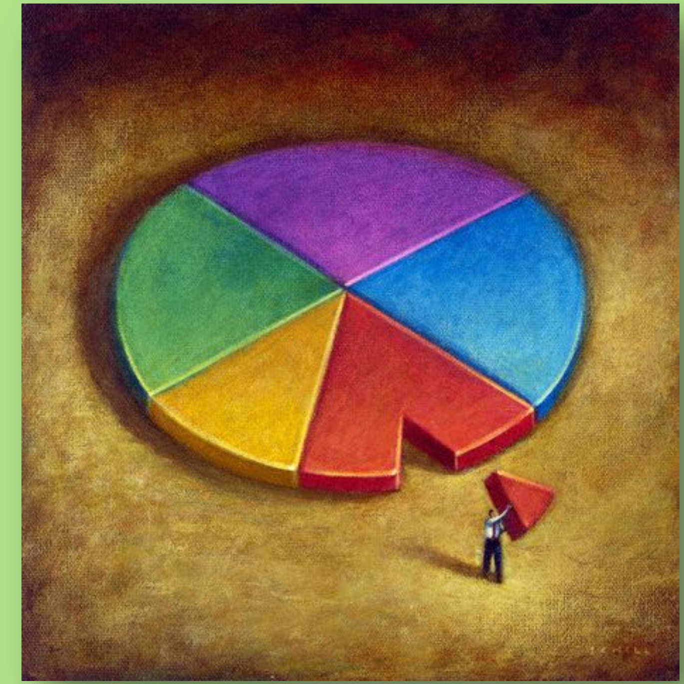
ESTADÍSTICA Y PROBABILIDAD



1. Estadística Bidimensional

B5.C1.1. Elabora tablas bidimensionales de frecuencias con variables discretas y continuas.
B5.C1.2. Calcula e interpreta parámetros estadísticos en variables bidimensionales.
B5.C1.3. Calcula las distribuciones marginales y diferentes distribuciones condicionadas, así como sus parámetros (media, varianza y desviación típica).
B5.C1.4. Decide si dos variables estadísticas son o no dependientes a partir de sus distribuciones condicionadas y marginales.
B5.C2.1. Distingue la dependencia funcional de dependencia estadística y estima si dos variables son o no dependientes mediante la representación de la nube de puntos.
B5.C2.2. Cuantifica el grado y sentido de la dependencia lineal entre dos variables mediante el cálculo e interpretación del coeficiente de correlación lineal.
B5.C2.3. Calcula las rectas de regresión de dos variables y obtiene predicciones.
B5.C2.4. Evalúa la fiabilidad de las predicciones a partir de la recta de regresión mediante el coeficiente de determinación lineal.

1. Estadística Bidimensional

Se ha hecho una encuesta a 500 varones, mayores de edad, sobre sus preferencias televisivas. Hay seis posibles respuestas (como puedes ver en el margen). Los resultados se dan en la siguiente tabla:

Edad ☐

De los siguientes tipos de programas televisivos, ¿cuál prefieres?:

Informativos .... ☐

Documentales .... ☐

Entretenimiento (concursos, famosos, corazón…) .... ☐

Deportes .... ☐

Películas y series .... ☐

Otros .... ☐

1. Estadística Bidimensional

Tablas de Contingencia

y \ x	18-25	26-35	36-50	51-65	MÁS DE 65	TOTAL
INF	4	6	15	11	25	61
DOC	5	15	21	15	21	77
ENT	10	10	22	15	16	73
DEP	20	34	56	40	50	200
PEL	11	15	20	16	11	73
OTR	0	5	6	3	2	16
TOTAL	50	85	140	100	125	500

La variable x describe las *edades* de los encuestados. Se presenta repartida en cinco tramos. Es una variable cuantitativa.

La variable y describe las *preferencias televisivas*. Es una variable cualitativa.

1. Estadística Bidimensional

Sumamos las filas y columnas

y \ x	18-25	26-35	36-50	51-65	MÁS DE 65	TOTAL
INF	4	6	15	11	25	61
DOC	5	15	21	15	21	77
ENT	10	10	22	15	16	73
DEP	20	34	56	40	50	200
PEL	11	15	20	16	11	73
OTR	0	5	6	3	2	16
TOTAL	50	85	140	100	125	500

1. Estadística Bidimensional

Distribuciones Marginales

Si a cada grupo de edades le asignamos la frecuencia correspondiente de la última fila (TOTAL) obtenemos la **distribución marginal de la** x.

x_i	18-25	26-35	36-50	51-65	MÁS DE 65	
f_i	50	85	140	100	125	500

Análogamente, la última columna (TOTAL) nos da las frecuencias de la **distribución marginal de la** y.

y_i	f'_i
INF	61
DOC	77
ENT	73
DEP	200
PEL	73
OTR	16
	500

1. Estadística Bidimensional

Distribuciones Marginales

Si a cada grupo de edades le asignamos la frecuencia correspondiente de la última fila (TOTAL) obtenemos la **distribución marginal de la** x.

x_i	18-25	26-35	36-50	51-65	MÁS DE 65	
f_i	50	85	140	100	125	500

Análogamente, la última columna (TOTAL) nos da las frecuencias de la **distribución marginal de la** y.

y_i	f_i'
INF	61
DOC	77
ENT	73
DEP	200
PEL	73
OTR	16
	500

Pag. 347

1 Calcula la media y la desviación típica de la distribución marginal de la x. Para ello, asigna a cada intervalo de edades su marca de clase (punto medio) y al último intervalo asígnale el valor 75.

1. Estadística Bidimensional

Distribuciones Condicionadas

¿Cómo se distribuyen las *preferencias televisivas* de los mayores de 65 años? Lo vemos en la quinta columna de la tabla: es la *distribución de la variable y* **condicionada** *a que x sea del intervalo "más de 65"*.

VAR	FREC
INF	25
DOC	21
ENT	16
DEP	50
PEL	11
OTR	2

1. Estadística Bidimensional

Distribuciones Condicionadas

La cuarta fila de la tabla nos da las frecuencias de la distribución de edades de los que prefieren deporte. Es la *distribución de la variable* x **condicionada** *a que* $y = DEP$.

VAR	18-25	26-35	36-50	51-65	MÁS DE 65
FREC	20	34	56	40	50

1. Estadística Bidimensional

Variables independientes

Vamos a comparar las **frecuencias relativas** de estas distribuciones:

— La distribución de la x condicionada a y = DEP.

— La distribución marginal de la x.

18-25	26-35	36-50	51-65	MÁS DE 65
$\frac{20}{200} = 0{,}10$	$\frac{34}{200} = 0{,}17$	$\frac{56}{200} = 0{,}28$	$\frac{40}{200} = 0{,}20$	$\frac{50}{200} = 0{,}25$
$\frac{50}{500} = 0{,}10$	$\frac{85}{500} = 0{,}17$	$\frac{140}{500} = 0{,}28$	$\frac{100}{500} = 0{,}20$	$\frac{125}{500} = 0{,}25$

Las frecuencias relativas son idénticas. Es decir, los que prefieren deporte se reparten por edades de igual manera que el colectivo total. Eso lo concretamos diciendo que el suceso PREFIERE DEPORTE es **independiente** de la edad.

1. Estadística Bidimensional

Variables independientes

Cuando todos los sucesos de una variable son independientes respecto de la otra, se dice que **las dos variables son independientes.**

En la tabla que nos ocupa, los demás sucesos (y = INF, y = DOC, …) no son independientes de la edad. Por eso, estas dos variables x e y no son independientes.

1. Estadística Bidimensional

Ejercicio 1

1. Para hacer un estudio sobre las preferencias sobre 2 modelos de impresoras 3D de reciente fabricación, se consideraron las ventas de un distribuidor durante 25 días.

Ventas Modelo Impresora A (X): 0 2 2 2 1 3 3 3 3 4 4 2 3 3 3 3 2 3 2 4 2 2 3 3 3
Ventas Modelo Impresora B (Y): 2 1 2 2 3 1 1 1 2 0 1 1 1 1 1 2 2 1 1 1 2 2 2 2 1

a) Completa la siguiente tabla de contingencia de frecuencias absolutas y las correspondientes distribuciones marginales de la X y de la Y.

x_i/y_j	0	1	2	3	4	n_i
0						
1						
2						
3						
4						
n_j						

x_i	n_i

y_j	n_j

1. Estadística Bidimensional

Ejercicio 1

b) Calcula la media y la varianza de las distribuciones marginales del apartado a).

c) Completa la siguiente tabla de contingencia de frecuencias relativas y las correspondientes distribuciones marginales de la X y de la Y. Mirando esas tablas, ¿podrías decir si las variables X e Y son independientes?.

x_i/y_j	0	1	2	3	4	n_i
0						
1						
2						
3						
4						
n_j						

x_i	n_i

y_j	n_j

d) Escribe la tabla de frecuencias de la X condicionada a que la Y=3 ó 4 y escribe la tabla de frecuencias de la Y condicionada a que la X=1.

e) Mirando las tablas de frecuencias relativas del apartado c) responde a las siguientes preguntas:
e.1) ¿Qué porcentaje de días se ha vendido 1 unidad del modelo A y 3 unidades del modelo B?
e.2) ¿Qué porcentaje de días se han vendido 3 unidades del modelo B?
e.3) ¿Qué porcentaje de los días se han vendido 2 unidades del modelo A?.
e.4) ¿Qué porcentaje de los días se han vendido más unidades del modelo B que del modelo A?.

1. Relación funcional y estadística

Relación Funcional

- Si tiramos una piedra con una cierta fuerza F hacia arriba se puede calcular mediante una fórmula la altura que alcanzará. Por tanto, hay una función que nos permite representar este hecho.

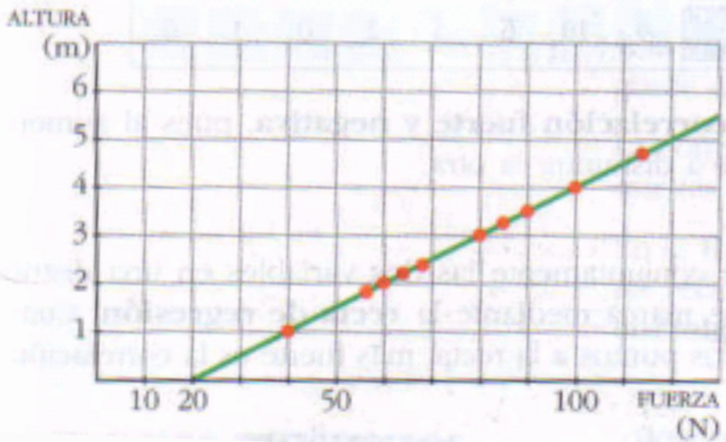
Relación Estadística

- Si tratamos de relacionar la altura y el peso no vamos a encontrar una fórmula exacta. En este caso no es una relación funcional, sino una relación estadística

1. Relación funcional y estadística

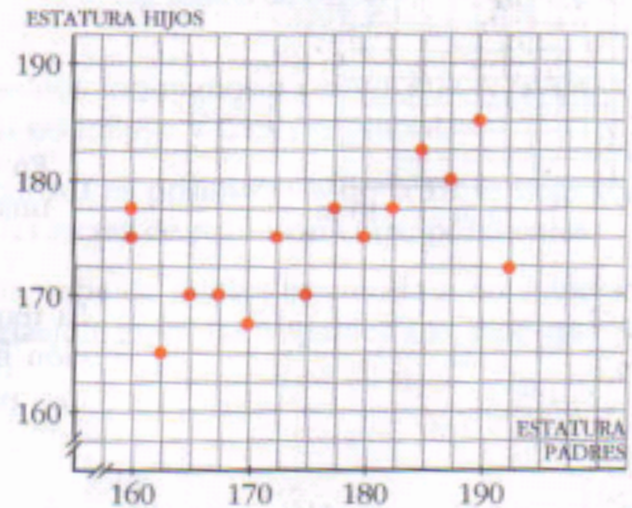
Ejemplo de relación funcional

Distintas personas lanzan hacia arriba una misma piedra de 2 kg de masa, que alcanza más o menos altura según la fuerza con que ha sido impulsada. (La fuerza actúa en un tramo de 1 m).



Ejemplo de relación estadística

En la siguiente gráfica, cada punto corresponde a un chico. La abscisa es la estatura de su padre, y la ordenada, su propia altura.



2. Nubes de puntos. Correlación

Tenemos un colectivo de n individuos. Estudiamos en ellos dos variables, x, y. Conocemos los valores de las variables para cada uno de los individuos.

El conjunto de pares de valores (x_1, y_1), (x_2, y_2), …, (x_n, y_n) se llama **distribución bidimensional**. Si interpretamos cada par de valores como las coordenadas de un punto, el conjunto de todos ellos se llama **nube de puntos** o **diagrama de dispersión**.

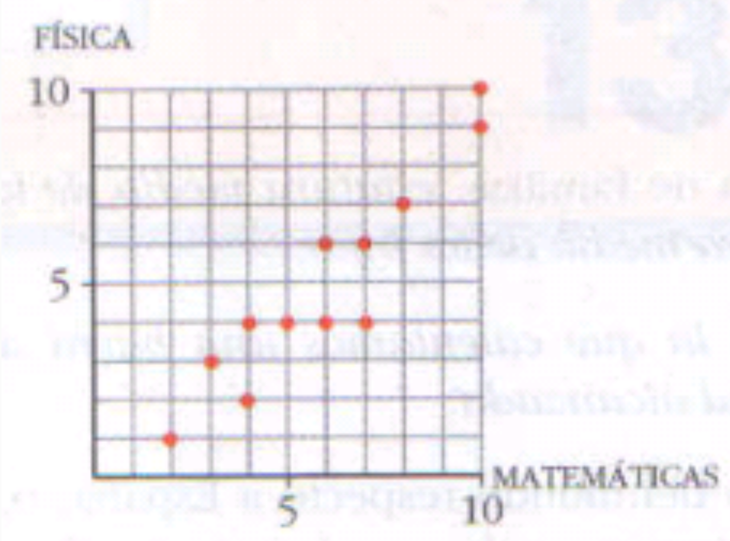
La **correlación** viene a representar la relación que existe entre esas dos variables para los n individuos. Puede ser más o menos fuerte según lo apretados que estén los puntos de la nube en torno a una recta que marca la tendencia y se llama **recta de regresión**.

Si la pendiente de la recta de regresión es positiva o negativa, la **correlación** se llama **positiva** o **negativa**, respectivamente.

2. Nubes de puntos. Correlación

Estas son las notas de 12 estudiantes en Matemáticas y en Física:

ALUMNO	a	b	c	d	e	f	g	h	i	j	k	l
MATEMÁTICAS	2	3	4	4	5	6	6	7	7	8	10	10
FÍSICA	1	3	2	4	4	4	6	4	6	7	9	10

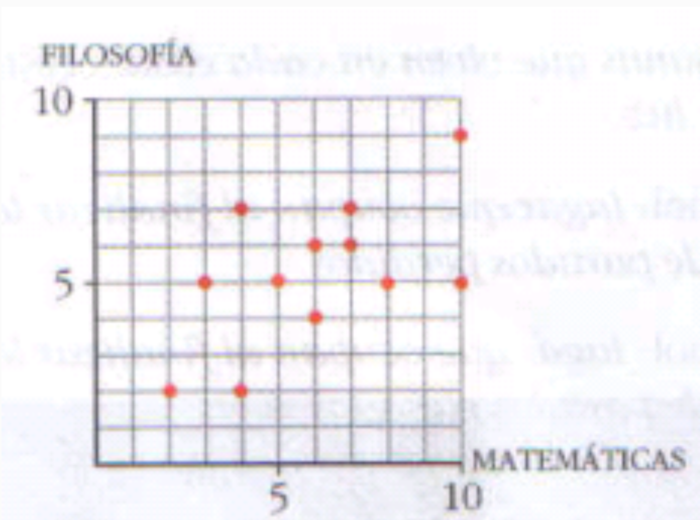


Cuando se aprecia que hay relación entre las notas se dice que hay correlación entre ellas

2. Nubes de puntos. Correlación

Relacionemos ahora las notas de *Matemáticas* de los mismos alumnos con las de otra asignatura, *Filosofía*.

ALUMNO	a	b	c	d	e	f	g	h	i	j	k	l
MATEMÁTICAS	2	3	4	4	5	6	6	7	7	8	10	10
FILOSOFÍA	2	5	2	7	5	4	6	6	7	5	5	9



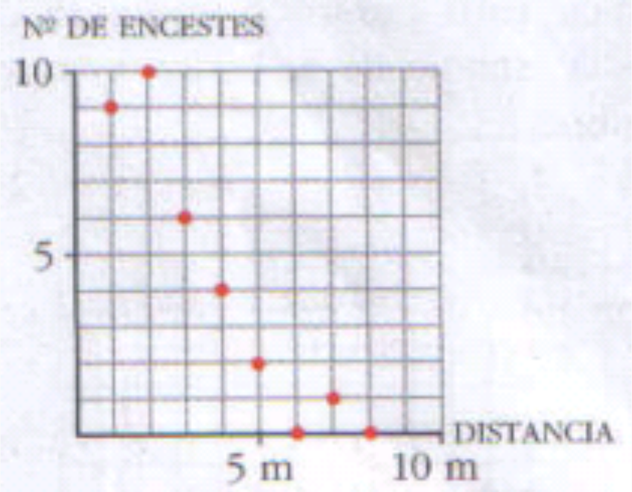
En este caso hay correlación pero es más débil.

2. Nubes de puntos. Correlación

Una jugadora de baloncesto lanza a canasta, desde distintas distancias, 10 balones cada vez. Lógicamente, encesta más cuanto más cerca está.

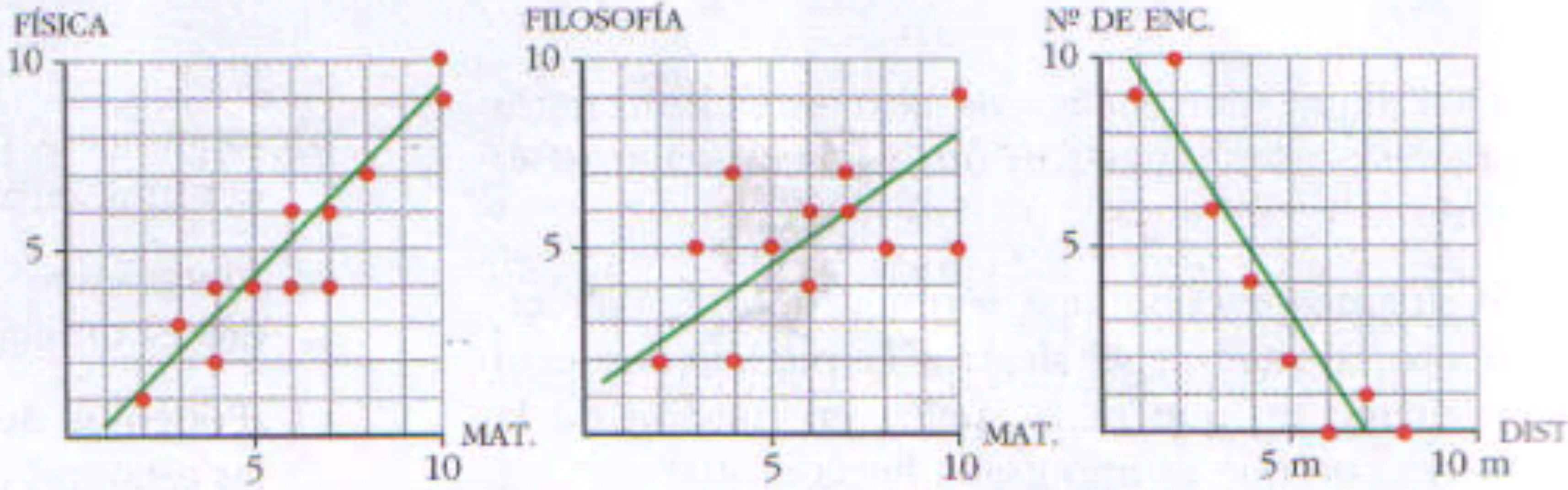
DISTANCIA (m)	1	2	3	4	5	6	7	8
ENCESTES	9	10	6	4	2	0	1	0

En este caso hay **correlación fuerte y negativa**, pues al aumentar una variable tiende a disminuir la otra.



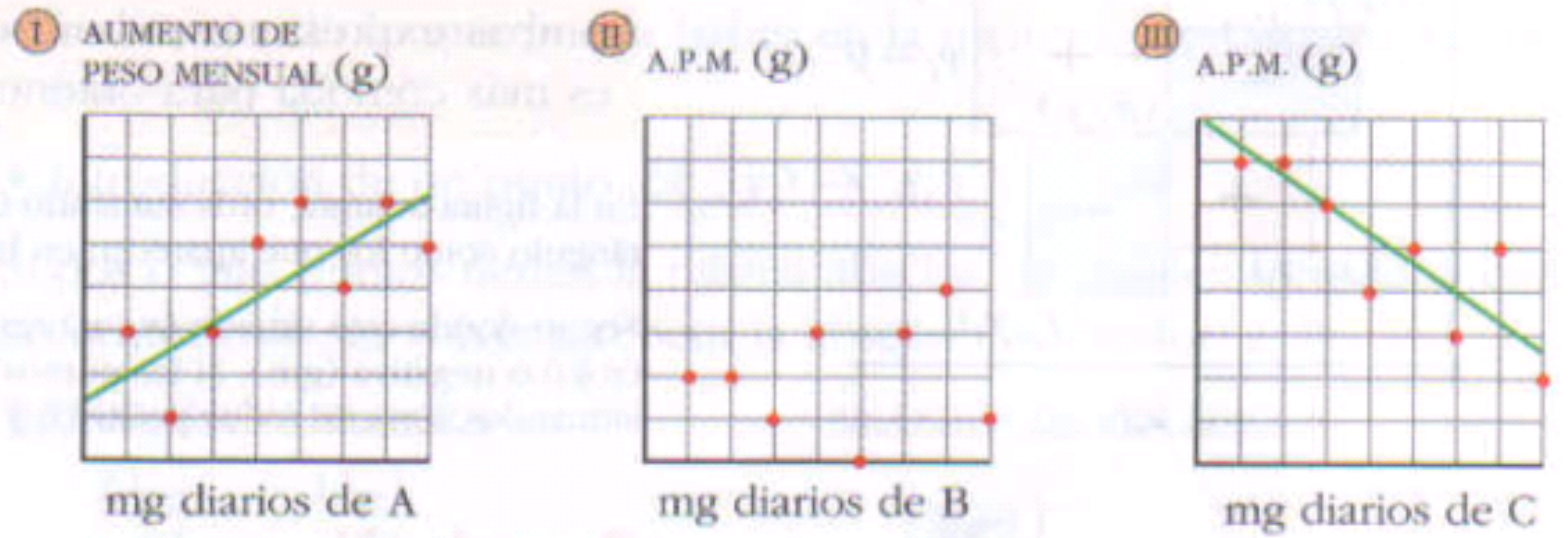
2. Nubes de puntos. Correlación

La tendencia a variar conjuntamente las dos variables en una distribución bidimensional se marca mediante la **recta de regresión**. Cuanto más próximos estén los puntos a la recta, más fuerte es la correlación.



2. Nubes de puntos. Correlación

Se suministra, a cada una de 8 cobayas, una dosis diaria de 1 mg, 2 mg, …, 8 mg, respectivamente, de un cierto fármaco A. Al cabo de un mes, calculamos el aumento de peso de cada cobaya. Repetimos el experimento con otras 8 cobayas y otro fármaco B y, por último, un tercer experimento con otras 8 cobayas y otro fármaco C.



2. Nubes de puntos. Correlación

Ejercicio Pag. 333

1. La tabla de la derecha muestra cómo se ordenan entre sí diez países, A, B, C..., según dos variables, R.P.C. (*renta per cápita*) e I.N. (*índice de natalidad*). Representa los resultados en una nube de puntos, traza la recta de regresión y di cómo te parece la correlación.

PAÍSES	A	B	C	D	E	F	G	H	I	J
R.P.C.	1	2	3	4	5	6	7	8	9	10
I.N.	10	6	9	5	7	4	1	3	8	2

2. Correlación

El punto $(\bar{x}, \bar{y})$ se llama **centro de gravedad** de la distribución.

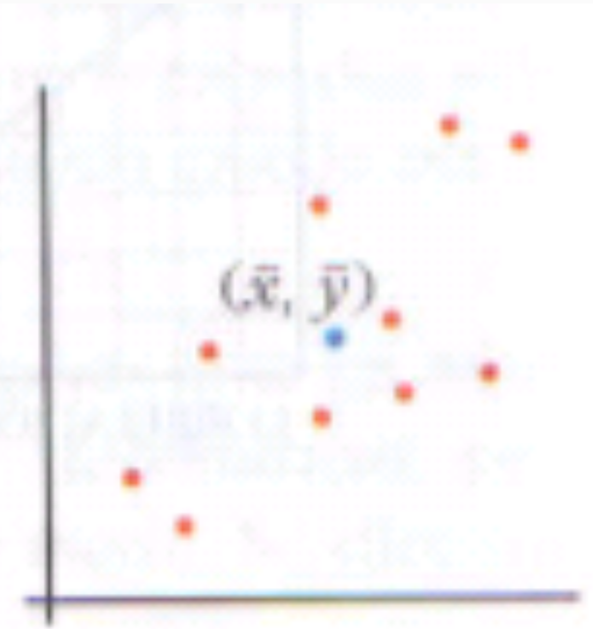
x_i	y_i
x_1	y_1
x_2	y_2
x_3	y_3
...	...
x_n	y_n

MEDIA DE LA VARIABLE x

$$\bar{x} = \frac{\Sigma x_i}{n}$$

MEDIA DE LA VARIABLE y

$$\bar{y} = \frac{\Sigma y_i}{n}$$



2. Correlación

El valor de la **correlación** entre las dos variables de una distribución bidimensional viene dado por la expresión:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

σ_{xy} es la covarianza

σ_x, σ_y son las desviaciones típicas de cada variable

- r no tiene unidad
- r próximo a 1 y -1 es correlación fuerte
- r próximo a 0 es correlación débil

$$\sigma_x = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}$$

Se llama **covarianza** al parámetro:

$$\sigma_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y}$$

2. Correlación

calcular la correlación entre las variables:

x: nota en Matemáticas

y: nota en Física

Para ello, calcular previamente $\bar{x}$, $\bar{y}$, σ_x, σ_y, σ_{xy}.

$$\bar{x} = \frac{\Sigma x_i}{n} =$$

$$\bar{y} = \frac{\Sigma y_i}{n} =$$

$$\sigma_x = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2} =$$

$$\sigma_y = \sqrt{\frac{\Sigma y_i^2}{n} - \bar{y}^2} =$$

$$\sigma_{xy} = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y} =$$

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	1	4	1	2
3	3	9	9	9
4	2	16	4	8
4	4	16	16	16
5	4	25	16	20
6	4	36	16	24
6	6	36	36	36
7	4	49	16	28
7	6	49	36	42
8	7	64	49	56
10	9	100	81	90
10	10	100	100	100
72	60	504	380	431

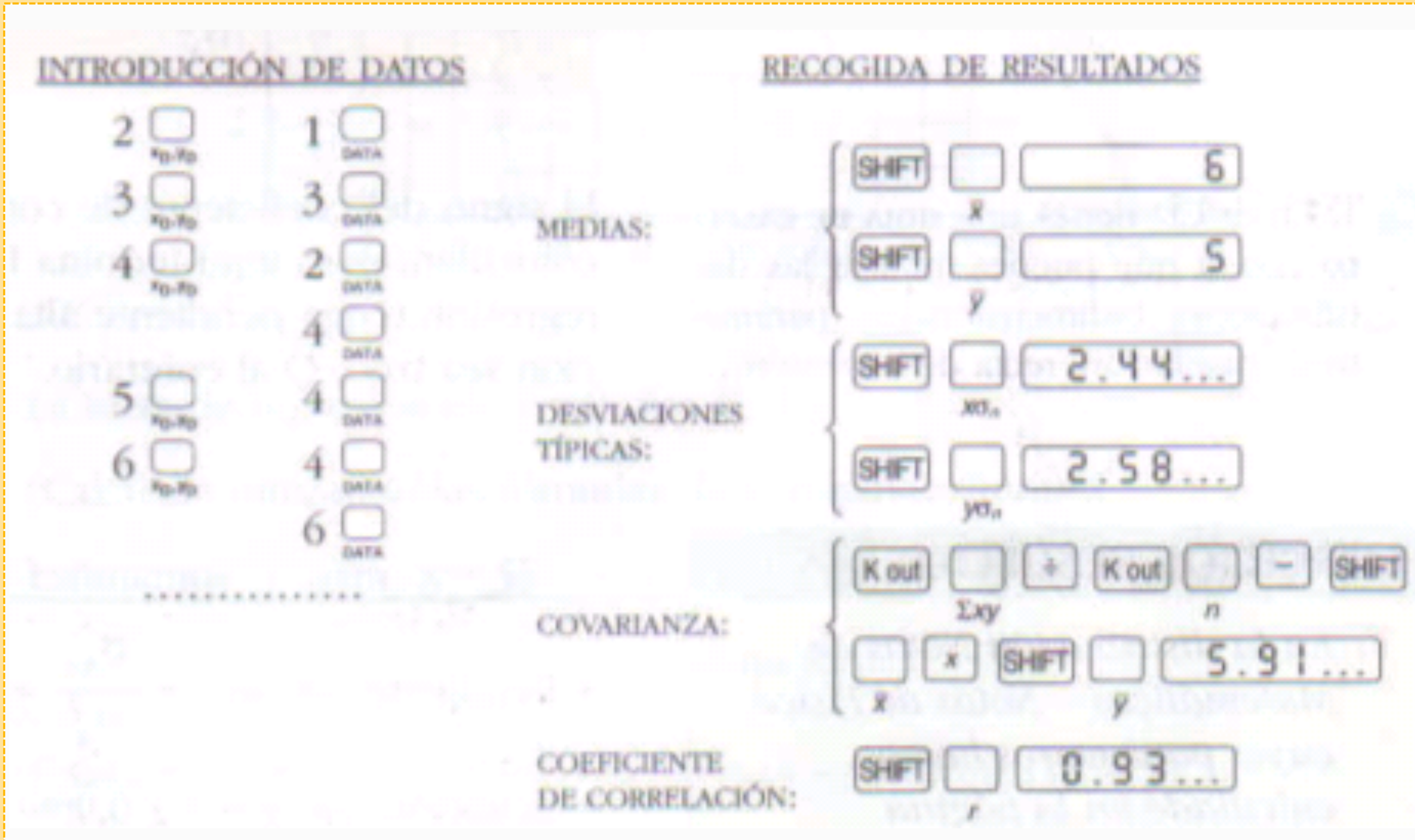
3. Correlación con la Calculadora

- Preparación de la calculadora para tratamiento bidimensional → MODO LR

- Se borra todo lo que pueda haber en la memoria de cálculos anteriores → SHIFT AC

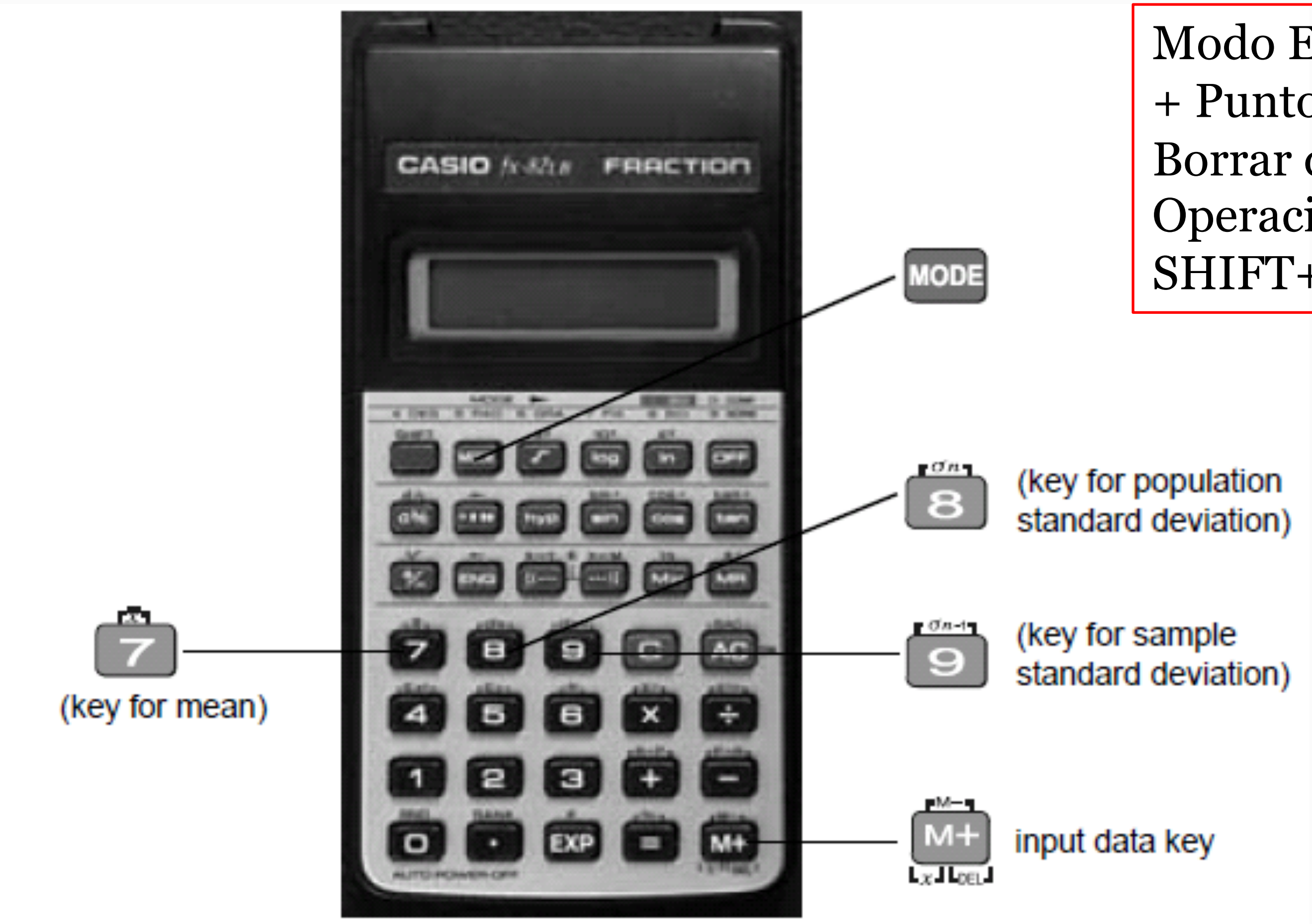
- Introducción de un punto $(x_i, y_i) \rightarrow x_i \; \boxed{} \; y_i$ DATA

3. Correlación con la Calculadora



3. Modelo fx-82l

En esta calculadora se trabaja en modo unidimensional…



Modo Estadística – Mode + Punto
Borrar datos –SHIFT + AC
Operaciones usar SHIFT+ Número

3. Modelo fx-82ms

15. Cálculos estadísticos (SD, REG*)
*sólo en las fx-82MS/85MS/300MS/350MS

Para seleccionar este tipo de cálculo estadístico: (La fórmula de regresión se muestra entre paréntesis)	Realice esta operación de tecla:
Una variable (X)	MODE 2 (SD)
Par de variables (X, Y), regresión lineal ($y = A + Bx$)	MODE 3 (REG) 1 (Lin)
Par de variables (X, Y), regresión logarítmica ($y = A + B\ln x$)	MODE 3 (REG) 2 (Log)
Par de variables (X, Y), regresión exponencial con base e ($y = Ae^{Bx}$)	MODE 3 (REG) 3 (Exp)
Par de variables (X, Y), regresión en potencias ($y = Ax^{B}$)	MODE 3 (REG) ▶ 1 (Pwr)
Par de variables (X, Y), regresión recíproca ($y = A + B/x$)	MODE 3 (REG) ▶ 2 (Inv)
Par de variables (X, Y), regresión cuadrática ($y = A + Bx + Cx^2$)	MODE 3 (REG) ▶ 3 (Quad)

3. Modelo fx-82ms

■ Ingreso de datos

- En los modos SD y REG, la tecla [M+] funciona como la tecla [DT].
- Inicie el ingreso de datos siempre con [SHIFT] [MODE] (CLR) [1] (Scl) [=] ([CLR] [1] (Scl) [=] en la fx-82SX PLUS/220 PLUS) para borrar la memoria estadística.
- Ingrese los datos utilizando la secuencia de teclas que se indica a continuación.

  Modo SD: <datos x> [DT]

  Modo REG: <datos x> [,] <datos y> [DT]

■ Precauciones en el ingreso de datos

- Durante el ingreso de los datos o una vez completado, puede utilizar las teclas [▲] y [▼] para desplazarse a través de los datos que haya ingresado.

3. Modelo fx-82ms

fx-82SX PLUS/220 PLUS:

Suma: Σx^2, Σx, **Cantidad de elementos:** n, **Valor medio:** $\bar{x}$, **Desviación estándar de la población:** σx, **Desviación estándar de la muestra:** sx

SHIFT 4 a 9

fx-82MS/85MS/300MS/350MS:

En el caso del cálculo estadístico de una variable, dispone de las variables marcadas con un asterisco (*).

Suma: Σx^{2*}, Σx^*, Σy^2, Σy, Σxy, Σx^3, Σx^2y, Σx^4, **Cantidad de elementos:** n^*

Σx^2, Σx, n .... SHIFT 1 (S-SUM) 1 a 3

Σy^2, Σy, Σxy .... SHIFT 1 (S-SUM) ▶ 1 a 3

Σx^3, Σx^2y, Σx^4 ..... SHIFT 1 (S-SUM) ▶ ▶ 1 a 3 (Sólo regresiones cuadráticas)

Valor medio: $\bar{x}^*$, $\bar{y}$, **Desviación estándar de la población:** σx^*, σy, **Desviación estándar de la muestra:** sx^*, sy

$\bar{x}$, σx, sx .... SHIFT 2 (S-VAR) 1 a 3

$\bar{y}$, σy, sy .... SHIFT 2 (S-VAR) ▶ 1 a 3

4. Recta de regresión

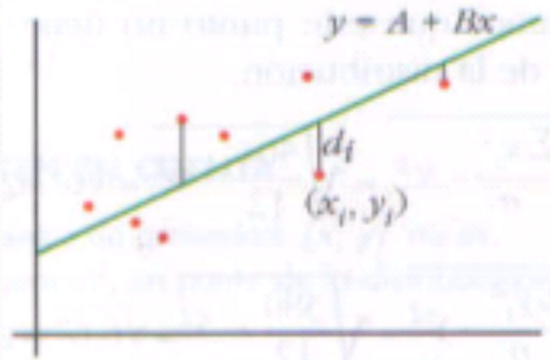
MÉTODO DE LOS MÍNIMOS CUADRADOS

Consideraremos todas las posibles rectas $y = A + Bx$ y nos quedaremos con aquella para la cual los cuadrados de las distancias, d_i, sumen lo menos posible, es decir, para la cual Σd_i^2 es mínimo.

La recta buscada pasa por el centro de gravedad $(\bar{x}, \bar{y})$ de la distribución.

Su pendiente es $m_{yx} = \dfrac{\sigma_{xy}}{\sigma_x^2}$.

$$y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$



d_i = diferencia entre las ordenadas del punto y de la recta.

4. Recta de regresión

EJERCICIO 2

Calcula el coeficiente de correlación entre las variables x, y de la tabla:

x: gastos en publicidad de un producto (en miles de euros)

y: ventas conseguidas (en miles de euros)

x	y
1	10
2	17
3	30
4	28
5	39
6	47

Haya también la recta de regresión de Y sobre X

Y=...

4. Recta de regresión

EJERCICIO 2

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	10	1	100	10
2	17	4	289	34
3	30	9	900	90
4	28	16	784	112
5	39	25	1 521	195
6	47	36	2 209	282
21	171	91	5 803	723