» **Ejercicio 1.** En una fábrica de producción de envases de productos lácteos se ha recogido la siguiente información sobre los productos defectuosos y el tipo de máquina que los ha elaborado. Para ello se ha recogido una muestra de 500 envases producidos por tres máquinas diferentes: A, B y C y se ha observado lo siguiente: la máquina A ha producido 195 productos no defectuosos y 4 defectuosos, la máquina B ha producido 125 productos correctos y 75 defectuosos, y la máquina C ha producido 95 productos sin defecto y 6 con defecto.

- Representa los datos de la muestra en una tabla de doble entrada y obtén las distribuciones de frecuencias absolutas de las variables marginales: calidad de los productos y tipo de máquina.

Tipo de máquina/Calidad de los productos	Defectuoso	No defectuoso	Distribución marginal del Tipo de máquina
A	4	195	199
B	75	125	200
C	6	95	101
Distribución marginal de la calidad de los productos	85	415	N=500

La última fila es la distribución de frecuencias absolutas marginal de la variable Calidad de los productos. La última columna es la distribución de frecuencias absolutas marginal de la variable Tipo de máquina.

- Obtén la distribución conjunta de frecuencias relativas.

Tipo de máquina/Calidad de los productos	Defectuoso	No defectuoso	Distribución marginal del Tipo de máquina
A	4/500=0,008	195/500=0,39	199/500=0,398
B	75/500=0,15	125/500=0,25	200/500=0,4
C	6/500=0,012	95/500=0,19	101/500=0,202
Distribución marginal de la calidad de los productos	85/500=0,17	415/500=0,83	1

Lo que hay que hacer es dividir por el total 500 para obtener frecuencias relativas.

» **Ejercicio 2.** Para estudiar la relación existente entre el precio y el número de habitaciones de una vivienda en Madrid disponemos de los datos referidos a 300 operaciones de venta, de los que se deduce que el precio medio es 250.000€, el coeficiente de variación del precio es 0'25, el número medio de habitaciones es de 3, el coeficiente de variación para esta última variable es 0'10 y finalmente el coeficiente de correlación entre ambas variables se sitúa en 0'8.

- Ajusta un modelo lineal entre las dos variables.

Datos del problema:
Sea la variable aleatoria Y el "Precio" y X el "número de habitaciones". Queremos poner el precio en función del número de habitaciones, es decir, Y en función de X.
El tamaño de la muestra es n=300.
El precio medio es $\bar{Y} = 250000$ y el coeficiente de variación de Y es: $CV_Y = 0{,}25$
El número medio de habitaciones es $\bar{X} = 3$ y el coeficiente de variación de X es: $CV_X = 0{,}10$
El coeficiente de correlación entre X e Y es $r_{XY} = 0{,}8$

De los coeficientes de variación hallamos las desviaciones típicas de cada variable.

$0{,}25 = CV_Y = \frac{s_Y}{|\bar{Y}|} = \frac{s_Y}{250000}$, de aquí se obtiene que

$$s_Y = 62500$$

$0{,}10 = CV_X = \frac{s_X}{|\bar{X}|} = \frac{s_X}{3}$, de aquí se obtiene que

$$s_X = 0{,}3$$

Los parámetros del modelo de regresión $Y = \beta_0 + \beta_1 X$, son:

La pendiente $\beta_1 = r_{XY}\frac{s_Y}{s_X}$ y el intercepto $\beta_0 = \bar{Y} - \beta_1 \bar{X}$

$$\beta_1 = r_{XY}\frac{s_Y}{s_X} = 0{,}8 * \frac{62500}{0{,}3} = 166666{,}\bar{6} \approx 166667$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} = 250000 - 166667 * 3 = -250001$$

El modelo quedaría: $\boldsymbol{Y = -250001 + 166667 * X}$

Como la pendiente es positiva, el precio aumenta según aumenta el número de habitaciones.

- Determina cuánto debería pedirse, de acuerdo con el mercado, por una vivienda de dos habitaciones.

$$Y = -250001 + 166667 * 2 = 83333$$

Debería pedirse 83.333€

- Efectúa una valoración del grado de ajuste del modelo a los datos disponibles.

  El coeficiente de determinación R^2 en regresión lineal simple es el cuadrado del coeficiente de correlación lineal de Pearson, por tanto $R^2 = 0{,}64$ lo cual quiere decir que este modelo explica el 64% de la variabilidad de la variable Y (Precio) en función de X (número de habitaciones). No es un porcentaje muy alto (cercano a 100%) pero es aceptable.

» **Ejercicio 3.** La siguiente tabla de correlación recoge los datos correspondientes a las variables en una muestra formada por 100 habitantes de una ciudad:
X: Renta anual (en miles de euros).
Y: Gasto anual en vacaciones (en cientos de euros).

	Y		
X	**[0,20)**	**[20,40)**	$n_{i.}$
[10,30)	15	5	20
[30,60)	15	35	50
[60,90)	5	25	30
$n_{.j}$	35	65	N=100

Predice (usando un modelo de regresión lineal) el gasto vacacional de un habitante de esta ciudad cuya renta es de 35.000 euros, sabiendo que $s_{xy} = 73$, y cuantifica la bondad del ajuste.

Para ello necesitamos las tablas de frecuencia marginales de cada variable

X	**Marca de clase** x_i	**Frecuencia absoluta** n_i	$n_i * x_i^2$
[10,30)	20	20	8000
[30,60)	45	50	101250
[60,90)	75	30	168750
		N=100	Suma=278000

Y	Marca de clase y_j	Frecuencia absoluta n_j	$n_j * y_j^2$
[0,20)	10	35	3500
[20,40)	30	65	58500
		N=100	Suma=62000

Las medias son

$$\bar{X} = \frac{20 * 20 + 45 * 50 + 75 * 30}{100} = \frac{4900}{100} = 49$$

$$\bar{Y} = \frac{10 * 35 + 30 * 65}{100} = \frac{2300}{100} = 23$$

Las desviaciones típicas son

$$s_X = \sqrt{\frac{\sum_{i=1}^{3} n_i * x_i^2}{N} - \bar{X}^2} = \sqrt{\frac{278000}{100} - 49^2} = \sqrt{379} = 19{,}47$$

$$s_Y = \sqrt{\frac{\sum_{j=1}^{2} n_j * y_j^2}{N} - \bar{Y}^2} = \sqrt{\frac{62000}{100} - 23^2} = \sqrt{91} = 9{,}54$$

El coeficiente de correlación: $r_{XY} = \frac{73}{19{,}47 * 9{,}54} = 0{,}39$

Los parámetros del modelo $Y = \beta_0 + \beta_1 X$, son:

$$\beta_1 = 0{,}39 * \frac{9{,}54}{19{,}47} = 0{,}19$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} = 23 - 0{,}19 * 49 = 13{,}69$$

El modelo queda $Y = 13{,}69 + 0{,}19 * X$

Para el habitante cuya renta es de 35.000 euros (X=35), el gasto anual en vacaciones:

$$Y = 13{,}69 + 0{,}19 * 35 = 20{,}34$$

Es de 2034 euros (recordemos que Y se mide en cientos de euros).

» **Ejercicio 4.** Para estudiar el nivel de vida de los hogares de una gran ciudad, se disponen los datos de 200 hogares relativos a las siguientes variables:

X: Nº de coches disponibles en el hogar.

Y: Nº de días al año, que pasan de vacaciones en el extranjero.

$\bar{X} = 1$ $\bar{Y} = 10{,}6$

S_x=5

S_Y=3

S_{xy}=15

- Obtén la ecuación de la recta de regresión considerando la variable Y como variable dependiente de X.

  El modelo sería $Y = \beta_0 + \beta_1 X$

  El coeficiente de correlación correlación: $r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{15}{3*5} = 1$

  Los parámetros del modelo:

$$\beta_1 = r_{XY}\frac{s_Y}{s_X} = 1 * \frac{3}{5} = 0{,}6$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X} = 10{,}6 - 0{,}6 * 1 = 10$$

  Quedaría $Y = 10 + 0{,}6 * X$

  La pendiente es positiva, por lo cual la relación lineal es positiva, a mayor número de coches disponibles en la vivienda, mayor número de días al año que pasan de vacaciones en el extranjero.

- Según este modelo, ¿cuántos días al año pasará una familia que dispone de tres coches, en el extranjero?

$$Y = 10 + 0{,}6 * 3 = 11{,}8 \approx 12$$

Pasarán al año aproximadamente 12 días de vacaciones en el extranjero.