



# EJERCICIO DE ESTADÍSTICA BIDIMENSIONAL

1.º de Bachillerato

## Enunciado

En una clase de 20 alumnos se quiere estudiar si existe relación entre el tiempo dedicado al estudio de un examen y la nota obtenida.

Se consideran las siguientes variables:

- **X**: horas de estudio
- **Y**: nota obtenida en el examen (sobre 10)

Los datos recogidos se resumen en la siguiente **tabla de frecuencias conjuntas**:

### Tabla de frecuencias conjuntas

X (horas) \ Y (nota)	4	6	8	Total
1	3	1	0	4
2	2	4	1	7
3	0	3	6	9
<b>Total</b>	5	8	7	20

---



# 1. Distribuciones marginales

## Distribución marginal de X

Se suman las frecuencias de cada fila:

X	Frecuencia
1	4
2	7
3	9
<b>Total</b>	<b>20</b>

---

## Distribución marginal de Y

Se suman las frecuencias de cada columna:

Y	Frecuencia
4	5
6	8



8      7

**Total    20**

## 2. Distribución condicional de Y sabiendo X = 2

La frecuencia condicional se calcula con la fórmula:

$$f(Y | X = 2) = \text{frecuencia conjunta} / \text{frecuencia total de } X = 2$$

Como hay 7 alumnos que han estudiado 2 horas:

| Y | Frecuencia | f(Y | X = 2) |

|—|———|———|

| 4 | 2 | 2 / 7 ≈ 0,286 |

| 6 | 4 | 4 / 7 ≈ 0,571 |

| 8 | 1 | 1 / 7 ≈ 0,143 |

## 3. Diagrama de dispersión

Para construir el **diagrama de dispersión**, se representan en un sistema de ejes cartesianos los pares (X,Y), repitiendo cada punto tantas veces como indique su frecuencia.

Los puntos representados son:

- (1,4) → 3 veces
- (1,6) → 1 vez
- (2,4) → 2 veces
- (2,6) → 4 veces



- (2,8) → 1 vez
- (3,6) → 3 veces
- (3,8) → 6 veces

Al representar estos puntos se observa que, a medida que aumentan las horas de estudio, las notas tienden a ser mayores, lo que indica una **relación positiva entre ambas variables**.

---

## 4. Cálculo de las medias

### Media de X

Se multiplica cada valor por su frecuencia y se divide entre el total:

$$\bar{X} = (1 \cdot 4 + 2 \cdot 7 + 3 \cdot 9) / 20$$

$$\bar{X} = 45 / 20 = \mathbf{2,25}$$

---

### Media de Y

$$\bar{Y} = (4 \cdot 5 + 6 \cdot 8 + 8 \cdot 7) / 20$$

$$\bar{Y} = 124 / 20 = \mathbf{6,2}$$

---

## 5. Covarianza

La covarianza se calcula con la fórmula:

$$\text{Cov}(X,Y) = (1 / N) \cdot \sum n_i (x_i - \bar{X})(y_i - \bar{Y})$$

Se calcula para cada pareja de valores (X,Y), se multiplica por su frecuencia y se suman todos los resultados.

Ejemplo:

Para X = 1, Y = 4, frecuencia = 3:



$$(1 - 2,25)(4 - 6,2) \cdot 3$$

$$= (-1,25)(-2,2) \cdot 3$$

$$= 8,25$$

Sumando todos los productos y dividiendo entre 20:

$$\text{Cov}(X,Y) = \mathbf{0,975}$$

---

## 6. Varianzas y desviaciones típicas

### Varianza y desviación típica de X

$$\sigma^2_x = (1 / 20)[$$

$$4(1 - 2,25)^2 + 7(2 - 2,25)^2 + 9(3 - 2,25)^2$$

]

$$\sigma^2_x \approx 0,54$$

$$\sigma_x = \sqrt{0,54} \approx \mathbf{0,73}$$

---

### Varianza y desviación típica de Y

$$\sigma^2_y = (1 / 20)[$$

$$5(4 - 6,2)^2 + 8(6 - 6,2)^2 + 7(8 - 6,2)^2$$

]

$$\sigma^2_y \approx 2,34$$

$$\sigma_y = \sqrt{2,34} \approx \mathbf{1,53}$$

---

## 7. Coeficiente de correlación lineal

La fórmula es:

$$r = \text{Cov}(X,Y) / (\sigma_x \cdot \sigma_y)$$



$$r = 0,975 / (0,73 \cdot 1,53)$$

$$r \approx \mathbf{0,87}$$

**Interpretación:** existe una correlación lineal positiva fuerte entre las horas de estudio y la nota obtenida.

---

## 8. Rectas de regresión

### Recta de regresión de Y sobre X

Pendiente:

$$a = \text{Cov}(X,Y) / \sigma^2_x = 0,975 / 0,54 \approx 1,83$$

Ordenada en el origen:

$$b = \bar{Y} - a \cdot \bar{X}$$

$$b = 6,2 - 1,83 \cdot 2,25 \approx 2,08$$

**Recta:**

$$Y = 1,83X + 2,08$$

---

### Recta de regresión de X sobre Y

Pendiente:

$$a' = \text{Cov}(X,Y) / \sigma^2_y = 0,975 / 2,34 \approx 0,42$$

Ordenada en el origen:

$$b' = \bar{X} - a' \cdot \bar{Y}$$

$$b' = 2,25 - 0,42 \cdot 6,2 \approx -0,35$$

**Recta:**

$$X = 0,42Y - 0,35$$

---



## Conclusión

El estudio muestra una clara relación entre las horas de estudio y la nota obtenida.

El diagrama de dispersión, la covarianza y el coeficiente de correlación confirman que la relación es **positiva y fuerte**, por lo que es razonable utilizar las rectas de regresión para hacer predicciones.