ELEMENTOS DE PROBABILIDAD Y ESTADÍSTICA

1. **Las estaturas en centímetros y pesos en kilogramos de 10 jugadores de baloncesto son:**

Estatura (X)	186	189	190	192	193	193	198	201	203	205
Peso (Y)	85	85	86	90	87	91	93	103	100	101

a) **Hallar la recta de regresión de Y sobre X.**
b) **Hallar el coeficiente de correlación.**
c) **¿Qué estatura, se puede predecir, según la recta obtenida, de un jugador que pesa 92 kilogramos? ¿Y de uno que pesa 103 kilogramos?**
d) **Interpreta la bondad de estas predicciones en función del coeficiente de correlación.**

a) Necesitamos calcular los siguientes datos:

- Media de X $(\overline{x})$
- Media de Y $(\overline{y})$
- Desviación típica de x (s_x)
- Desviación típica de y (s_y)
- Covarianza (s_{xy})

Se tiene:

$$\overline{x} = \frac{\sum x_i f_i}{N} = \frac{186+189+190+192+193+193+198+201+203+205}{10} = 195 \quad cm$$

$$\overline{y} = \frac{\sum y_i f_i}{N} = \frac{85+85+86+90+87+91+93+103+100+101}{10} = 92{,}1 \quad kg.$$

$$s_x = \sqrt{\frac{\sum x_i^2 f_i}{N} - \overline{x}^2} =$$

$$= \sqrt{\frac{186^2+189^2+190^2+192^2+193^2+193^2+198^2+201^2+203^2+205^2}{10} - 195^2} = 6{,}07$$

$$s_y = \sqrt{\frac{\sum y_i^2 f_i}{N} - \overline{y}^2} =$$

$$= \sqrt{\frac{85^2+85^2+86^2+90^2+87^2+91^2+93^2+103^2+100^2+101^2}{10} - 92{,}1^2} = 6{,}56$$

$$s_{xy} = \frac{\sum x_i y_j f_{ij}}{N} - \overline{xy} =$$

$$= \frac{\begin{array}{c}186\cdot 85+189\cdot 85+190\cdot 86+192\cdot 90+193\cdot 87+\\ 193\cdot 91+198\cdot 93+201\cdot 103+203\cdot 100+202\cdot 101\end{array}}{10} - 195\cdot 92{,}1 =$$

$$= \frac{179971}{10} - 17959{,}5 = 37{,}6$$

Con todo ello, la recta de regresión de Y sobre X es:

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x}) \Rightarrow y-92{,}1=\frac{37{,}6}{6{,}07^2}(x-195) \Rightarrow y-92{,}1=1{,}02(x-195) \Rightarrow$$

$$\Rightarrow y=1{,}02x-106{,}8$$

Que es la forma explícita de la recta pedida:

b) Para el coeficiente de correlación tenemos:

$$r=\frac{s_{xy}}{s_x \cdot s_y}=\frac{37{,}6}{6{,}07\cdot 6{,}56}=0{,}94$$

c) En la recta de regresión obtenida en a) para y=92 kg., se espera una estatura de:

$$92=1{,}02x-106{,}8 \Rightarrow x=\frac{92+106{,}8}{1{,}02}=194{,}9 \quad cm$$

Para y=103 kg. se espera:

$$103=1{,}02x-106{,}8 \Rightarrow x=\frac{103+106{,}8}{1{,}02}=205{,}69 \quad cm$$

d) Dado el coeficiente de correlación obtenido en b) positivo (correlación directa) y muy próximo a 1, se puede afirmar que los valores obtenidos en c) son muy buenos.

---

2. Se han pasado dos test A y B a un mismo grupo de personas obteniendo los siguientes parámetros: medias aritméticas 56,8 y 62,2 (respectivamente), desviaciones típicas 15,6 y 10,8 (respectivamente) y covarianza 164,4

a) Si una persona ha obtenido 60 en el test A y 64 en el B ¿En qué test ha obtenido una puntuación mayor en relación al grupo?
b) Calcula el coeficiente de correlación lineal e interprétalo.
c) Utiliza la recta de regresión apropiada para predecir qué puntuación debe obtener una persona en el test B si en el A ha obtenido 73

a) Vamos a tipificar las variables X (puntuación del test A) e Y (puntuación en el test B) para tener acceso a las tablas de la distribución normal N(1, 0). Calculando con ella las probabilidades de que una persona cualquiera haya obtenido puntuaciones por debajo de 60 en el A y de 64 en el B sabremos en cuál de los dos la puntuación obtenida es más significativa.

Se tienen los siguientes datos:

$$\bar{x}=56{,}8$$
$$\bar{y}=62{,}2$$
$$s_x=15{,}6$$
$$s_y=10{,}8$$
$$s_{xy}=164{,}64$$

Tipificando las puntuaciones obtenidas por nuestro sujeto con la expresión:

$$z_1=\frac{x-\bar{x}}{s_x}=\frac{60-56{,}8}{15{,}6}=0{,}205$$

$$z_2=\frac{y-\bar{y}}{s_y}=\frac{64-62{,}2}{10{,}8}=0{,}166$$

Con las tablas de la normal obtenemos:

$$p(X \le 60) = p(z \le z_1) = 0{,}5987$$
$$p(Y \le 64) = p(z \le z_2) = 0{,}5636$$

O sea, es más probable que una persona obtenga una puntuación por debajo de 60 en el test A que obtenga una por debajo de 64 en el test B, luego la puntuación obtenida es mayor con relación al grupo A que al B

b) El coeficiente de correlación es:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{164{,}64}{15{,}6 \cdot 10{,}8} = 0{,}97$$

Siendo la correlación directa (r>0) y muy significativa (r próximo a 1)

c) Calcularemos la recta de regresión de Y sobre X para despejar en ella directamente la puntuación correspondiente del test A. Será:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \Rightarrow y - 62{,}2 = \frac{164{,}64}{15{,}6^2}(x - 56{,}8) \Rightarrow y - 62{,}2 = 0{,}68(x - 56{,}8) \Rightarrow$$
$$\Rightarrow y = 0{,}68x + 23{,}58$$

Y para x=73

$$y = 0{,}68 \cdot 73 + 23{,}58 = 73{,}22$$

---

3. Las notas obtenidas por cinco alumnos en Latín y Griego son:

Latín	6	4	8	5	3,5
Griego	6,5	4,5	7	5	4

Determinar la recta de regresión de Y sobre X y calcular la nota esperada en Griego de un alumno que tiene 7,5 en Latín.

a) r=1 significa que la correlación es directa (al crecer X crece Y y viceversa) y que ambas variables son totalmente dependientes (dependencia funcional).
b) r=-1 igual que antes (dependencia funcional) pero con correlación inversa (al crecer X decrece Y y viceversa).
c) R=0,75 la correlación es directa pero poco significativa, apenas hay dependencia entre las variables (empieza a ser considerada a partir de r=0,8)

Para la segunda parte del problema determinamos:

$$\bar{x} = \frac{\sum x_i f_i}{N} = \frac{6 + 4 + 8 + 5 + 3{,}5}{5} = 5{,}3$$
$$\bar{y} = \frac{\sum y_i f_i}{N} = \frac{6{,}5 + 4{,}5 + 7 + 5 + 4}{5} = 5{,}4$$
$$s_x = \sqrt{\frac{\sum x_i^2 f_i}{N} - \bar{x}^2} = \sqrt{\frac{6^2 + 4^2 + 8^2 + 5^2 + 3{,}5^2}{5} - 5{,}3^2} = 1{,}16$$
$$s_{xy} = \frac{\sum x_i y_j f_{ij}}{N} - \overline{xy} = \frac{1}{5}\left(6 \cdot 6{,}5 + 4 \cdot 4{,}5 + 8 \cdot 7 + 5 \cdot 5 + 3{,}5 \cdot 4\right) - 5{,}3 \cdot 5{,}4 = 1{,}78$$

La recta pedida es:

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}\left(x-\bar{x}\right)\Rightarrow y-5{,}4=\frac{1{,}78}{1{,}16^2}\left(x-5{,}3\right)\Rightarrow y-5{,}4=1{,}32(x-5{,}3)\Rightarrow$$

$$\Rightarrow y=1{,}32x-1{,}6$$

Y la puntuación estimada en Latín para una de Griego de 7,5 es:

$$7{,}5=1{,}32x-1{,}6\Rightarrow x=\frac{7{,}5+1{,}6}{1{,}32}=6{,}89$$

---

4. Las notas obtenidas por 10 alumnos en matemáticas y en inglés son:

Alumnos	1	2	3	4	5	6	7	8	9	10
Matemáticas.	6	4	8	5	3,5	7	5	10	5	4
Inglés	6,5	4,5	7	5	4	8	7	10	6	5

a) Calcular la covarianza, las varianzas y el coeficiente de correlación lineal.
b) ¿Existe correlación entre las dos variables? Razónese la respuesta.
c) Calcular la recta de regresión. ¿Cuál sería la nota esperada en inglés para un alumno que hubiera obtenido un 8,3 en matemáticas?

a) Se tiene:

$$\bar{x}=\frac{\sum x_i f_i}{N}=\frac{6+4+8+5+3{,}5+7+5+1+5+4}{10}=\frac{57{,}5}{10}=5{,}75$$

$$\bar{y}=\frac{\sum y_i f_i}{N}=\frac{6{,}5+4{,}5+7+5+4+8+7+10+6+5}{10}=\frac{63}{10}=6{,}3$$

$$s_x^2=\frac{\sum x_i^2 f_i}{N}-\bar{x}^2=\frac{36+16+64+25+12{,}25+49+25+100+25+16}{10}-33{,}0625=$$

$$=\frac{368{,}25}{10}-33{,}0625=36{,}825-33{,}0625=3{,}7625$$

$$s_y^2=\frac{\sum y_i^2 f_i}{N}-\bar{y}^2=\frac{42{,}25+20{,}25+49+25+16+64+49+100+36+25}{10}-39{,}69=$$

$$=\frac{426{,}5}{10}-39{,}69=42{,}65-39{,}69=2{,}96$$

$$s_{xy}=\frac{\sum x_i y_j f_{ij}}{N}-\overline{xy}=\frac{39+18+56+25+14+56+35+100+30+20}{10}-5{,}75\cdot 6{,}3=$$

$$\frac{393}{10}-36{,}225=39{,}3-36{,}225=3{,}075$$

$$r=\frac{s_{xy}}{s_x\cdot s_y}=\frac{3{,}075}{\sqrt{3{,}7625}\cdot\sqrt{2{,}96}}=\frac{3{,}075}{1{,}94\cdot 1{,}72}=0{,}92$$

b) La correlación que existe es directa (r>0) y muy significativa (r próximo a 1).

c) La recta de regresión de Y sobre X es:

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}\left(x-\bar{x}\right)\Rightarrow y-6{,}3=\frac{3{,}075}{3{,}7625}(x-5{,}75)\Rightarrow y-6{,}3=0{,}82(x-5{,}75)\Rightarrow$$

$$\Rightarrow y=0{,}82x+1{,}59$$

Y la nota esperada en inglés para un alumno que tuviera 8,3 en matemáticas (x=8,3) sería:

$$y=0{,}82\cdot 8{,}3+1{,}59=8{,}4$$

---

5. La nota media del expediente (X) y la nota obtenida en las pruebas de acceso (Y) de ocho personas ha sido:

X	**6,24**	**7,91**	**7,04**	**6,13**	**6,38**	**6,48**	**6,44**	**5,99**
Y	**4,20**	**4,65**	**6,51**	**6,73**	**5,20**	**4,60**	**5,69**	**3,42**

a) Obtener el coeficiente de correlación lineal entre las dos variables e interpretar el resultado.
b) Calcular la recta de regresión de Y sobre X.
c) Según el ajuste que ofrece la recta de regresión, ¿Qué nota sería esperable que obtuviera en las pruebas de acceso una persona con nota media de expediente de 7,31?

a) Procediendo en los cálculos como en problemas similares anteriores se llega a los siguientes resultados:

$$\bar{x}=\frac{\sum x_i f_i}{N}=6{,}58$$

$$\bar{y}=\frac{\sum y_i f_i}{N}=5{,}13$$

$$s_x=\sqrt{\frac{\sum x_i^2 f_i}{N}-\bar{x}^2}=0{,}54$$

$$s_y=\sqrt{\frac{\sum y_i^2 f_i}{N}-\bar{y}^2}=1{,}04$$

$$s_{xy}=\frac{\sum x_i y_j f_{ij}}{N}-\overline{xy}=0{,}02$$

$$r=\frac{0{,}02}{0{,}54\cdot 1{,}04}=0{,}04$$

El resultado obtenido para r indica una correlación insignificante (prácticamente nula) entre ambas variables. Al ser r positivo la correlación es directa.

b) La recta de regresión será:

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x})\Rightarrow y-5{,}13=\frac{0{,}02}{0{,}54^2}(x-6{,}58)\Rightarrow y=0{,}07x+4{,}68$$

c) Para x=7,31, la recta obtenida da:

$$y=0{,}07\cdot 7{,}31+4{,}68=5{,}19$$
