INTRODUCCIÓN A LA ESTADÍSTICA
LADE, LEC, LADE-DER, LEC-DER, PER-DER
Solución al examen del 15 de Enero de 2009

Problema 1.

a) En la tabla aparecen las frecuencias absolutas n_{ij}:

Edad \ Periódico	El Universal	Últimas Noticias	TOTAL
[15, 25)	4	6	10
[25, 40)	32	10	42
[40, 60)	48	60	108
[60, 80)	16	24	40
TOTAL	100	100	200

Las frecuencias relativas se calculan: $f_{ij} = n_{ij}/n$.

Edad \ Periódico	El Universal	Últimas Noticias	TOTAL
[15, 25)	0.02	0.03	0.05
[25, 40)	0.16	0.05	0.22
[40, 60)	0.24	0.30	0.54
[60, 80)	0.08	0.12	0.20
TOTAL	0.5	0.5	1

b) Las distribuciones de frecuencias marginales son las que aparecen en los márgenes de la tabla anterior:

Periódico	El Universal	Últimas Noticias	TOTAL
$f_{\bullet j}$	0.5	0.5	1

Edad	[15, 25)	[25, 40)	[40, 60)	[60, 80)	TOTAL
$f_{i\bullet}$	0.05	0.22	0.54	0.20	1

c) Se trata de hallar la distribución de frecuencias de la variable "Edad" condicionada a que la variable "Periódico leído" toma el valor "El Universal". Las frecuencias de la variable X condicionada a $Y = j$ se obtienen: $f_{i|Y=j} = \frac{f_{ij}}{f_{\bullet j}}$.

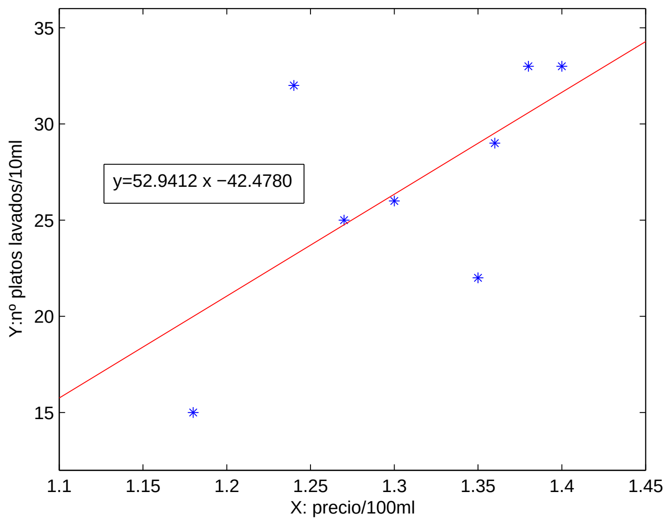
Edad	[15, 25)	[25, 40)	[40, 60)	[60, 80)	TOTAL
$f_{i\|Y=EU}$	0.04	0.32	0.48	0.16	1

d) Media para datos agrupados:

$$\overline{x}_{|Y=EU} = 0.04 \times 20 + 0.32 \times 32.5 + 0.48 \times 50 + 0.16 \times 70 = 46.4$$

Problema 2.

a) Diagrama de dispersión (y recta de regresión calculada en c)):



b) Coeficiente de correlación: $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$.

$$\begin{aligned}
s_{xy} &= \tfrac{1}{8}\textstyle\sum_{i=1}^{8} x_i \cdot y_i - \overline{x} \cdot \overline{y} = \frac{283.81}{8} - \frac{10.48}{8} \cdot \frac{215}{8} = 0.27 \\
s_x &= \sqrt{\tfrac{1}{8}\textstyle\sum_{i=1}^{8} x_i^2 - \overline{x}^2} = \sqrt{\frac{13.77}{8} - \left(\frac{10.48}{8}\right)^2} = 0.0714 \\
s_y &= \sqrt{\tfrac{1}{8}\textstyle\sum_{i=1}^{8} y_i^2 - \overline{y}^2} = \sqrt{\frac{6053}{8} - \left(\frac{215}{8}\right)^2} = 5.8617 \\
r_{xy} &= \frac{s_{xy}}{s_x \cdot s_y} = \frac{0.27}{0.0714 \cdot 5.8617} = 0.6451
\end{aligned}$$

Al ser el coeficiente de correlación lineal positivo pero no muy próximo a 1, podemos decir que la asociación lineal entre las dos variables es creciente pero no muy fuerte.

c) Ecuación de la recta de regresión: $y = ax + b$ donde $a = \frac{s_{xy}}{s_x^2}$ y $b = \overline{y} - a\overline{x}$.

$$\begin{aligned}
a &= \frac{s_{xy}}{s_x^2} = \frac{0.27}{0.0714^2} = 52.9412 \\
b &= \overline{y} - a\overline{x} = 26.8750 - 52.9412 \cdot 1.31 = -42.4780
\end{aligned}$$

La recta de regresión es: $y = 52.9412 \cdot x - 42.4780$.

d) El producto número 4 (1.24, 32), ya que es el que tiene mejor relación calidad/precio: ningún producto más caro tiene un poder de limpieza significativamente mayor, y el único producto más barato tiene un poder de limpieza muchísimo menor.

Problema 3.

a) Por tratarse de una función de densidad

$$1 = \int_{\mathbb{R}} f(x)dx = \int_{100}^{200} kdx = [kx]_{100}^{200} = 200k - 100k = 100k \quad \Longrightarrow k = \frac{1}{100}.$$

b)

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} \int_{-\infty}^{x} 0\,du = 0 & \text{si} \quad x < 100 \\ \int_{-\infty}^{100} 0\,du + \int_{100}^{x} \frac{1}{100}\,du = 0 + \left[\frac{u}{100}\right]_{100}^{x} = \frac{x-100}{100} & \text{si} \quad 100 \leq x \leq 200 \\ \int_{-\infty}^{100} 0\,du + \int_{100}^{200} \frac{1}{100}\,du + \int_{200}^{\infty} 0\,du = 1 & \text{si} \quad x > 200 \end{cases}$$

c) Por tratarse de una distribución uniforme $E[X] = \frac{100+200}{2} = 150$ (ó $E[X] = \int_{100}^{200} x \cdot f(x) = 150$).
Por ser una distribución simétrica $Me_X = E[X] = 150$ (ó $0.5 = F(Me_X) = \frac{Me_X - 100}{100} \Rightarrow Me_X = 150$).

d) $P(X < 125) = F(125) = \frac{125-100}{100} = 0.25$

Problema 4. X = "puntuación en el test de habilidades". $X \sim N(95, 10) \Rightarrow Z = \frac{X-95}{10} \sim N(0, 1)$.

a) $P(X > 114.6) = P(Z > \frac{114.6-95}{10}) = P(Z > 1.96) = P(Z < -1.96) = 0.025$.

b) $P(X > 120) = P(Z > \frac{120-95}{10}) = P(Z > 2.5) = P(Z < -2.5) = 0.0062$.

c) $P(85 < X < 120) = P(\frac{85-95}{10} < Z < \frac{120-95}{10}) = P(-1 < Z < 2.5) = P(Z < 2.5) - P(Z < -1) = 0.9938 - 0.1587 = 0.8351$.

d) Teniendo en cuenta la simetria de la distribucion normal respecto a su media y el apartado a), tenemos que $P(X > 114.6) = 0.025$ y como el punto simétrico de 114.6 con respecto a 95 es 75.4 entonces $P(X < 75.4) = 0.025$.
El intervalo buscado es $(75.4, 114.6)$.

e) $P(X < 85) = P(Z < \frac{85-95}{10}) = P(Z < -1) = 0.1587$. Por tanto, el 15.87% de los alumnos necesitaría asistencia al aula de refuerzo. El numero esperado de alumnos sería: $0.1587 \times 500 = 79.35$, lo que implicaría la necesidad de 4 aulas de refuerzo. Por lo tanto el presupuesto esperado sería 4×1000 euros $= 4000$ euros.