Dpto. Matemática Aplicada y Estadística

Grado en IIAA y Grado en IHJ
Asignatura: **Estadística Aplicada. Curso 2011-2012**
Asignatura: Estadística Aplicada. **Problemas del Modelo B.**

PROBLEMA 1.- El volumen máximo de agua alcanzado en un embalse durante 30 años hidrológicos se recoge en la siguiente tabla:

50	51	51	52	53	54	54	55	55	56
57	58	59	60	61	62	63	67	68	69
70	73	74	74	80	84	86	94	99	110

1. Realizar el diagrama de caja y bigotes para estos datos. Colocar en cada línea del gráfico su valor numérico. Comentar las características más relevantes del diagrama. Determinar, a partir de los valores observados, entre qué valores se encuentran aquellas observaciones que pueden considerarse no atípicas.
2. Agrupar los datos en intervalos de la misma amplitud y representar gráficamente la distribución de frecuencias. Describir las características más relevantes de dicha gráfica. ¿Qué medidas de posición central y de dispersión son más adecuadas para resumir los datos? Razona tu respuesta.
3. Calcular la media y la desviación típica del volumen máximo de agua alcanzado.
4. ¿En qué intervalo se sitúa la mediana? ¿Cuál es el intervalo modal?
5. Para que un año hidrológico no se considere seco el volumen de agua alcanzado debe de estar por encima de 70, ¿cuál es el porcentaje de años hidrológicamente secos de la muestra?

PROBLEMA 3.- Se estudia la relación entre la renta familiar disponible, y , y el consumo familiar en alimentación, c, (tanto renta como consumo expresados en miles de euros anuales). Para una muestra de 25 familias de una determinada ciudad, se obtuvo los siguientes valores muestrales:

$$\overline{c} = 14 \quad s_c = 6$$
$$\overline{y} = 24 \quad s_y = 7$$

Como conclusión de estudio se establece el siguiente modelo lineal:

$$y = b + 1.1 \times \ c$$

1. Calcula el valor de la covarianza muestral entre las variables c e y.
2. Obtener el coeficiente de determinación y comentar la bondad del ajuste.
3. Para una renta anual de 25000 euros, ¿cuál sería el consumo en alimentación estimado por el modelo?

PROBLEMA3.- Los ladrillos fabricados por una empresa pueden tener dos tipos de defectos D1 y D2. El defecto D1 ocurre con una probabilidad de 0.09, el defecto D2 con una probabilidad de 0.05 y en el 3% de los ladrillos ocurren ambos tipos de defectos. Se pide:

1. Calcular la probabilidad de que al seleccionar un ladrillo al azar de la producción sea defectuoso (esto es, tenga al menos un tipo de defecto).
2. Calcular la probabilidad de que el ladrillo seleccionado tenga el defecto D2 y no tenga el defecto D1.
3. Cada ladrillo se somete de manera automática a un test de ruptura, con las siguientes probabilidades:
    - Si el ladrillo tiene alguno de los defectos tiene una probabilidad de romperse del 0.99

- Si el ladrillo no tiene ninguno de los defectos tiene una probabilidad de romperse de 0.04.

Se pide:

3.1. Calcular la probabilidad de que un ladrillo seleccionado al azar se rompa durante el test.

3.2. Si un ladrillo escogido al azar no se ha roto durante el test, ¿cuál es la probabilidad de que sea defectuoso?

PRUEBA ESCRITA 1: $\boxed{\text{MODELO B}}$

PROBLEMA 1:

X ≡ Volumen de agua alcanzado en el embalse

$n = 30$

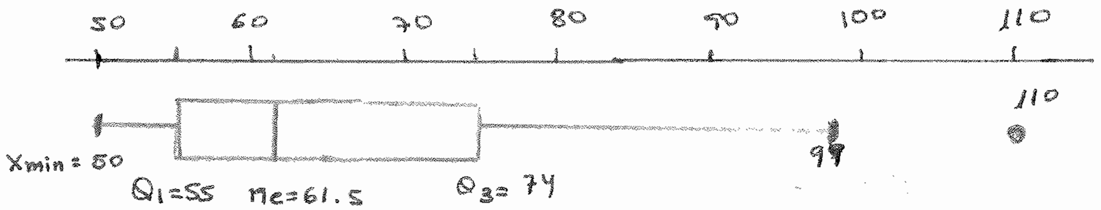
(1): $Me = \dfrac{X_{(15)} + X_{(16)}}{2} = \dfrac{61+62}{2} = 61.5$

$$\left.\begin{array}{l} Q_1 = 55 \\ Q_3 = 74 \end{array}\right| \Rightarrow RIQ = 19$$

$$\begin{cases} L_{INF.} = 55 - 1.5 * 19 = 26.5 \\ L_{SUP} = 74 + 1.5 * 19 = 102.5 \end{cases}$$

⇒ los $x_i \notin [26.5, 102.5]$ son datos atípicos

⇒ 110 es un dato atípico.



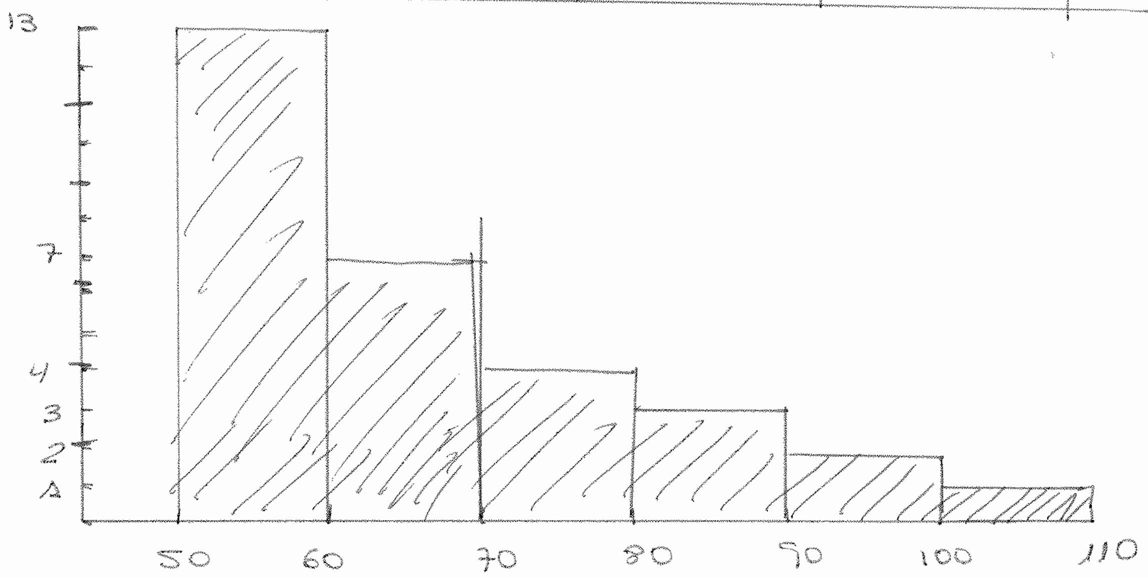
→ ∃ un dato atípico, asimetría a la derecha. Presenta mayor variabilidad en la zona alta de la tabla.

→ A partir de los valores observados, los datos no atípicos se encuentran en el intervalo $[26.5, 102.5]$

② : $n = 30 \Rightarrow \bullet\ k \geq \sqrt{30} \simeq 5.5 \Rightarrow k = 6$

$\bullet\ h = \dfrac{110 - 50}{6} = 10$

límites	m_i	n_i	f_i	N_i
[50, 60)	55	13	0.43	13
(60, 70	65	7	0.23	20
(70, 80]	75	4	0.13	24
(80, 90]	85	3	0.10	27
(90, 100]	95	2	0.07	29
(100, 110]	105	1	0.03	30
TOTALES	–	30	1	–



→ Unimodal, asimétrico a la derecha.

→ No se percibe el dato atípico, pero en el apartado anterior hemos obtenido que $x = 110$ es dato atípico.

⇒ Me = como medida de centro

RIQ = " " " dispersión

③ :- <u>Con datos AGRUPADOS:</u>

$$\bar{X} = \frac{1}{30}\left(55*13 + 65*7 + 75*4 + 85*3 + 95*2 + 105*1 \right) =$$

$$= \boxed{67.333 \ m^3}$$

$$S_X^2 = \frac{n}{n-1}\left(\overline{X^2} - \bar{X}^2 \right) = \frac{30}{29}\left[\frac{55^2*13 + 65^2*7 + 75^2*4 + 85^2*3 + 95^2*2 + 105^2}{30} \right.$$

$$\left. - (67.333)^2 \right] = 211.6152 \Rightarrow \boxed{S_X = 14.547 \ m^3}$$

También se puede calcular <u>con los datos NO AGRUPADOS:</u>

$$\bar{X} = \frac{50 + 51 + 51 + \cdots + 94 + 99 + 110}{30} = \boxed{66.633 \ m^3}$$

$$S_X^2 = \frac{30}{29}\left(\frac{50^2 + 51^2 + 51^2 + \cdots + 94^2 + 99^2 + 110^2}{300} - (66.633)^2 \right) =$$

$$= 238.10 \Rightarrow \boxed{S_X = 15.431 \ m^3}$$

④ :-
- Int. mediano = [60, 70)
- Int. modal = [50, 60)

⑤ :- <u>Con datos agrupados</u> :

% de años secos = % de $X \leq 70$ =

$$= \frac{(13+7)}{30} * 100\,\% = 66.67\%.$$

Datos no agrupados $= \frac{21}{30} * 100\% = 70\%$

PROBLEMA 2:

y ≡ Renta familiar
c ≡ Consumo alimentación

$\Rightarrow \bar{c} = 14,\ S_c = 6$
$\bar{y} = 24,\ S_y = 7$

$$y = b + 1.1 * c$$

(a): Covarianza $= S_{c,y} = ?$

$$\hat{a} = 1.1 = \frac{S_{c,y}}{S^2_c} = \frac{S_{c,y}}{6^2} \Rightarrow \boxed{S_{c,y} = 36 * 1.1 = 39.6}$$

(b): Coef. de determinación?

$$R^2 = \frac{S^2_{c,y}}{S^2_c * S^2_y} = \frac{(39.6)^2}{6^2 * 7^2} = 0.89$$

Ajuste bueno

(c): $y = 25 * 10^3$ € $\Rightarrow$ ¿ $\hat{c}$?

Necesitamos ajustar $c = f(y) = \alpha + \beta y$

$$\hat{\beta} = \frac{S_{c,y}}{S^2_y} = \frac{39.6}{49} = 0.81$$

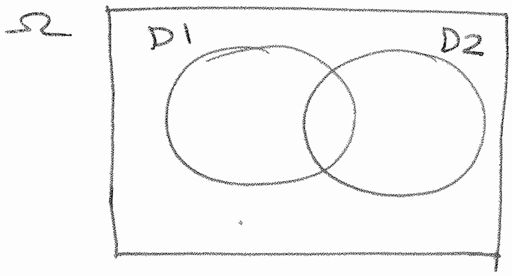
$$\hat{\alpha} = \bar{c} - \hat{\beta} \cdot \bar{y} = 14 - 0.81 * 24 = -5.44$$

$$\Rightarrow c = -5.44 + 0.81 y$$

De donde, $\hat{c} = -5.44 + 0.81 * 25 = 14.81 * 10^3$ €

Problema 3 :

$$\begin{cases} D1 \rightarrow p(D1) = 0.09 \\ D2 \rightarrow p(D2) = 0.05 \\ \rightarrow p(D1 \cap D2) = 0.03 \end{cases}$$



(1) : $p(\overbrace{D1 \cup D2}^{D}) = p(D1) + p(D2) - p(D1 \cap D2) =$

$= 0.09 + 0.05 - 0.03 = \boxed{0.11}$

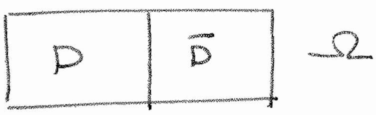
(2) : $p(D2 \cap \overline{D1}) = p(D2) - p(D1 \cap D2) =$

$= 0.05 - 0.03 = \boxed{0.02}$

(3) : $R \equiv$ El ladrillo se rompe.

$p(R|D) = 0.99$

$p(R|\overline{D}) = 0.04$



$p(D) = 0.11$
$p(\overline{D}) = 0.89$

(3.1) : $p(R) = p(D)\,p(R|D) + p(\overline{D})\,p(R|\overline{D}) =$

$= 0.11 * 0.99 + 0.89 * 0.04 = \boxed{0.1445}$

(3.2) : $p(D|\overline{R}) = \dfrac{p(D \cap \overline{R})}{p(\overline{R})} =$

$= \dfrac{p(D)\,p(\overline{R}|D)}{p(\overline{R})} = \dfrac{p(D)(1 - p(R|D))}{1 - p(R)} = \dfrac{0.11 * (1 - 0.99)}{1 - 0.1445} =$

$= \boxed{0.0013}$