Dpto. Matemática Aplicada y Estadística

Grado en IIAA y Grado en IHJ
Asignatura: **Estadística Aplicada. Curso 2011-2012**
Febrero 2012

NOMBRE:...**APELLIDOS:**..
ESPECIALIDAD:..

1. **[1 punto]** En la tabla adjunta se presentan los datos relativos al volumen de agua (en miles de litros) alcanzado en un embalse durante 30 años hidrológicos:

Volumen de agua	Nº de años
$[60, 77]$	12
$(77, 94]$	10
$(94, 111]$	5
$(111, 128]$	3

Se pide:

(a) Calcular el porcentaje de años en los que el volumen de agua es mayor a 85000 litros. (**0.25 puntos**)

(b) Si un año lo clasificamos en el grupo en el que se encuentra el 50% de los años con menos volumen de agua en el ambalse, ¿cuál será el máximo de volumen de agua que habrá en el embalse? (**0.25 puntos**)

(c) Se quiere estudiar la relación entre el volumen de agua alcanzado en el embalse y la precipitación anual (en l/m^2) caída en la zona. Para ello, se observan para estos 30 años, conjuntamnete el volumen de agua embalsada y la cantidad de precipitación anual caída en la zona. El estudio concluye con la siguiente relación:

$$V = -42.732 + a * P$$

donde V= volumen y P= precipitación.
Sabiendo que

$$s_v^2 = 318.12 \qquad s_p^2 = 510.27$$
$$s_{v,p} = 366.906$$

(c1) Determinar el valor de la pendiente de la recta. (**0.25 puntos**)

(c2) Determinar el coeficiente de determinación y comentar la bondad del ajuste realizado. (**0.25 puntos**)

2. **[0.75 puntos]** En una cooperativa agrícola se comercializa la producción de cítricos de tres fincas A, B y C. Las fincas B y C tiene el mismo volumen de producción, mientras que la finca A produce lo mismo que las otras dos fincas juntas. Por experiencias previas, se sabe que aproximadamente el 5% de la fruta recolectada por la finca A se encuentra en mal estado, el 4% para la finca B y el 3% para la finca C. Se elige una pieza de fruta al azar de la producción, se pide:

(a) Calcular la probabilidad de que la fruta no se encuentre en mal estado. (**0.25 puntos**)

(b) Se comprueba que la fruta seleccionada se encuentre en mal estado, ¿de cuál de las tres fincas es más probable que proceda y con qué probabilidad? (**0.25 puntos**)

(c) Determinar la probabilidad de que la pieza de fruta no esté en mal estado y no proceda de la finca A. (**0.25 puntos**)

3. **[1.5 puntos]** El peso de los huevos que ponen las gallinas de una determinada granja se puede modelizar como una distribución desconocida de media 61 gramos y de desviación típica 4 gramos. En toda la Unión Europea, para que un huevo sea clasificado de categoría mediana su peso debe de situarse en el intervalo (53 gr., 69 gr.) Se pide:

(a) ¿Podrías dar una cota de la proporción de huevos que son clasificados de categoría mediana en dicha granja? Razonar la respuesta. (**0.25 puntos**)

(b) Asumiendo a partir de este momento que el peso de los huevos que ponen las gallinas de dicha granja se puede modelizar como una distribución normal de media 61 gramos y de desviación típica 4 gramos, dar respuesta a las siguientes cuestiones

(b1) Determinar el porcentaje exacto de huevos de categoría mediana que se producen en la granja. (**0.25 puntos**)

(b2) Elegida una docena de huevos al azar, determinar la probabilidad de que al menos 10 sean de categoría mediana. (**0.5 puntos**)

(b3) Si al día se producen 500 huevos, determinar el número esperado de huevos de categoría mediana que se obtiene y la probabilidad de que entre los 500 a lo sumo 485 sean de categoría mediana. (**0.5 puntos**)

4. [**1.25 puntos**] El tiempo que transcurre antes de que una persona sea atendida en una cafetería es una variable aleatoria T con función de densidad:

$$f(t) = \begin{cases} 0, & t \leq 0 \\ k \cdot e^{-t/4}, & t > 0 \end{cases}$$

Se pide.

(a) Determinar el valor de k para que $f(t)$ sea una función de densidad. (**0.25 puntos**)

(b) Si se sabe que han pasado tres minutos y una determinada persona no ha sido atendida, determinar la probabilidad de que dicha persona sea atendida en los tres minutos siguientes.(**0.5 puntos**)

(c) El número de personas, al día, que en esta cafetería se olvida de pagar se puede modelizar como una distribución de Poisson de media 0.5. Si esta cafetería abre seis días a la semana, determinar la probabilidad de que en una semana el número de personas que se olvidan de pagar sea al menos 3 personas. (**0.5 puntos**)

5. [**1.5 puntos**] Cuando la cantidad de semillas se soja que quedan en el suelo después de pasar la cosechadora es menor de 80 semillas/m^2, la pérdida de producción en qq/ha. es aceptable. Un productor decide probar el funcionamiento de su máquina y para ello, después de cosechar una parcela cuenta en 10 unidades de un m^2 cada una, cuántas semillas quedan en el suelo después de cosechar. Los resultados en semillas/m^2 fueron:

77 73 82 82 79 81 78 76 76 75

Suponiendo que el número de semillas que quedan en el suelo después de cosechar sigue un modelo normal, se pide:

0.3 + 0.1 → (0.2 + 0.2)

(a) Construir de manera detallada un intervalo de confianza al 98% para el número medio de semillas que quedan después de cosechar ¿Cuál es el error de la estimación? (**0.5 puntos**)

(b) ¿Se puede concluir que la cosechadora está funcionando bien, esto es, la cantidad de semillas se soja promedio que quedan en el suelo después de pasar la cosechadora es menor de 80 semillas/m^2? Plantear y llevar a cabo el contraste adecuado para responder a esta pregunta tomando $\alpha = 0.05$ (**0.5 puntos**)

(c) Si se desea construir un intervalo de confianza al 98% para el número medio de semillas que quedan después de cosechar, qué tamaño de la muestra será necesario utilizar para estimar el número medio de semillas que quedan después de cosechar, con un error menor de 2.5 . (Tomar como desviación típica el valor $\sigma = 4$). (**0.5 puntos**)

ESTADÍSTICA APLICADA: Grado en IIAA y en IHJ. 1/

FEBRERO 2012

P.1:

m_i	Volumen = Y	Nº años = n_i	N_i
68.5	[60, 77]	12	12
85.5	(77, 94]	10	22
102.5	(94, 111]	5	27
119.5	(111, 128]	3	30
		$n = 30$	–

$Y > 80$

(a): % de $Y > 85 = \frac{(10+5+3)}{30} * 100\% =$

$= 60\%$

(b): El 50% de los años con menos volumen de agua en el embalse

$$\underbrace{Y_{min}, \ldots, }_{50\% \text{ de los datos}} Me, \ldots \ldots Y_{max}$$

$\frac{n}{2} = 15 \longrightarrow$ 1er Int-de clase tal que $N_i \geq 15$ es $(77, 94]$

$\Rightarrow$ El máximo volumen de agua que habrá en el embalse es 94000 litros.//

(c): $\underbrace{\text{Volumen}}_{V} = -42.732 + a * \underbrace{\text{precipitación}}_{P}$

$s_v^2 = 318.12$

$s^2_p = 510.27$

$s_{v,p} = 366.906$

(c.1): $\hat{a} = \dfrac{s_{v,p}}{s^2_p} = \dfrac{366.906}{510.27} = 0.719$

(c.2): $R^2 = \dfrac{s^2_{v,p}}{s_v^2 * s_p^2} = \dfrac{(366.906)^2}{318.12 * 510.27} = 0.829$

$\Rightarrow$ Ajuste bueno.

P. 2:

$$\begin{cases} A \longrightarrow p(A) = \alpha \\ B \longrightarrow p(B) = \beta \\ G \longrightarrow p(G) = \gamma \end{cases}$$

$$\left.\begin{array}{l} \alpha + \beta + \gamma = 1 \\ \beta = \gamma \\ \alpha = \beta + \gamma \end{array}\right\} \begin{array}{l} \longrightarrow 2\beta + \beta + \beta = 4\beta = 1 \Rightarrow \boxed{\beta = 1/4} \\ \rightsquigarrow \boxed{\gamma = 1/4} \\ \longrightarrow \alpha = \beta + \beta = 2\beta \Rightarrow \boxed{\alpha = \frac{1}{2}} \end{array}$$

$$\Rightarrow \begin{cases} p(A) = \frac{1}{2} \\ p(B) = 1/4 \\ p(G) = 1/4 \end{cases}$$

M = Fruta en mal estado.

$$\begin{cases} p(M/A) = 0.05 \\ p(M/B) = 0.04 \\ p(m/C) = 0.03 \end{cases}$$

(a): $p(\bar{M}) = 1 - p(M) =$

$= 1 - [p(A)\,p(m/A) + p(B)\,p(m/B) + p(C)\,p(m/C)] =$

$= 1 - \left[\frac{1}{2} * 0.05 + \frac{1}{4} * 0.04 + \frac{1}{4} * 0.03\right] =$

$= 1 - \frac{(0.1 + 0.04 + 0.03)}{4} = 1 - \frac{0.17}{4} = \boxed{0.9575}$

$\rightsquigarrow p(m) = 0.0425$

(b): $p(A/m) = \frac{p(A)\,p(m/A)}{p(M)} = \frac{\frac{1}{2} * 0.05}{0.0425} = \boxed{0.5882}$

$p(B/M) = \frac{p(B)\,p(m/B)}{p(m)} = \frac{\frac{1}{4} * 0.04}{0.0425} = 0.2353$

$p(C/m) = \frac{p(C)\,p(m/C)}{p(m)} = \frac{\frac{1}{4} * 0.03}{0.0425} = 0.1765$

$\Rightarrow$ Es más probable que proceda de la finca A.

4/13

(e):- $p(\bar{M}\cap\bar{A}) = p(\overline{M\cup A}) = 1 - p(M\cup A) =$

$= 1 - \left[p(M) + p(A) - p(M\cap A)\right] =$

$= 1 - \left[p(M) + p(A) - p(A)\,p(M/A)\right] =$

$= 1 - \left[0.0425 + 0.5 - 0.5 \ast 0.05\right] =$

$= 1 - 0.5175 = \boxed{0.4825}$

P. 3: ▸ $X \equiv$ Peso de los huevos

$E(X) = 61$

$\sigma_X = 4$

▸ Categoría MEDIANA si $X \in (53, 69)$

(a): Como NO conocemos la distribución de probabilidad de X vamos a intentar aplicar la D. de Tchebychev:

$p(\text{cat. mediana}) = p(53 < X < 69) =$

$= p(|X - 61| < \underbrace{8}) \geq 1 - \frac{1}{K^2} = 1 - \frac{1}{4}$

$\left(\begin{array}{l} 8 = K * \sigma_X = K * 4 \\ \Rightarrow K = 8/4 = 2 \end{array} \right)$

$\Rightarrow p(\text{cat. mediana}) \geq \underline{\underline{0.75}}$.

(b): $X \sim N(\mu = 61, \sigma = 4)$

(b.1): $p(\text{cat. mediana}) = p(53 < X < 69) =$

$= \phi\left(\frac{69-61}{4}\right) - \phi\left(\frac{53-61}{4}\right) = \phi(2) - \underbrace{\phi(-2)}_{1-\phi(2)} =$

$= 2\phi(2) - 1 = 2 * 0.9772 - 1 = \boxed{0.9544}$

(b.2): $Y \equiv$ Nº de huevos de cat. mediana en una docena $\sim B(n=12, p=0.9544)$

$$\Rightarrow \begin{cases} f_Y(y) = \binom{12}{y} 0.9544^y \, 0.0456^{12-y} \\ y = 0, 1, 2, 3, \dots, \underbrace{10, 11, 12}_{\text{Al menos } 10} \end{cases}$$

$$P(Y \geq 10) = \binom{12}{10} 0.9544^{10} \, 0.0456^2 +$$

$$+ \binom{12}{11} 0.9544^{11} \, 0.0456^1 + \binom{12}{12} 0.9544^{12} \, 0.0456^0 =$$

$$= 66 * 0.9544^{10} * 0.0456^2 + 12 * 0.9544^{11} * 0.0456 +$$

$$+ 0.9544^{12} = 0.9544^{10} \left(66 * 0.0456^2 + 12 * 0.9544 * 0.0456 + \right.$$

$$\left. + 0.9544^2 \right) = 1.57 * 0.9544^{10} = \boxed{0.9847}$$

(b.3): $T \equiv$ Nº de huevos de categoría mediana entre 500 huevos $\sim B(n = 500, p = 0.9544)$

$\Rightarrow E(T) = n * p = 500 * 0.9544 = 477.2$

$$P(T \leq 485) \cong$$

$$\begin{cases} np = 500 * 0.9544 = 477.2 \\ np(1-p) = 500 * 0.9544 * 0.0456 = 21.76032 \\ \text{si } W \sim N(477.2,\ \sigma_w = 4.66) \Rightarrow T \approx W \end{cases}$$

corrección por continuidad

$$\cong P(W \leq 485 + 0.5) =$$

$$= \phi\left(\frac{485.5 - 477.2}{4.66}\right) = \phi(1.78) = \boxed{0.9625}$$

P.4: $T \equiv$ Tiempo que transcurre antes de que una persona sea atendida

$$f(t) = \begin{cases} 0, & t \leq 0 \\ K e^{-\frac{t}{4}}, & t > 0 \end{cases}$$

(a): Notar que es la fd de la dist. exponencial de parámetro $\lambda = \frac{1}{4}$

$$\Rightarrow K = \frac{1}{4} = 0.25$$

entonces la f. de DISTRIBUCIÓN de T es:

$$F_T(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-0.25 \cdot t}, & t > 0 \end{cases}$$

(b): $P(T \leq 6 \mid T > 3) =$

$$= \frac{P(3 < T \leq 6)}{P(T > 3)} = \frac{P(3 < T \leq 6)}{1 - P(T \leq 3)} =$$

$$= \frac{F_T(6) - F_T(3)}{1 - F_T(3)} =$$

$$= \frac{\left(1 - e^{-0.25*6}\right) - \left(1 - e^{-0.25*3}\right)}{1 - \left(1 - e^{-0.25*3}\right)} =$$

$$= \frac{e^{-0.75} - e^{-1.5}}{e^{-0.75}} = 1 - e^{-1.5+0.75} =$$

$$= 1 - e^{-0.75} = \boxed{0.5276}$$

(c): $X \equiv$ Nº de personas que se olvidan de pagar en UN DÍA $\sim Po(\lambda = 0.5)$

Definimos $Y \equiv$ Nº de personas que se olvidan de pagar en SEIS DÍAS $= \sum_{i=1}^{6} X_i \sim Po(\underbrace{6*0.5}_{3})$

$$\left(\begin{array}{l} X_i \sim Po(0.5) \\ i = 1 - 6 \\ \text{IND} \end{array} \right)$$

$$\Rightarrow f_Y(y) = e^{-3} \frac{3^y}{y!}, \quad y = 0, 1, 2, \underbrace{3, 4, \ldots}_{\text{AL MENOS } 3}$$

$$P(Y \geq 3) = 1 - P(Y \leq 2) =$$

$$= 1 - \sum_{y=0}^{2} e^{-3} \frac{3^y}{y!} = 1 - e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right)$$

$$= 1 - e^{-3}\left(1 + 3 + \frac{9}{2}\right) = 1 - 8.5 * e^{-3} = \boxed{0.5768}$$

P. 5 : $X \equiv$ Nº de semillas que quedan en el suelo después de pasar la cosechadora $\sim N(\mu, \sigma)$.

Datos muestrales: $n = 10$

$$77, 73, \ldots, 76, 75 \longrightarrow \bar{x} = \frac{77 + 73 + \cdots + 76 + 75}{10} = 77.9$$

$$\rightarrow s^2 = \frac{10}{9}\left(\frac{77^2 + 73^2 + \cdots + 76^2 + 75^2}{10} - (77.9)^2\right) = 9.43$$

$$\rightsquigarrow s = \sqrt{9.43} = 3.071$$

(a) : IC al 98% para μ.

Tomamos el estadístico $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow$

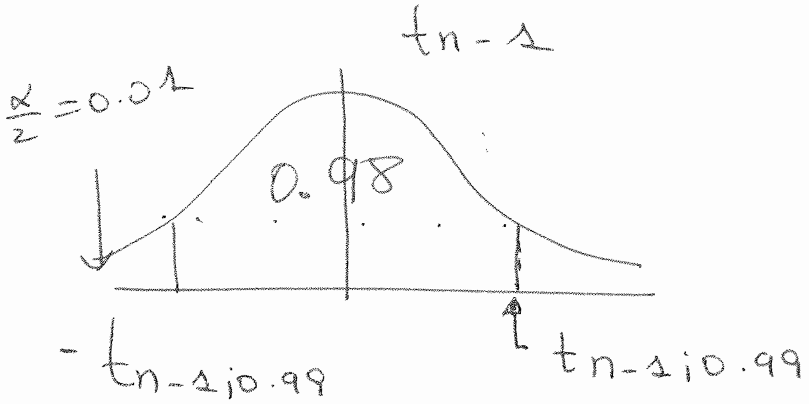
$$\Rightarrow Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Como σ es desconocida, la estimamos mediante la desviación típica muestral S

$\Longrightarrow$

$\Rightarrow$ la v.a. $T = \dfrac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \sim t_{n-1}$

en la t_{n-1}, buscamos los percentiles:



$$\Rightarrow 0.98 = P\left(-t_{n-1;0.99} \leq \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \leq t_{n-1;0.99}\right) =$$

Despejamos adecuadamente μ

$$= P\left(\bar{X} - t_{n-1;0.99} * \frac{\hat{S}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1;0.99} * \frac{\hat{S}}{\sqrt{n}}\right)$$

$\Rightarrow$ el IC al 98% para μ es:

$$\bar{X} \pm t_{n-1;0.99} * \frac{\hat{S}}{\sqrt{n}}$$

Para nuestra muestra quedaría:

$$\bar{X} \pm t_{9;0.99} * \frac{S}{\sqrt{n}} = 77.9 \pm 2.82 * \frac{3.071}{\sqrt{10}} =$$

$$= 77.9 \pm 2.74 = (75.16,\ 80.64).$$

▸ ERROR de la estimación $= 2.74$.

(b): ▸ $H_0: \mu = 80$

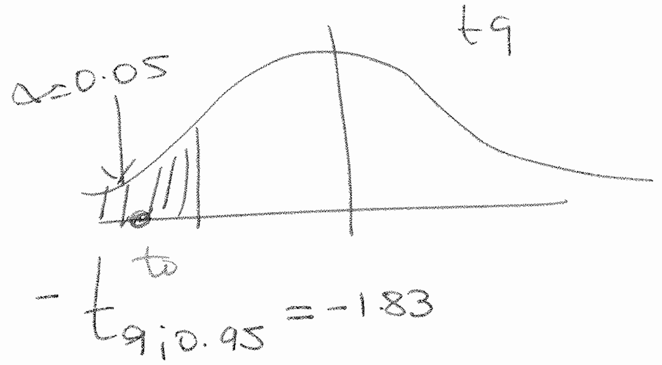
$H_1: \mu < 80$

▸ $\alpha = 0.05$

▸ EST. del CONTRASTE bajo H_0:

$$T_0 = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t_{n-1} \quad (\equiv t_9)$$

▸ $R =$

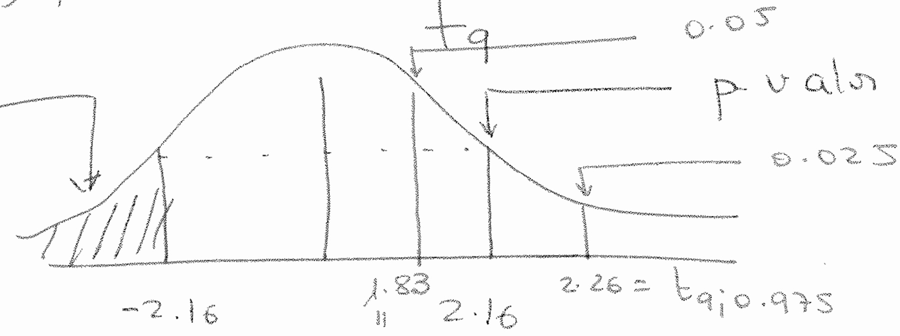


▸ $t_0 = \dfrac{77.9 - 80}{3.071 / \sqrt{10}} = -2.16 \in R$

⇒ Rechazamos H_0 al 95% de confianza

⇒ La pérdida de producción es significativamente menor a 80 qq/ha. esto es, es aceptable.

▸ p-valor =



⇒ $\boxed{0.025 < \text{p-valor} < 0.05}$

(c) : Sup. $X \sim N(\mu, \sigma = 4)$

¿ n | Margen de error < 2.5 al 98%?

$$\left(\begin{array}{l} 1-\alpha = 0.98 \\ \alpha = 0.02 \\ \frac{\alpha}{2} = 0.01 \end{array} \middle| \Rightarrow Z_{1-\frac{\alpha}{2}} = Z_{0.99} = 2.33 \right)$$

$$\Rightarrow Z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} < 2.5$$

$$\Rightarrow 2.33 * \frac{4}{\sqrt{n}} < 2.5$$

$$\Rightarrow n > \left(\frac{2.33 * 4}{2.5} \right)^2 = 13.9$$

$$\Rightarrow \boxed{n = 14 \text{ observaciones}}$$