Dpto. Matemática Aplicada y Estadística

Grado en IIAA y Grado en IHJ
Asignatura: **Estadística Aplicada. Curso 2012-2013**
FEBRERO 2013

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

1. [**0.5 puntos**] Las trazas de metales en el agua potable afectan su sabor y en altas concentraciones pueden representar un riesgo para la salud. En un artículo aparece un estudio en el cual se seleccionan muestras de agua en distintas zonas (X) y se determinó la concentración de zinc (en mg/l) en estas muestras de agua (Y). Con los datos obtenidos se decide ajustar el siguiente modelo lineal entre las características X e Y:

$$Y = 0.56 - 0.07X \qquad \text{con } R^2 = 0.89$$

   (a) Calcular el coeficiente de correlación lineal entre X e Y. Interpretar el signo del coeficiente de correlación entre X e Y. (**0.25 puntos**)

   (b) Sabiendo que la varianza de la característica Y es 0.0253, determinar la covarianza entre X e Y. (**0.25 puntos**)

2. [**0.75 puntos**] En una determinada explotación agraria se cultivan 3 variedades de pimientos P1, P2 y P3. El 36% de la producción total de pimientos es de la variedad P1 y las otras dos variedades de pimientos tienen el mismo volumen de producción. Se sabe que la probabilidad de que en la variedad P1 aparezca un determinado tipo de parásito es del 0.80 y la probabilidad de que en la variedad P2 no aparezca el citado parásito es del 15%. Además, por experiencias previas, se sabe que el 24% de la producción es de la variedad P3 y presenta este tipo de parásito. Se pide:

   (a) Determinar la probabilidad de que elegido un pimiento al azar de la producción de la explotación no presente el citado parásito. (**0.25 puntos**)

   (b) Si el pimiento elegido presenta el citado parásito, determinar la probabilidad de que no sea de la variedad P2. (**0.25 puntos**)

   (c) Determinar la probabilidad de que el pimiento no presente el citado parásito y no sea de la variedad P2. (**0.25 puntos**)

3. [**1.25 puntos**] La fecha de cosecha de una determinada hortaliza (en escala juliana : 1-365 días) se puede modelizar como una distribución normal centrada sobre el 12 de septiembre (día 255) y con una desviación típica de 10 días. Se pide:

   (a) Si se considera primicia a los frutos obtenidos antes del 1 de septiembre (día 244), ¿qué proporción de cosecha será primicia? (**0.25 puntos**)

   (b) Si las hortalizas se distribuyen en cajones de 100 piezas, determinar la probabilidad de que al menos el 20% de las piezas del cajón se hayan cosechado antes del 1 de septiembre? (**0.5 puntos**)

   (c) La aplicación de un regulador del crecimiento permite adelantar 3 días la fecha de cosecha y reduce la desviación típica. ¿Cuál será el nuevo valor de σ si sabemos que el 18% de la cosecha será primicia? (**0.5 puntos**)

4. [**1 punto**] Una empresa de electricidad asegura que la proporción de hogares con aparatos eléctricos estropeados cuando se produce una subida de tensión en la zona se puede modelizar como una variable de media 0.25 y de varianza 0.0375. Se pide:

   (a) Determinar una cota para que la proporción de hogares con aparatos eléctricos estropeados debido a una subida de tensión se sitúe en el intervalo [0,0.5]. (**0.25 puntos**)

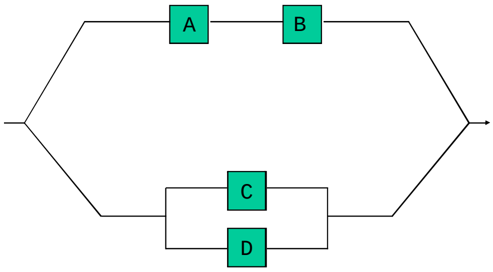
(b) Supongamos que la empresa ha comprobado empíricamente que la proporción de hogares con aparatos eléctricos estropeados cuando se produce una subida de tensión en la zona se puede modelizar como una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} k(x^2 - 2x + 1), & 0 < x < 1 \\ 0, & \text{en otro caso} \end{cases}$$

(b1) Determinar el valor de k para que $f(x)$ sea una función de densidad. (**0.25 puntos**)

(b2) Determinar la función de distribución de la variable aleatoria X. (**0.5 puntos**)

5. [**1.25 puntos**] Un sistema formado por cuatro componentes independientes sigue el esquema siguiente:



Cada una de las componentes del sistema tiene un tiempo de vida útil que sigue una distribución exponenecial de media 8 años. Se pide:

(a) Determinar la probabilidad de que una determinada componente del sistema dure más de 2 años. (**0.25 puntos**)

(b) Si el funcionamiemto de una determinada componente está por encima de los dos años, determinar la probabilidad de que no falle antes del siguiente año. (**0.25 puntos**)

(c) Determinar la probabilidad de que el sistema completo dure más de 2 años. (**0.5 puntos**)

(d) Estos sistemas pasan un control de calidad para garantizar su buen funcionamiento. El técnico que efectúa el control encuentra que, el número de sistemas que se rechazan a la semana se puede modelizar como una distribución de Poisson de media igual a 2. Hallar la probabilidad de que como máximo se rechacen 5 sistemas en el lapso de cuatro semanas. (**0.25 puntos**)

6. [**1.25 puntos**] Se considera que la fibra de un tipo de algodón es de buena calidad si su longitud media es mayor a 203 mm. Para saber si un lote cumple con esta norma de calidad se toman 19 fibras de algodón del lote y se mide la longitud de las mismas. Los resultados obtenidos vienen en la tabla siguiente:

$$\sum_{i=1}^{19} x_i = 3909.82$$

$$\sum_{i=1}^{19} x_i^2 = 805313.36$$

Admitiendo la hipótesis de normalidad para la longitud de la fibra de este tipo de algodón, se pide:

(a) Construir de manera detallada un intervalo de confianza al 98% para la longitud media de la fibra de algodón. ¿Cuánto vale el margen de error de este intervalo? **(0.5 ptos.)**

(b) A partir de los datos muestrales, ¿se puede concluir al 95% de confianza que el algodón de este lote es de buena calidad, esto es, que la longitud media de la fibra de algodón es mayor a 203 mm.? Plantear y llevar a cabo el contraste adecuado para responder a esta pregunta. Calcular de manera aproximada el $p - valor$ de la prueba.**(0.5 ptos.)**

(c) Supongamos que la desviación típica de la v.a. longitud de la fibra de este tipo de algodón es igual a $\sigma = 5.9$ mm, determinar qué tamaño de la muestra será necesario utilizar si se desea reducir a la mitad el margen de error obtenido en la estimación dada por el intervalo de confianza del primer apartado al 98% de confianza. **(0.25 ptos.)**

GIIAA
y
GIICJ

FEBRERO 2013

P.1: $X \equiv$ Zonas en las que se recoge el agua
$Y \equiv$ Concentración de zinc en el agua

$$y = 0.56 - 0.07X$$

$$R^2 = 0.89$$

(a): $r = -\sqrt{R^2} = -\sqrt{0.89} = -0.9434$

ya que la dependencia entre X e Y es negativa

Si $r > 0 \Rightarrow S_{xy} > 0 \Rightarrow$ Dependencia POSITIVA entre X e Y
Si $r < 0 \Rightarrow S_{xy} < 0 \Rightarrow$ Dependencia NEGATIVA entre X e Y

(b): $s_y^2 = 0.0253 \Rightarrow$ ¿S_{xy}?

$$R^2 = 0.89 = \frac{S_{xy}^2}{S_x^2 * S_y^2} = \frac{S_{xy}^2}{S_x^2 * 0.0253} \Rightarrow$$

$$\Rightarrow \frac{S_{xy}^2}{S_x^2} = 0.89 * 0.0253 = 0.022517$$

$$\hat{a} = \frac{S_{xy}}{S_x^2} = -0.07$$

$\Rightarrow$

$$\Rightarrow 0.022517 = \boxed{\frac{S_{xy}}{S_x^2}} \cdot S_{xy} = -0.07 * S_{xy}$$

$$\Rightarrow S_{xy} = -\frac{0.022517}{0.07} = \boxed{-0.3217}$$

P.2:

$$\left.\begin{matrix} P1 \\ P2 \\ P3 \end{matrix}\right\} \quad \left\{\begin{matrix} p(P1) = 0.36 \\ p(P2) = \alpha = p(P3) \end{matrix}\right\} \Rightarrow 0.36 + 2\alpha = 1 \Rightarrow \boxed{\alpha = 0.32}$$

$$\Rightarrow \left.\begin{matrix} p(P1) = 0.36 \\ p(P2) = 0.32 \\ p(P3) = 0.32 \end{matrix}\right\}$$

$A \equiv$ Aparezca una variedad de parásito

- $p(A/P1) = 0.80$
- $p(\bar{A}/P2) = 0.15 \longrightarrow p(A/P2) = 0.85$
- $p(A \cap P3) = 0.24 = p(P3)\,p(A/P3) = 0.32 * p(A/P3)$

$$\Rightarrow p(A/P3) = \frac{0.24}{0.32} = 0.75$$

(a): $p(\bar{A}) = 1 - p(A) = \boxed{0.20}$

$$\left\{\begin{aligned} p(A) &= p(P1)\,p(A/P1) + p(P2)\,p(A/P2) + p(P3)\,p(A/P3) = \\ &= 0.36 * 0.80 + 0.32 * 0.85 + 0.32 * 0.75 = 0.80 \end{aligned}\right.$$

(b): $p(\overline{P2}/A) = 1 - p(P2/A) =$

$$= 1 - \frac{p(P2 \cap A)}{p(A)} = 1 - \frac{p(P2)\,p(A/P2)}{p(A)} =$$

$$= 1 - \frac{0.32 * 0.85}{0.80} = \boxed{0.66}$$

(c): $p(\bar{A} \cap \overline{P2}) = p(\overline{A \cup P2}) = 1 - p(A \cup P2) =$

$$= 1 - [\,p(A) + p(P2) - p(P2)\,p(A/P2)\,] =$$

$$= 1 - [\,0.8 + 0.32 - 0.32 * 0.85\,] = \boxed{0.152}$$

P.3:- $X \equiv$ Fecha de la cosecha de la hortaliza $\sim N(\mu = 255, \sigma = 10)$

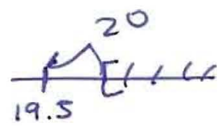
(a):- $p(X < 244) = \phi\left(\frac{244 - 255}{10}\right) = \phi(-1.1) =$

$= 1 - \phi(1.1) = 1 - 0.8643 = \boxed{0.1357}$

$\Rightarrow$ El 13.57% de la cosecha será primicia.

(b):- $Y \equiv$ Nº de hortalizas que son primicias en un cajón de 100 hortalizas $\sim B(n = 100, p = 0.1357)$.

$P(Y \geq 20) \cong$

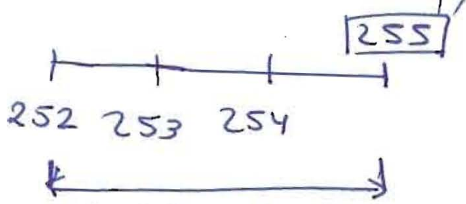


$$\begin{cases} np = 13.57 \\ np(1-p) = 11.728551 \rightarrow \sigma_W = 3.42 \\ \Rightarrow \text{si } W \sim N(\mu = 13.57, \sigma_W = 3.42) \text{ entonces} \\ \quad Y \approx W \end{cases}$$

$\cong P(W \geq 19.5) = 1 - \phi\left(\frac{19.5 - 13.57}{3.42}\right) =$

$= 1 - \phi(1.73) = 1 - 0.9582 = \boxed{0.0418}$

(c): $T \equiv$ Fecha de cosecha de la hortaliza con el regulador $\sim N(\mu = 252, \sigma)$



Adelanta 3 días

¿ σ / $P(T < 244) = 0.18$?

$$0.18 = \phi\left(\frac{244-252}{\sigma}\right) = \phi\left(\frac{-8}{\sigma}\right) = 1 - \phi\left(\frac{8}{\sigma}\right)$$

$$\Rightarrow \phi\left(\frac{8}{\sigma}\right) = 0.82 \Rightarrow \frac{8}{\sigma} = Z_{0.82} \cong 0.92$$

$$\boxed{\sigma = \frac{8}{0.92} = 8.7}$$

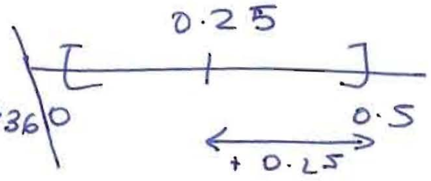
La desviación típica se reduce un poco menos de 2 días.

P.4:

$X \equiv$ Proporción de hogares con aparatos eléctricos estropeados por una subida de tensión

$E(X) = 0.25$

$\sigma_X^2 = 0.0375 \rightarrow \sigma_X = 0.1936$



(a): $P[0 \leq X \leq 0.5] = P(|X - \mu| \leq 0.25) \geq$

$$\left(0.25 = K * \sigma_X = K * 0.1936 \Rightarrow K = \frac{0.25}{0.1936} = 1.29\right)$$

$$\geq 1 - \frac{1}{(1.29)^2} = 0.399\,1$$

Aplicando la D. de Tchebychev.

(b): Sup. $X \sim$

$$f(x) = \begin{cases} k(x^2-2x+1), & 0<x<1 \\ 0, & \text{———} \end{cases}$$

(b.1):

- $f(x) = k(x^2-2x+1) = k\underbrace{(x-1)^2}_{\geq 0} \geq 0 \iff k \geq 0$

- $1 = \int_{\mathbb{R}} f(x)\,dx = k\int_0^1 (x^2-2x+1)\,dx =$

$= k\left[\frac{x^3}{3}\Big|_0^1 - x^2\Big|_0^1 + x\Big|_0^1\right] =$

$= k\left(\frac{1}{3} - \cancel{1} + \cancel{1}\right) = \frac{k}{3} \Rightarrow \boxed{k=3}$

(b.2): $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, \ \forall x \in \mathbb{R}$

$$\underset{0}{(}\text{———————}\underset{1}{)}$$

- Si $x \leq 0 \Rightarrow F(x) = 0.$

- Si $0<x<1 \Rightarrow F(x) = 3\int_{-\infty}^{x}(t^2-2t+1)\,dt =$

$= 3\int_0^x (t^2-2t+1)\,dt = 3\left[\frac{t^3}{3}\Big|_0^x - t^2\Big|_0^x + t\Big|_0^x\right] =$

$= x^3 - 3x^2 + 3x.$

- Si $x \geq 1$, $F(x) = 1.$

$$\Rightarrow F(x) = \begin{cases} 0, & x \leq 0 \\ x^3 - 3x^2 + 3x, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

P.5:– $X \equiv$ tiempo de vida de cada componente $\sim Exp(\lambda = 1/8)$

$E(X) = \frac{1}{\lambda} = 8 \Rightarrow \lambda = 1/8$

$\Rightarrow$ f. densidad: $f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{8} e^{-\frac{x}{8}}, & x > 0 \end{cases}$

f. DISTRIBUCIÓN: $F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\frac{x}{8}}, & x > 0 \end{cases}$

(a):– $P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) =$

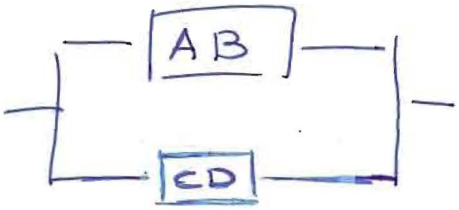
$= 1 - (1 - e^{-2/8}) = e^{-2/8} = \boxed{0.7788}$

(b):– $P(X > 3 / X > 2) = \frac{P[(X>2) \cap (X>3)]}{P(X>2)} =$

$= \frac{P(X>3)}{P(X>2)} = \frac{e^{-3/8}}{e^{-2/8}} = e^{-3/8 + 2/8} = e^{-1/8} =$

$= \boxed{0.8825}$

(c) :



▶ p(AB funcione más de 2 años) = $p[(A>2) \cap (B>2)] =$

$\overset{IND}{=} p(A>2) * p(B>2) = (0.7788)^2 = 0.6065$

▶ p(CD dure más de 2 años) = $p[(C>2) \cup (D>2)] =$

$= p(C>2) + p(D>2) - p[(C>2) \cap (D>2)] \overset{IND}{=}$

$= p(C>2) + p(D>2) - p(C>2) * p(D>2) =$

$= 0.7788 + 0.7788 - (0.7788)^2 = 0.9511$

▶ p(sistema COMPLETO dure más de 2 años) = $p[(AB>2) \cup (CD>2)] \overset{IND}{=}$

$= p(AB>2) + p(CD>2) - p(AB>2) * p(CD>2) =$

$= 0.6065 + 0.9511 - 0.6065 * 0.9511 = \boxed{0.9808}$

(d) : $Y \equiv$ Nº de sistemas que se rechazan a la semana $\sim Po(\lambda = 2)$

$T \equiv$ Nº de sistemas que se rechazan en cuatro semanas $= Y_1 + Y_2 + Y_3 + Y_4 \sim Po(\lambda = 8)$

$(Y_i \sim Po(\lambda = 2)\ i = 1, 2, 3, 4\ ; IND)$

$$\Rightarrow f_T(t) = e^{-8} \frac{8^t}{t!} \quad , \quad t = 0, 1, 2, 3, 4, 5, 6 \dots$$

$$P(T \leq 5) = e^{-8} \left(\frac{8^0}{0!} + \frac{8^1}{1!} + \frac{8^2}{2!} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!} \right) =$$

$$= e^{-8} \left(1 + 8 + 32 + \frac{512}{6} + \frac{4096}{24} + \frac{32768}{120} \right) =$$

$$= 570.067 * e^{-8} = \boxed{0.1912}$$

P6 :- $X \equiv$ longitud de la fibra de algodón $\sim N(\mu, \sigma)$

(a): IC al 98% para μ:

Tomamos el estimador natural de μ dado por la media muestral

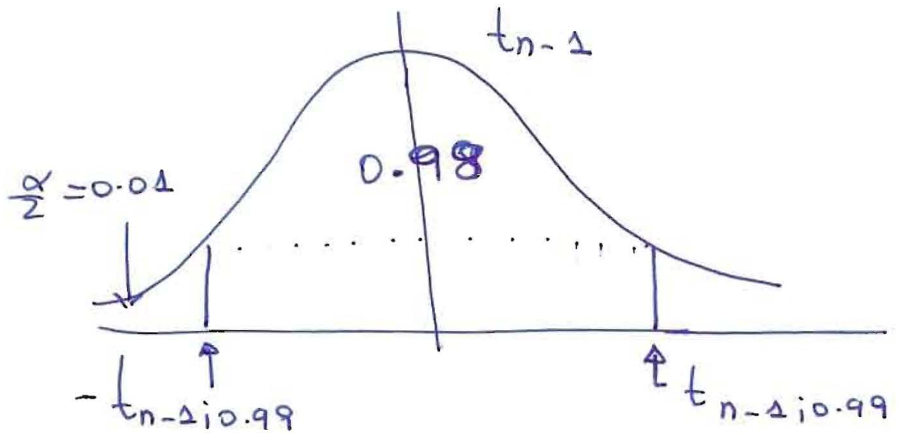
$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ por la aditividad del modelo normal

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

Ahora bien, como σ no es conocida la estimamos mediante la desviación típica muestral $\hat{S}$

$\Rightarrow$ Aplicando el Tª de Fisher tenemos que la v.a. $T = \dfrac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} \sim t_{n-1}$

$\longrightarrow$ en la t_{n-1} tomamos los valores:



$\rightarrow$ entonces

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

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

$\Rightarrow$ El IC al 98% para μ es:

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

Para nuestra muestra quedaría:

$$\bar{x} \pm t_{n-1;0.99} * \frac{s}{\sqrt{n}}$$

Datos muestrales: $n = 19$

$$\sum x_i = 3909.82 \Rightarrow \bar{x} = \frac{3909.82}{19} = 205.78$$

$$\sum x_i^2 = 805313.36 \Rightarrow s_x^2 = \frac{19}{18}\left[\frac{805313.36}{19} - \left(205.78^2\right)\right] = 41.70 \Rightarrow s_x = 6.4576$$

entonces el intervalo quedaría:

$$\bar{x} \pm t_{18;0.99} * \frac{S}{\sqrt{n}} = 205.78 \pm 2.55 * \frac{6.4576}{\sqrt{19}} =$$

$$= 205.78 \pm \underbrace{3.78}_{} = \boxed{(202,\ 209.56)}$$

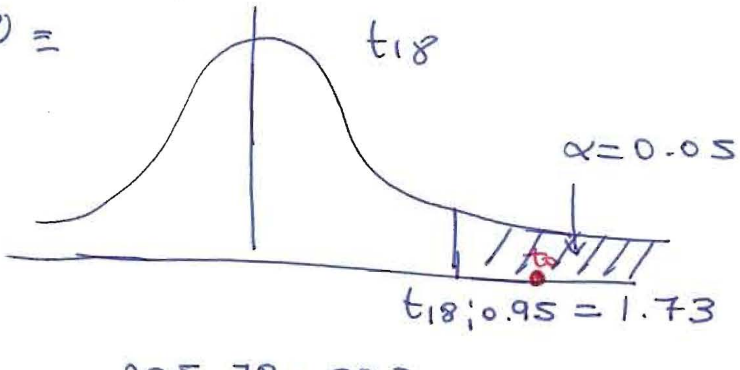
MARGEN de ERROR del intervalo.

(b):

- $H_0: \mu = 203$
  $H_A: \mu > 203$
- $\alpha = 0.05$
- EST. del CONTRASTE bajo H_0:

$$T_0 \equiv \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \ (\equiv t_{18})$$

- $R \equiv$

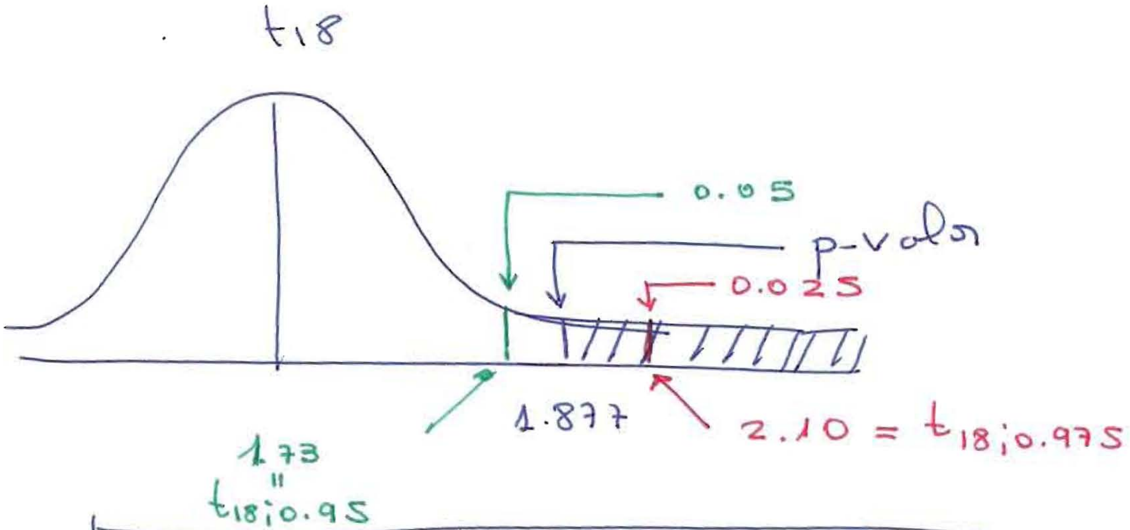


- $t_0 \equiv \dfrac{205.78 - 203}{6.4576/\sqrt{19}} \equiv 1.877$

- Como $t_0 \in R \Rightarrow$ Rechazamos H_0 al 95% de confianza

$\Rightarrow$ la longitud media de la fibra del algodón del lote es significativamente mayor a 203 mm., con lo cual, el algodón de este lote es de buena calidad.

- p-valor $= p(t_{18} > 1.877) \Rightarrow$



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

(c): Sup. $X \sim N(\mu, \sigma = 5.9)$

¿n / $\underbrace{\text{Margen de error}}_{= z_{0.99} * \frac{\sigma}{\sqrt{n}}} = \frac{1}{2} 3.78$ al 98%?

$\Rightarrow 2.33 * \frac{5.9}{\sqrt{n}} < 1.89$

$\Rightarrow n > \left(\frac{2.33 * 5.9}{1.89} \right)^2 = 52.9$

$\Rightarrow$ Necesitamos al menos $\boxed{53 \text{ observaciones}}$