Dpto. Matemática Aplicada y Estadística

Grado en IIAA _Grado en IHJ_Grado en IASB
Asignatura: **Estadística Aplicada. Curso 2014-2015**
FEBRERO 2015

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

1. [**0.5 puntos**] Un determinado estudio intenta relacionar la dureza del agua de un pozo acuífero (Y, en gramos por litro) con la temperatura (X, en ^{o}C). Tras analizar 40 experimentos de manera independiente en el laboratorio, se obtuvieron los siguients resultados:

$$\sum x_i = 1054.4 \qquad \overline{x^2} = 1685.021 \qquad \sum y_i = 20.752 \qquad \sum y_i^2 = 6269.096 \qquad \overline{x \cdot y} = 393.356$$

A partir de estos datos se pide:

(a) Determinar la recta de regresión de mínimos cuadrados que permite predecir la dureza del agua del pozo acuífero en función de la temperatura. (**0.25 puntos**)

(b) Dar una medida de la bondad del ajuste realizado e interpretar dicha medida. (**0.25 puntos**)

2. [**0.75 puntos**] En una determinada explotación agraria se cultivan 3 variedades de tomates, T1, T2 y T3 que se destinan a la exportación. El 32% de la producción de la explotación es de la variedad T1 y la producción de la variedad T3 es un tercio de la producción de la variedad T2 ya que es la variedad más costosa de producir. Por experiencia propia, se sabe que el 5% de los tomates de la variedad T1 no se pueden destinar a la exportación y el 2% de los tomates de la variedad T2 tampoco se pueden exportar. Por otra parte, de las campañas anteriores, se estima que el 16.5% de la producción es de la variedad T3 y se puede destinar a la exportación. Se pide:

(a) Calcular la probabilidad de que en un tomate seleccionado al azar de toda la producción de la explotación se pueda destinar a la exportación. (**0.25 puntos**)

(b) Elegido un tomate al azar de toda la producción que no se puede exportar, determinar la probabilidad de que sea de la variedad T1 o de la variedad T2. (**0.25 puntos**)

(c) Determinar la probabilidad de que un tomate elegido al azar de la producción se pueda destinar a la exportación y sea de la variedad T1 o de la variedad T3. (**0.25 puntos**)

3. [**1 punto**] El tiempo de vida útil, en años, de la batería de una cosechadora se comporta como una variable aleatoria X con función de distribución dada por:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{x^2}{2}, & x \in (0,1] \\ -\dfrac{x^2}{2} + \alpha x - 1, & x \in (1,2) \\ 1, & x \geq 2 \end{cases}$$

Se pìde:

(a) Determinar el valor de α para que $F(x)$ sea verdaderamente una función de distribución. **(0.25 puntos)**

(b) Determinar la función de densidad de la v.a. X y el tiempo medio de vida útil de la batería de la cosechadora.**(0.5 puntos)**

(c) Determinar la probabilidad de que la batería dure al menos año y medio. **(0.25 puntos)**

4. [**1 punto**] El peso de cierta fruta recolectada en una explotación agraria se comporta aproximadamente normal con un peso medio $\mu = 157$ gramos. Se pide:

   (a) Determinar el valor de la desviación típica del peso de la fruta, σ, si se sabe que el 80% de las frutas tiene un peso superior a 154.48 gramos. **(0.25 puntos)**

   (b) Si las frutas salen al mercado en mallas de 8 unidades, determinar la probabilidad de que al menos 6 de estas frutas tengan un peso superior a 154.48 gramos. **(0.25 puntos)**

   (c) El encargado de envasarla en las mallas encuentra que el número de fruta que no puede salir al mercado se puede modelizar como una Poisson de media 8 unidades en una hora. Determinar la probabilidad de que se retiren más de 3 unidades en el lapso de media hora. **(0.5 puntos)**

5. [**1.5 puntos**] El tiempo de vida útil, en horas, de cierto componente electrónico se considera como una variable aleatoria con media igual a 500 horas y con desviación típica igual a 400 hras. Se pide:

   (a) ¿Podrías dar una cota de la probabilidad de que el tiempo de vida del componente se encuentre en el intervalo [0,1000]? Razonar la respuesta.(**0.25 puntos**)

   (b) Supongamos a partir de ahora, que el tiempo de vida útil, en horas, de dicho componente electrónico se considera como una variable aleatoria exponencial de media igual a 500 horas.

      (b1) Determinar la probabilidad de que el componente funcione por encima de las 400 horas. (**0.25 puntos**)

      (b2) Si un componente de este tipo a las 400 horas aún funciona, determinar la probabilidad de que deje de funcionar antes de las 600 horas siguientes. (**0.25 puntos**)

      (b3) Si se consideran defectuosos aquellos componentes cuya duración es inferior a 400 horas y en un lote hay 75 componentes, determinar la probabilidad de que el número de componenetes defectuosos en el lote sea inferior a 35 componentes. (**0.5 puntos**)

      (b4) Una máquina consta de 2 de estos componentes y la máquina funciona cuando funciona al menos uno de sus dos componentes (**sistema en paralelo**). Determinar la probabilidad de que la máquina funcione al menos 400 horas sabiendo que las componentes funcionan de manera independiente entre sí. (**0.25 puntos**)

6. [**1.25 puntos**] Se quiere analizar el contenido de celulosa de una determinada variedad de alfalfa. Para ello, se seleccionan 7 cortes de alfalfa de la citada variedad y se analiza el contenido de celulosa obteniéndose los siguientes resultados en mg./g.:

							media muestral	varianza muestral
135.4	137.5	140.4	128.6	131.5	133.5	155	137.46	74.7

   (a) Construir de manera detallada un intervalo de confianza al 95% para el contenido promedio de celulosa de la variedad de alfalfa. Indicar todos los resultados teóricos necesarios para obtener el citado intervalo. **(0.5 puntos.)**

   (b) A partir de los datos muestrales, ¿podemos asumir al 95% de confianza que el contenido promedio de celulosa de dicha variedad de alfalfa es significativamente mayor a 131 mg./g. ? Plantear y llevar a cabo el contraste adecuado para responder a esta pregunta. Calcular el $p - valor$ de la prueba. **(0.5 puntos.)**

   (c) Supongamos que la desviación típica de la variable aleatoria "contenido de celulosa de esta variedad de alfalfa" es $\sigma = 8.5$ mg./g., determinar qué tamaño de la muestra será necesario utilizar si queremos estimar el contenido promedio de celulosa de esta variedad de alfalfa con un error menor a 3 mg./g. y garantizando una confianza del 95%. **(0.25 puntos).**

Grado de AGRÍCOLAS : $\boxed{\text{FEBRERO - 2015}}$

P.1 :

(a) : $Y \equiv$ Dureza del agua

$X \equiv$ Temperatura.

$n = 40 \longrightarrow \sum x_i = 1054.4 \longrightarrow \bar{x} = \frac{1054.4}{40} = 26.36$

$\overline{X^2} = 1685.021 \rightarrow S_x^2 = \frac{40}{39}(1685.021 - 26.36^2) = 1015.56$

$\sum y_i = 20.752 \rightarrow \bar{y} = 0.519$

$\sum y_i^2 = 6269.096 \rightarrow S_y^2 = \frac{40}{39}\left(\frac{6269.096}{40} - 0.519^2\right) = 160.47$

$\overline{xy} = 393.356 \rightarrow S_{xy} = \frac{40}{39}(393.356 - 26.36 * 0.519) = 389.41$

$\Rightarrow Y = f(X) = b + aX$ donde

$$\hat{a} = \frac{S_{xy}}{S_x^2} = \frac{389.41}{1015.56} = 0.3834$$

$$\hat{b} = \bar{y} - \hat{a}\cdot\bar{x} = 0.519 - 0.3834 * 26.36 = -9.5874$$

$$\Rightarrow \boxed{Y = -9.5874 + 0.3834 * X}$$

(b) : $R^2 = \frac{S_{xy}^2}{S_x^2 \cdot S_y^2} = \frac{389.41^2}{1015.56 * 160.47} = \boxed{0.93}$

Ajuste muy bueno.

P.21:

$$\begin{cases} T1 \longrightarrow p(T1) = 0.32 \\ T2 \longrightarrow p(T2) = \alpha \\ T3 \longrightarrow p(T3) = \frac{1}{3}\alpha \end{cases} \Bigg\} \; 0.32 + \alpha + \frac{1}{3}\alpha = 1$$

$$\Rightarrow \frac{3\alpha + \alpha}{3} = 1 - 0.32 = 0.68 \Rightarrow 4\alpha = 2.04$$

$$\alpha = 0.51$$

$$\Rightarrow \begin{cases} p(T1) = 0.32 \\ p(T2) = 0.51 \\ p(T3) = 0.17 \end{cases}$$

$E \equiv$ Tomate que se destina a la exportación

$$p(\bar{E} \mid T1) = 0.05 \longleftrightarrow p(E \mid T1) = 0.95$$

$$p(\bar{E} \mid T2) = 0.02 \longleftrightarrow p(E \mid T2) = 0.98$$

$$p(T3 \cap E) = 0.165 = p(T3)\,p(E \mid T3) = 0.17\, p(E \mid T3)$$

$$\Rightarrow p(E \mid T3) = \frac{0.165}{0.17} = 0.971 \longleftrightarrow p(\bar{E} \mid T3) = 0.029$$

(a): $p(E) = \sum_{i=1}^{3} p(T_i)\, p(E \mid T_i) =$

$$= 0.32 * 0.95 + 0.51 * 0.98 + 0.17 * 0.971 =$$

$$= \boxed{0.9689} \quad \rightsquigarrow \quad p(\bar{E}) = 0.0311$$

(b): $p(T1 \cup T2 \mid \bar{E}) = \dfrac{p[(T1 \cup T2) \cap \bar{E}]}{p(\bar{E})} =$

$$= \frac{p[(T1 \cap \bar{E}) \overset{\text{Incompatibles}}{\widehat{\cup}} (T2 \cap \bar{E})]}{p(\bar{E})} = \frac{p(T1 \cap \bar{E}) + p(T2 \cap \bar{E})}{p(\bar{E})} =$$

$$= \frac{p(T1)\, p(\bar{E} / T1) + p(T2)\, p(\bar{E} / T2)}{p(\bar{E})} =$$

$$= \frac{0.32 * 0.05 + 0.51 * 0.02}{0.0311} = \boxed{0.8424}$$

(c): $p[E \cap (T1 \cup T3)] = p[(E \cap T1) \overset{\text{Incompatibles}}{\widehat{\cup}} (E \cap T3)] =$

$$= p(E \cap T1) + p(E \cap T3) = p(T1)\, p(E \mid T1) + p(T3)\, p(E \mid T3) =$$

$$= 0.32 * 0.95 + 0.17 * 0.971 = \boxed{0.4691}$$

P.3: $X \equiv$ Tiempo de vida de la batería, en años

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

(a) : $F(x)$ tiene que ser una función continua en $\mathbb{R}$ $\Rightarrow$ $\lim_{x \to 2^-} F(x) = \lim_{x \to 2^+} F(x) = F(2) = 1$

$$\Rightarrow \bullet \lim_{x\to 2^-} F(x) = \lim_{x \to 2} \left(-\frac{x^2}{2} + \alpha x - 1 \right) = -\frac{2^{\cancel{2}}}{\cancel{2}} + 2\alpha - 1 =$$

$$= -2 + 2\alpha - 1 = 2\alpha - 3$$

$$\bullet\ F(1) = \lim_{x \to 1^+} F(x) = 1$$

$\Rightarrow 2\alpha - 3 = 1$

$\Rightarrow 2\alpha = 4 \Rightarrow \boxed{\alpha = 2}$

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

F. DISTRIBUCIÓN de X

(b) : • $\forall x$ en los que existe la derivada de $F(x)$ se tiene que $F'(x) = f(x)$

$$\Rightarrow f(x) = \begin{cases} \frac{2x}{2} = x, & 0 < x \leq 1 \\ -\frac{2x}{2} + 2 = 2 - x, & 1 < x < 2 \\ 0, & \text{resto} \end{cases}$$

f. de densidad de la v.a. X

$$E(X) = \int_{\mathbb{R}} x \cdot f(x)\,dx = \int_0^1 x \cdot x\,dx + \int_1^2 x(2-x)\,dx = \int_0^1 x^2\,dx + \int_1^2 (2x - x^2)\,dx =$$

$$= \frac{x^3}{3}\Big|_0^1 + x^2\Big|_1^2 - \frac{x^3}{3}\Big|_1^2 = \frac{1}{3} + (4-1) - \frac{1}{3}(8-1) = \frac{1}{3} + 3 - \frac{7}{3} = 3 - \frac{6}{3} = \boxed{1}$$

(b): $P(X \geq 1.5) = 1 - P(X < 1.5) =$

AÑO y MEDIO = 1.5 años

$$= 1 - F(1.5) = 1 - \left[-\frac{(1.5)^2}{2} + 2 \cdot 1.5 - 1\right] =$$

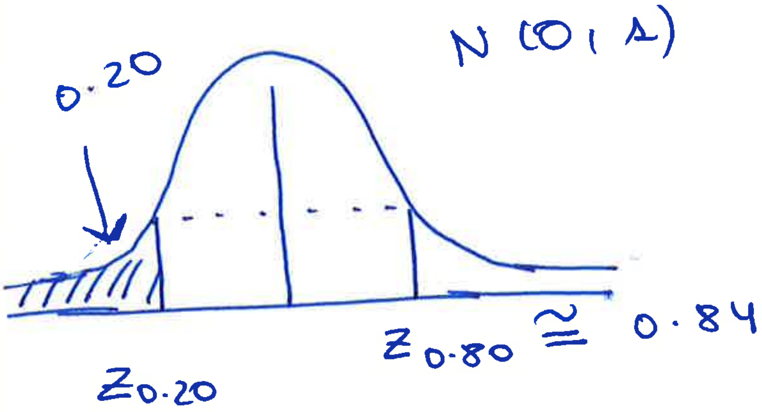
$$= 1 + 1.125 - 3 + 1 = \boxed{0.125}$$

P.4 : $X \equiv$ Peso de la fruta $\sim N(\mu = 157, \sigma)$

(a): ¿ σ | $P(X > 154.48) = 0.80$?

$$0.80 = P(X > 154.48) = 1 - \phi\left(\frac{154.48 - 157}{\sigma}\right)$$

$$\Rightarrow \phi\left(-\frac{2.52}{\sigma}\right) = 0.20 \Rightarrow Z_{0.20} = \frac{-2.52}{\sigma}$$



$$\Rightarrow \quad z_{0.20} = -z_{0.80} = -0.84 = \frac{-2.52}{\sigma}$$

$$\Rightarrow \quad \boxed{\sigma = \frac{2.52}{0.84} = 3}$$

(b): $Y \equiv$ Nº de frutas que en una malla tienen un peso superior a 154.48

$$\sim B(n = 8,\ p = p(X > 154.48) = 0.80)$$

$$\Rightarrow \begin{cases} f_Y(y) = P(Y = y) = \binom{8}{y} 0.80^y \, 0.20^{8-y} \\ y = 0,1,2,3,4,5,6,7,8 \end{cases}$$

$$P(Y \geq 6) = \binom{8}{6} 0.80^6 \, 0.20^2 + \binom{8}{7} 0.80^7 \, 0.20^1 +$$

$$+ \binom{8}{8} 0.80^8 \, 0.20^0 = 28 * 0.80^6 * 0.20^2 +$$

$$+ 8 * 0.80^7 * 0.20 + 0.80^8 =$$

$$= 0.80^6 \left(28 * 0.20^2 + 8 * 0.80 * 0.20 + 0.80^2 \right) = \boxed{0.7969}$$

(c) : $N =$ Nº de frutas que NO pueden salir al mercado en 1 HORA

$\sim Po(\lambda = 8)$

➜ La v.a. $T \equiv$ Nº de frutas que NO pueden salir al mercado en $\frac{1}{2}$ HORA

$\sim Po(\lambda = 4)$

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

$$P(T > 3) = P(T \geq 4) = 1 - P(T \leq 3) =$$

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

$$= 1 - e^{-4} \left(1 + 4 + 8 + \frac{64}{6} \right) = \boxed{0.5665}$$

P.5 : $X \equiv$ Tiempo de vida útil, en horas, de cierto comp. electrónico $\sim$?

con $E(X) = 500$ h.

$\sigma_X = 400$ h.

(a) : $P(0 \leq X \leq 1000) = P(|X - \mu| \leq 500) \geq$

0 500 1000

$500 = K * \sigma_X = k * 400 \Rightarrow k = \frac{500}{400} = 1.25$

$\geq 1 - \frac{1}{(1.25)^2} = 0.36$ y aplicamos la D. de Tcheby chev pues no conocemos la distribución de probabilidad de la v.a. X

(b): Sup. $X \sim Exp(\lambda = \frac{1}{500})$

$E(X) = \frac{1}{\lambda} = 500 \Rightarrow \lambda = \frac{1}{500}$

La f. de DISTRIBUCIÓN de la v.a. X es:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\frac{x}{500}}, & x > 0 \end{cases}$$

(b.1): $P(X > 400) = e^{-\frac{400}{500}} = e^{-0.8} =$

$= \boxed{0.4493}$

(b.2): $P(X \leq 1000 / X \geq 400) = \frac{P[(X \leq 1000) \cap (X \geq 400)]}{P(X \geq 400)} =$

$= \frac{P(400 \leq X \leq 1000)}{P(X \geq 400)} = \frac{F(1000) - F(400)}{e^{-0.8}} =$

$$= \frac{\left(1 - e^{-\frac{1000}{500}}\right) - \left(1 - e^{-\frac{400}{500}}\right)}{e^{-0.8}} = \frac{e^{-0.8} - e^{-2}}{e^{-0.8}} =$$

$$= 1 - e^{-2+0.8} = 1 - e^{-1.2} = \boxed{0.6988}$$

(b.3): ▸ $p(\text{comp. DEFECTUOSO}) = p(X \leq 400) = 0.5507$

▸ Y = Nº de comp. defectuosos en un lote de 75 componentes $\sim B(n = 75,\ p = 0.5507)$

$p(Y < 35) =$

$np = 41.3025$

$np(1-p) = 18.5572$

Si $W \sim N(\mu = 41.3025,\ \sigma = 4.3078) \Rightarrow Y \approx W$

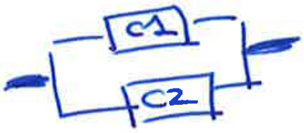
$$\cong P(W \leq \underset{\uparrow}{34.5}) = \phi\left(\frac{34.5 - 41.3025}{4.3078}\right) =$$

CORRECCION POR CONTINUIDAD

$$= \phi(-1.58) = 1 - \phi(1.58) = 1 - 0.9429 =$$

$$= \boxed{0.0571}$$

(b.4): SISTEMA en PARALELO



$p(\text{MÁQUINA FUNCIONE AL MENOS } 400\text{ h}) =$

$= p[(C1 \geq 400) \cup (C2 \geq 400)] =$

$= p(C1 \geq 400) + p(C2 \geq 400) - p((C1 \geq 400) \cap (C2 \geq 400)) =$

IND. $= 0.4493 + 0.4493 - (0.4493)^2 = \boxed{0.6967}$

P. 6 : $X \equiv$ Contenido de celulosa de la variedad de alfalfa $\sim N(\mu, \sigma)$

Datos muestrales:

$n = 7 \sim 135.4, \ldots, 155$ $\Rightarrow \begin{cases} \bar{x} = 137.41 \\ s^2 = 75.04 \Rightarrow s = 8.66 \end{cases}$

(a) : IC al 95% para μ:

Tenemos una m.a.s. de X : $X_1, X_2, \ldots X_n$ y tomamos la v.a. $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ por la aditividad del modelo normal

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

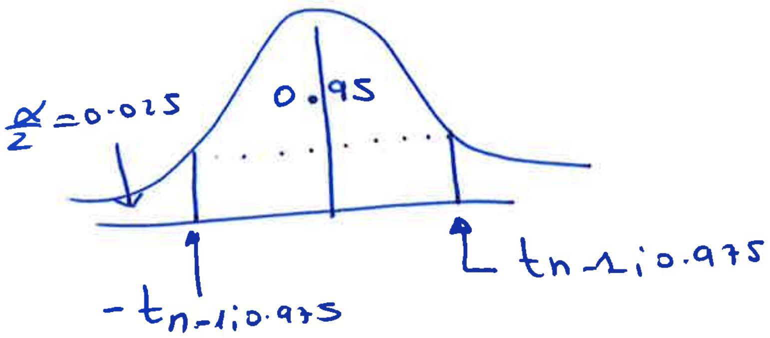
Como σ^2 es desconocida, la estimamos mediante la varianza muestral: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$\Rightarrow$ Aplicando el T^{a} de Fisher tenemos que la v.a.

$$T \equiv \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

donde S es la desviación típica muestral dada por $S \equiv \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$

▸ Tomamos la t_{n-1}



y tenemos que $0.95 = p\left(-t_{n-1;0.975} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{n-1;0.975}\right) =$

= Despejando μ

$$= p\left(\bar{X} - t_{n-1;0.975} * \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1;0.975} * \frac{S}{\sqrt{n}}\right)$$

➡ El IC al 95% para μ viene dado por:

$$\bar{X} \pm t_{n-1;0.975} * \frac{S}{\sqrt{n}}$$

Y, para la muestra que tenemos, toma la forma:

$$\bar{x} \pm t_{6;0.975} * \frac{s}{\sqrt{7}} =$$

$$= 137.41 \pm \underbrace{2.45 * \frac{8.66}{\sqrt{7}}}_{8.02} = 137.41 \pm 8.02 =$$

$$\boxed{= (129.39, 145.43)}$$

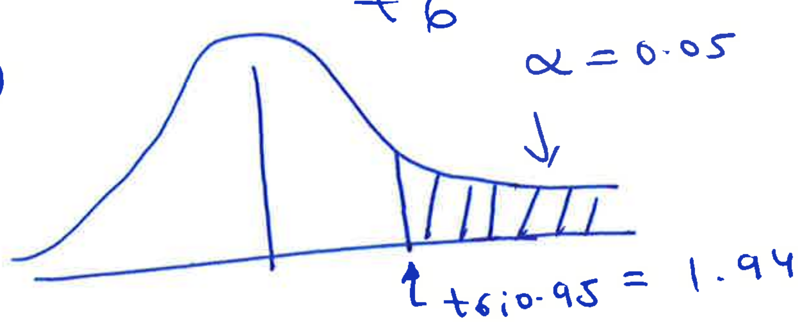
(b) $\therefore$ ▸ $H_0: \mu = 131$
$H_1: \mu > 131$

▸ $\alpha = 0.05$

▸ EST. del CONTRASTE bajo H_0:

$$T \equiv \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t_{n-1} \; (= t_6)$$

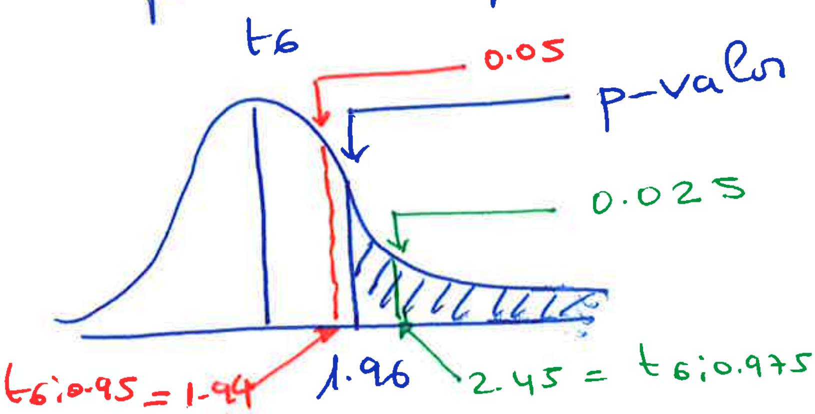
▸ R



▸ $t_0 \equiv \dfrac{137.41 - 131}{8.66 / \sqrt{7}} \simeq 1.96 \in R \Rightarrow$ Rechazamos H_0 al 95%

$\Rightarrow$ El contenido medio de celulosa de esta variedad de alfalfa es significativamente mayor a 131 mg/g.

▸ p-valor $= p(t_6 > 1.96)$



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

Además estará muy próximo a 0.05 ya que $t_{6;0.95} = 1.95$ y $t_0 = 1.96$

(c): Sup. $X \sim N(\mu, \sigma = 8.5)$

¿ n / $z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} < 3$ al 95% ?

$z_{0.975} * \frac{\sigma}{\sqrt{n}} < 3 \Rightarrow 1.96 * \frac{8.5}{\sqrt{n}} < 3$

$\Rightarrow n > \left(\frac{1.96 * 8.5}{3} \right)^2 = 30.84$

$\Rightarrow$ $\boxed{n \geq 31 \text{ cortes de alfalfa.}}$