Problema 2

$X \sim N(\mu, \sigma = 2) \qquad n = 30 \qquad \sum_{i=1}^{n} x_i = 56 \qquad \bar{x} = \frac{56}{30} = 1'8667$

a) Paso 1: Hipótesis

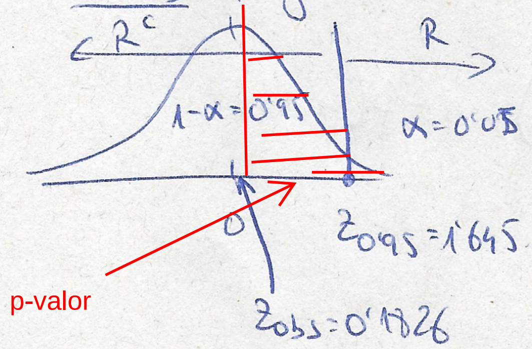
$H_0: \mu = 1'8 = \mu_0$ $\qquad \alpha = 0'05$

$H_1: \mu > 1'8$

Paso 2 Estadístico de contraste

$\left.\begin{array}{l}\cdot \text{ Población normal} \\ \cdot \text{ Varianza conocida}\end{array}\right\} \Rightarrow \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

Paso 3 Región de rechazo



$z_{1-\alpha} = z_{0'95} = 1'645$

$R = (1'645, \infty)$

$R^c = (-\infty, 1'645]$

$z_{obs} = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{1'8667 - 1'8}{2/\sqrt{30}} = 0'1826$

Paso 3: Decisión.

p-valor $= P(Z > z_{obs}) = P(Z > 0'1826) = 1 - \text{pnorm}(0'1826) = 0'4276$

No Rechazamos Ho porque $\begin{cases} z_{obs} = 0'1826 < z_{1-\alpha} = 1'645 \\ \text{p-valor} = 0'4276 \geq 0'05 \end{cases}$

Paso 5: Conclusión

No hay evidencia que sugiera que la media es mayor de 1'8 con un nivel de significación del 5%.

b)

$$\beta(\mu_1=3) = P(\text{error tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ Falsa})$$

$$= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \le z_{0'95} \;\middle|\; \bar{X}\sim N\left(3, \frac{2}{\sqrt{30}}\right)\right) =$$

$$= P\left(\bar{X} \le \mu_0 + z_{0'95}\frac{\sigma}{\sqrt{n}} \;\middle|\; \bar{X}\sim N\left(3, \frac{2}{\sqrt{30}}\right)\right)$$

$$= P\left(\bar{X} \le 1'8 + 1'645\cdot\frac{2}{\sqrt{30}} = 2'4007 \;\middle|\; \bar{X}\sim N\left(3, \frac{2}{\sqrt{30}}\right)\right)$$

$$= P\left(\frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}} \le \frac{2'4007-3}{2/\sqrt{30}}\right) = P(Z \le -1'6415) = \text{pnorm}(-1'6415)$$

$$= \underline{\underline{0'0503}}$$

$$\text{Potencia}(\mu_1=3) = P(\text{rechazar } H_0 \mid H_0 \text{ Falsa}) = 1-\beta(\mu_1=3)$$

$$= 1 - 0'0503 = \underline{0'9497}$$

Problema 3

$X \sim N(\mu, \sigma)$ $\qquad \sum_{i=1}^{n} x_i = 100 \qquad \sum_{i=1}^{n} x_i^2 = 600 \qquad n = 25$

$\bar{x} = \frac{100}{25} = 4 \qquad s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{600 - 25 \cdot 4^2}{24} = 8'3333$

$s = \sqrt{s^2} = \sqrt{8'3333} = 2'8868$

Paso 1

$H_0: \mu = 2'5 \equiv \mu_0$

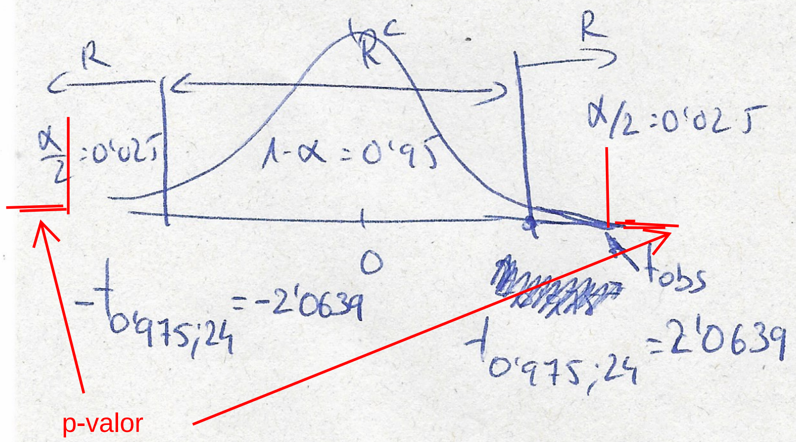
$H_1: \mu \neq 2'5$ $\qquad \alpha = 0'05$

Paso 2

- Población normal
- Varianza desconocida

$\Big\} \Rightarrow \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$

Paso 3



$t_{1-\frac{\alpha}{2}; n-1} = t_{0'975; 24} = 2'0639$

$t_{obs} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{4 - 2'5}{2'8868/\sqrt{25}} = 2'5981$

Paso 4

Rechazamos Ho porque $\begin{cases} t_{obs} = 2'5981 > t_{0'975;24} = 2'0639 \\ \text{p-valor} = 0'0158 < \alpha = 0'05 \end{cases}$

$\text{p-valor} = 2 \cdot P(t_{24} > t_{obs}) = 2\left(1 - pt(2'5981, 24)\right) = 0'0158$

Paso 5: Hay evidencia que sugiere que la media no es igual a 2'5. con un nivel de significación del 5%

b) Potencia($\mu_1 = 4$)

Calculamos usando R

power.t.test (n = 25, delta = 1'5, sd = 2'5, type = "one.sample", strict = TRUE)

Potencia = 0'8207

En la práctica veremos el uso del power.t.test.

$$delta = |\mu_1 - \mu_0| = |4 - 2'5| = 1'5$$

Problema 4

X: puntuación en el test de los estudiantes de ingeniería

Y: " " " inglés.

$n_x = 64 \quad X \sim N(\mu_x, \sigma_x = 100) \qquad n_y = 144 \quad Y \sim N(\mu_y, \sigma_y = 108)$

Paso 1:

$H_0: \mu_x = \mu_y \Rightarrow \mu_x - \mu_y = 0 = \delta_0$

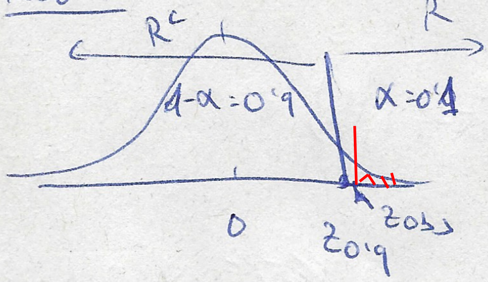
$H_1: \mu_x > \mu_y \Rightarrow \mu_x - \mu_y > 0$

Paso 2

- 2 ~~poblaciones independientes~~ poblaciones muestras independientes: porque los estudiantes son cogidos al azar de un grupo y de otro
- Asumimos X, Y normales.
- Varianzas conocidas.

$$\Rightarrow \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim N(0,1)$$

Paso 3



$z_{1-\alpha} = z_{0'9} = 1'2816 \qquad \bar{x} - \bar{y} = 20$

$$z_{obs} = \frac{(\bar{x} - \bar{y}) - \delta_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{20 - 0}{\sqrt{\frac{100^2}{64} + \frac{108^2}{144}}} = 1'2985$$

Paso 4

p-valor $= P(Z > z_{obs} = 1'2985) = 1 - \text{pnorm}(1'2985) = 0'0971$

Rechazamos Ho porque
$$\begin{cases} z_{obs} = 1'2985 > z_{1-\alpha} = 1'2816 \\ \text{p-valor} = 0'0971 < \alpha = 0'1 \end{cases}$$

Paso 5: Hay evidencia que sugiere que la diferencia entre la nota media en el test para los alumnos en ingeniería es mayor que la nota media de los estudiantes de inglés, con un nivel de significación del 5%

(b) ~~$H_0: \mu_x-\mu_y=0$ ($\mu_0(X,Y)=0$) $\Rightarrow \bar{X}-\bar{Y} \sim N\left(\mu_0(X,Y), \sigma_{\bar{X}-\bar{Y}}=\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right)$~~

~~$\beta(\mu_1(X,Y)=40)=$~~ ~~$\sigma_{\bar{X}-\bar{Y}}=\sqrt{\frac{100^2}{64}+\frac{108^2}{144}}=15'4029$~~

~~$\beta(\mu_1(X,Y)=40)=P(\text{no Rechazar } H_0 \mid H_0 \text{ Falsa})$~~

~~$=P\left(\frac{(\bar{X}-\bar{Y})-0}{\sigma_{\bar{X}-\bar{Y}}} \le z_{0'9} \mid H_1\right)=P(\bar{X}-\bar{Y} \le z_{0'9}\,\sigma_{\bar{X}-\bar{Y}} \mid H_1)$~~

~~$=P\left(\bar{X}-\bar{Y} \le 1'2816\sqrt{\frac{100^2}{64}+\frac{108^2}{144}}\right)$~~

$H_0: \mu_x-\mu_y=0=\delta_0 \Rightarrow \bar{X}-\bar{Y} \sim N\left(\mu_0(X,Y), \sigma_{\bar{X}-\bar{Y}}=\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}\right)$

$H_1: \mu_x-\mu_y > 0$

$\sigma_{\bar{X}-\bar{Y}}=\sqrt{\frac{100^2}{64}+\frac{108^2}{144}}=15'4029$

Suponemos que la verdadera diferencia de medias es 40

$\mu_1(X,Y)=$ $\mu_x-\mu_y=40=\delta_1$

$\beta(\delta_1=40)=P(\text{No Rechazar } H_0 \mid H_0 \text{ Falsa})$

$=P\left(\frac{(\bar{X}-\bar{Y})-0}{\sigma_{\bar{X}-\bar{Y}}} \le z_{0'9} \mid H_1\right)=P(\bar{X}-\bar{Y} \le z_{0'9}\cdot\sigma_{\bar{X}-\bar{Y}} \mid H_1)$

$=P(\bar{X}-\bar{Y} \le 1'2816\cdot 15'4029 \mid H_1)=P(\bar{X}-\bar{Y} \le 19'7404 \mid \bar{X}-\bar{Y} \sim N(\underbrace{40}_{\delta_1}, \underbrace{15'40}_{\sigma_{\bar{X}-\bar{Y}}}$

$=P\left(\frac{(\bar{X}-\bar{Y})-\delta_1}{\sigma_{\bar{X}-\bar{Y}}} \le \frac{19'7404-40}{15'4029}=-1'3154\right)=P(Z\le -1'3154)$

$=\text{qnorm}(-1'3154)=\underline{0'0942}$

Problema 5

X: nivel de satisfacción de los graduados de la facultad X

Y: " " " Y

$n_x = 11 \qquad n_y = 15$

a) Paso 1

$H_0: \mu_x = \mu_y \rightarrow \mu_x - \mu_y = 0 = \delta_0 \qquad \alpha = 0'05$

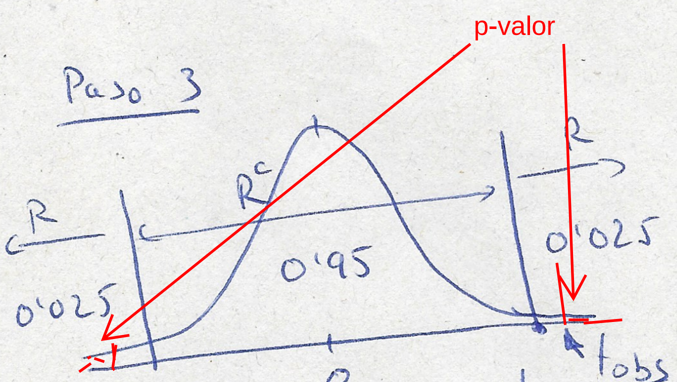
$H_1: \mu_x \neq \mu_y \rightarrow \mu_x - \mu_y \neq 0$

Paso 2

- 2 poblaciones. Muestras independientes porque los graduados se cogen al azar de X e Y.
- Asumimos X e Y normales (deberiamos comprobar con un QQ-plot
- Variantas desconocidas e iguales.

$$\rightarrow \frac{(\bar{X}-\bar{Y}) - \delta_0}{\sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}} \sim t_{n_x+n_y-2} \qquad \text{con} \quad S_p^2 = \frac{(n_x-1)S_x^2 + (n_y-1)S_y^2}{n_x+n_y-2}$$

Paso 3



$-t_{0'975;24} = -2'0639$

$t_{0'975;24} = 2'0639$

$S_x = 8'4067 \qquad S_y = 9'4491$
(Calculado de la muestra)

$S_p = 9'0294$

$$t_{obs} = \frac{(\bar{X}-\bar{Y}) - \delta_0}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{(83'4545 - 76) - 0}{9'4491\sqrt{\frac{1}{11} + \frac{1}{15}}} = 2'0798$$

Paso 4

$$p\text{-valor} = 2 \cdot P(t_{24} \geq 2'0798) = 2 \cdot (1 - pt(2'0798, 24)) = \underline{0'0484}$$

Rechazamos Ho porque $\begin{cases} |t_{obs}| = 2'0794 > 2'0639 \\ p\text{-valor} = 0'0484 < \alpha = 0'05 \end{cases}$

Paso 5: Hay evidencia que sugiere que los niveles de satisfacc son distintos para X e Y. con un nivel de significaci del 5%.

b) No lo podemos calcular con R porque necesitariamos que $n_x = n_y$. Este apartado no lo vamos a considerar.

Problema 6

Lo resolveremos con R en la práctica (Ver práctica 5-6: contrastes).

Problema 7

X: Produccion de cebada en 1932 en Morris

Y: Produccion de cebada en 1932 en Crookston

Paso 1

$H_0: \mu_X - \mu_Y = 0 = \delta_0$

$H_1: \mu_X - \mu_Y > 0$

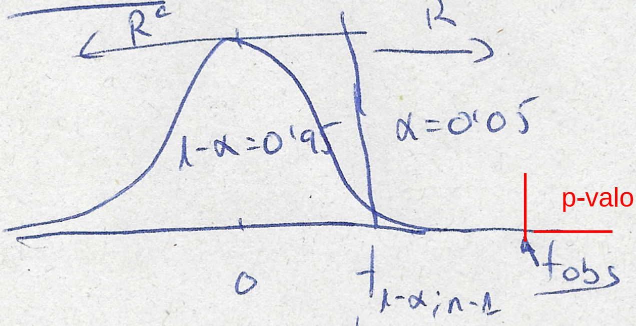
$\alpha = 0'05$

Paso 2

- 2 Poblaciones. Muestras relacionadas porque medimos la producción de una misma variedad de cebada en 2 sitios distintos.
- Varianza desconocida
- Asumimos $X-Y \sim$ Normal (Deberíamos comprobar). con un QQ-plot o eda

$\Rightarrow \dfrac{\bar{D} - \delta_0}{S_D/\sqrt{n}} \sim t_{n-1} = t_9$

Paso 3



$t_{0'95;9} = 1'8331$

$S_D = 5'1931$, $\bar{d} = 10'333$ } (calculado de la muestra)

$t_{obs} = \dfrac{\bar{d} - \delta_0}{S_D/\sqrt{n}} = \dfrac{10'333 - 0}{5'1931/\sqrt{10}} = 6'2924$

Paso 4

$$p\text{-valor} = P(t_9 \geq 6'2924) = 0'0001$$

Rechazamos Ho porque $\begin{cases} t_{obs} = 6'2924 > t_{0'95;9} = 1'8331 \\ p\text{-valor} = 0'0001 < \alpha = 0'05 \end{cases}$

Paso 5: Tenemos evidencia que sugiere que la producción ~~de ceba~~ media de cebada en Morris es mayor que la de Crookston con un nivel de significación del 5%

Problema 8

X: Diámetro de las arandelas de 4 cm

Paso 1

$H_0: \sigma_X^2 = 0'004 = \sigma_0^2$

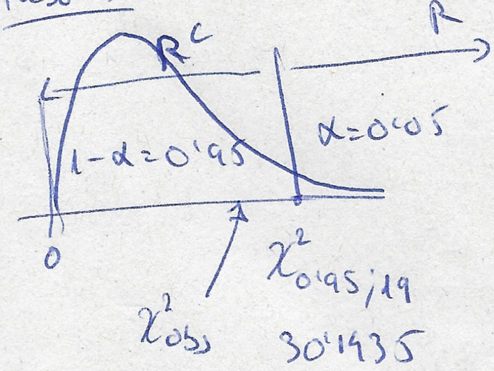
$H_1: \sigma_X^2 > 0'004$ $\qquad \alpha = 0'05$

Paso 2

· Asumimos $X \sim$ Normal (deberíamos justificar con un QQ-plot, eda, test de shapiro-wilk, etc.)

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Paso 3



$s^2 = 0'0053$ (calculado de los datos)

$$\chi^2_{obs} = \frac{(n-1)\cdot s^2}{\sigma_0^2} = \frac{(20-1)\cdot 0'0053}{0'004} = 25'2637$$

Paso 4

$p\text{-valor} = P(\chi^2_{19} \geq \chi^2_{obs}) = P(\chi^2_{19} \geq 25'2637) = 1 - \text{pchisq}(25'2637, 19)$
$= 0'152$

No rechazamos Ho porque $\begin{cases} \chi^2_{obs} < \chi^2_{0'95;19} \\ p\text{-valor} = 0'152 < \alpha = 0'05 \end{cases}$

Paso 5: Con un nivel de significación del 5% no hay evidencia suficiente para sugerir que la varianza del diametro de las arandela haya aumentado.

Problema 9

X: nivel de alcohol medido con alcoholímetro de la marca X

Y: nivel de alcohol medido con alcoholímetro de la marca Y

Paso 1

$H_0: \sigma_x^2 = \sigma_y^2 \Rightarrow \sigma_x^2 / \sigma_y^2 = 1$

$H_1: \sigma_x^2 < \sigma_y^2 \Rightarrow \sigma_x^2 / \sigma_y^2 < 1$

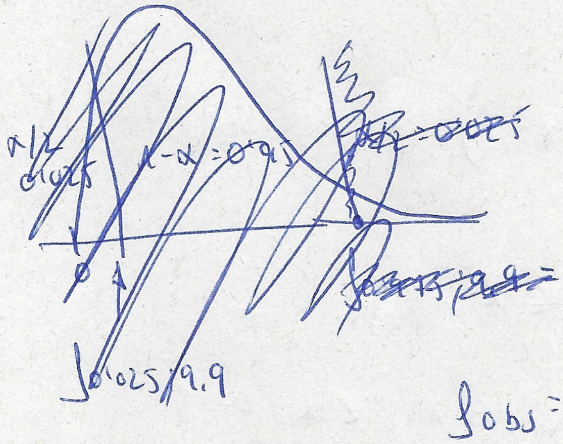
$\alpha = 0'05$

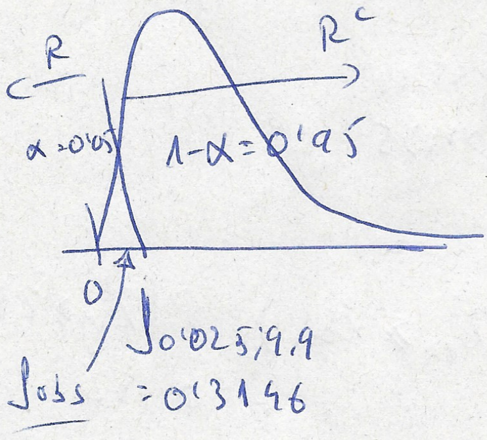
Paso 2

- 2 Poblaciones. Muestras independientes porque son aparatos cogidos al azar de cada compañía.
- Asumimos que X, Y ~ Normal (Deberíamos comprobar)

$$\Rightarrow \frac{S_x^2}{S_y^2} \sim F_{n_x-1;\, n_y-1} = F_{9,9}$$

Paso 3





$$f_{obs} = \frac{s_x^2}{s_y^2} = \frac{0'000133}{0'0006} = 0'222$$

$\left.\begin{array}{l} S_x^2 = 0'000133 \\ S_y^2 = 0'0006 \end{array}\right\} \Rightarrow$ Calculados de la muestra.

Paso 4

$$p\text{-valor} = P(F_{9,9} \leq 0'222) = \underline{0'0176}$$

Rechazamos Ho porque $\begin{cases} f_{obs} = 0'222 < F_{0'05;9,9} = 0'3146 \\ p\text{-valor} = 0'0176 < \alpha = 0'05 \end{cases}$

Paso 5

Con un nivel de significación del 5% tenemos evidencia suficiente que indica que la variabilidad del valor de alcohol en sangre medido con el alcoholímetro de la compañía X es menor que la de la compañía Y.

Problema 10

X: nº de graduados de un total de 500 que encuentran trabajo en su campo de estudios.

$X \sim B(n=500, \pi_x)$ $\qquad p = \frac{90}{500}$

a)

<u>Paso 1</u>

$H_0: \pi = 0'20 = \pi_0$

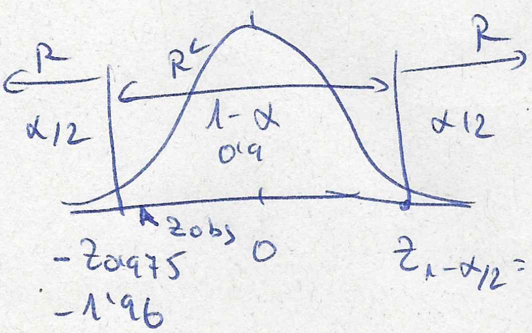
$H_1: \pi \neq 0'20$ $\qquad \alpha = 0'05$

<u>Paso 2</u>

Muestra grande $n \cdot p = 90 \geq 10$ $\quad n \cdot (1-p) = 410 \geq 10$

$$\frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \stackrel{\cdot}{\sim} N(0,1)$$

<u>Paso 3</u>



$Z_{1-\alpha/2} = z_{0'975} = 1'96$

$$z_{obs} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{\frac{90}{500} - 0'2}{\sqrt{\frac{0'2 \cdot 0'8}{500}}} = -1'118$$

<u>Paso 4</u>

$p\text{-valor} = P(Z < z_{obs}) \cdot 2 = P(Z \leq -1'118) \cdot 2 = 0'2636$

No rechazar H_0 porque
$$\begin{cases} \cdot \ |z_{obs}| < z_{1-\frac{\alpha}{2}} \\ \cdot \ p\text{-valor} = 0'2636 \geq \alpha = 0'05 \end{cases}$$

Paso 5

Con un nivel de significación del 5% no hay evidencia suficiente que sugiera que la proporción de graduados que encuentran trabajo en su campo sea distinto del 0'2.

B) Condiciones $\left\{\right.$ Muestras grandes (ya comprobado)

$$CI_{95\%}(\pi) = p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} = \frac{90}{500} \pm 1'96 \sqrt{\frac{\frac{90}{500} \cdot \frac{410}{500}}{500}}$$

$$= [0'1488 , 0'2161]$$

Con una confianza del 95% ~~el verd~~ ~~la verd~~ la proporción de graduados que encuentran trabajo en su campo está entre 0'1488 y 0'2161.

Problema 11

X: num. de pacientes que sufren ataques al corazon tomando aspirina de un total de $n_x = 11037$

Y: " " " " tomando placebo de un total de $n_y = 11034$

$X \sim B(n_x = 11037, \pi_x)$ $\quad$ $Y \sim B(n_y = 11034, \pi_y)$

$p_x = 104/11037 = 0'0094$ $\quad$ $p_y = 189/11034 = 0'0171$

Paso 1

$H_0: \pi_x = \pi_y \rightarrow \pi_x - \pi_y \geq 0$

$H_1: \pi_x < \pi_y \rightarrow \pi_x - \pi_y < 0$

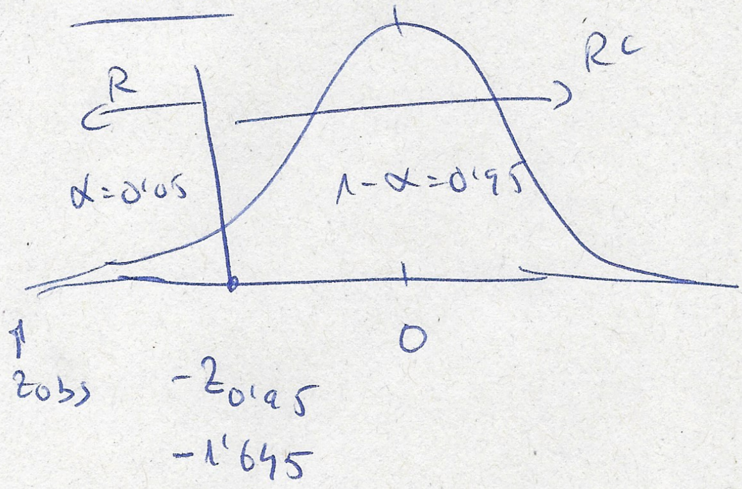
Paso 2

· 2 poblaciones independientes

· Muestras grandes $\begin{cases} n_x \cdot p_x = 104 \geq 10 & n_x(1-p_x) = 10933 \geq 10 \\ n_y \cdot p_y = 189 \geq 10 & n_y(1-p_y) = 10845 \geq 10 \end{cases}$

$$\frac{(p_x - p_y) - 0}{\sqrt{\frac{p(1-p)}{n_x} + \frac{p(1-p)}{n_y}}} \approx N(0,1) \qquad p = \frac{X+Y}{n_x+n_y} = \frac{p_x \cdot n_x + p_y \cdot n_y}{n_x + n_y}$$

Paso 3



$$z_{obs} = \frac{(p_x - p_y) - 0}{\sqrt{\frac{p(1-p)}{n_x} + \frac{p(1-p)}{n_y}}} = -5'0014$$

$$p = \frac{x+y}{n_x+n_y} = \frac{104+189}{22071} = 0'0133$$

Paso 4

$p\text{-valor} = P(Z < z_{obs}) = P(Z \leq -5'0014) \simeq 0$

Rechazamos Ho porque $\begin{cases} \bullet\ z_{obs} = -5'0014 < -z_{0'95} = -1'645 \\ \bullet\ p\text{-valor} \simeq 0 < \alpha = 0'05 \end{cases}$

Paso 5

Existe evidencia suficiente que indica que tomar aspirina es beneficioso para la prevención de los ataques al corazon en los médicos, con un nivel de significación del 5%.