UNIVERSIDAD NACIONAL DE COLOMBIA

Estadística I

Pruebas de Hipótesis para diferencia de medias $\mu_x - \mu_y$

Paso 1: Hipótesis

1. $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x > \mu_y$ $\longleftrightarrow$ 1. $H_0 : \mu_x - \mu_y = \delta_0$ vs $H_1 : \mu_x - \mu_y > \delta_0$
2. $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x < \mu_y$ $\longleftrightarrow$ 2. $H_0 : \mu_x - \mu_y = \delta_0$ vs $H_1 : \mu_x - \mu_y < \delta_0$
3. $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$ $\longleftrightarrow$ 3. $H_0 : \mu_x - \mu_y = \delta_0$ vs $H_1 : \mu_x - \mu_y \neq \delta_0$

Paso 2: Estadístico de prueba → I.C para $\mu_x - \mu_y$

Paso 3: Regiones de rechazo → I.C para $\mu_x - \mu_y$

Mirar si las m.a

- No normal, $n, m \geq 30$
  - σ^2 conocidas
  - σ^2 desconocidas
- Normal
  - σ^2 conocidas
  - σ^2 desconocidas
    - Si $\sigma_x^2 = \sigma_y^2$
    - Si $\sigma_x^2 \neq \sigma_y^2$

P.H para diferencia de medias $\mu_x - \mu_y$

Paso 2: Estadísticos de prueba.

Caso I: m.as de poblaciones no normales con $n,\ m$ grandes ($n, m \geq 30$)

De lo visto en clases anteriores se puede decir que la estadística de prueba asociada a las hipótesis anteriores se relaciona con la diferencia de las medias muestrales $\bar{X} - \bar{Y}$ (estimador puntual de $\mu_x - \mu_y$), así los estadísticos de prueba para el Caso I están dados por:

- Si σ_x^2 y σ_y^2 son conocidas, el estadístico de prueba está dado por:

$$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim_{aprox} N(0,1) \text{ bajo } H_0$$

- Si σ_x^2 y σ_y^2 son desconocidas, el estadístico de prueba está dado por:

$$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim_{aprox} N(0,1) \text{ bajo } H_0$$

P.H para diferencia de medias $\mu_x - \mu_y$

Paso 2: Estadísticos de prueba.

Caso II: m.as de poblaciones normales

- Si σ_x^2 y σ_y^2 son conocidas, el estadístico de prueba está dado por:

$$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim N(0,1) \text{ bajo } H_0$$

- Si σ_x^2 y σ_y^2 son desconocidas, es necesario usar una distribución que tenga en cuenta el desconocimiento de las varianzas poblacionales bajo normalidad. Esa es la distribución T de student. Dependiendo de la relación entre las varianzas poblacionales, se distinguen dos casos:

P.H para diferencia de medias $\mu_x - \mu_y$

Paso 2: Estadísticos de prueba.

Caso II: m.as de poblaciones normales

- Si $\sigma_x^2 = \sigma_y^2$, el estadístico de prueba está dado por:

$$T_c = \frac{\bar{X} - \bar{Y} - \delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T_{(n+m-2)} \text{ bajo } H_0, \text{ con } S_p = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$$

- Si $\sigma_x^2 \neq \sigma_y^2$, el estadístico de prueba está dado por:

$$T_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim T_v \text{ bajo } H_0, \text{ con } v = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{\left(\frac{S_x^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_y^2}{m}\right)^2}{m-1}}$$

P.H para diferencia de medias $\mu_x - \mu_y$

Paso 1: Hipótesis

1. $H_0: \mu_x - \mu_y = \delta_0$ **vs** $H_1: \mu_x - \mu_y > \delta_0$
2. $H_0: \mu_x - \mu_y = \delta_0$ **vs** $H_1: \mu_x - \mu_y < \delta_0$
3. $H_0: \mu_x - \mu_y = \delta_0$ **vs** $H_1: \mu_x - \mu_y \neq \delta_0$

Paso 2: Estadístico de prueba

P.H para $\mu_x - \mu_y$ → Mirar si las m.a

- No normal, $n, m \geq 30$
  - σ^2 conocidas: $$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$
  - σ^2 desconocidas: $$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$
- Normales
  - σ^2 conocidas: $$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$
  - σ^2 desconocida:
    - Si $\sigma_x^2 = \sigma_y^2$, $T_c = \dfrac{\bar{X} - \bar{Y} - \delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T_{(n+m-2)}$
    - $\sigma_x^2 \neq \sigma_y^2$, $T_c = \dfrac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim T_v$

Paso 3: Regiones de rechazo

1. $R_c: \{Z_c \mid Z_c > Z_\alpha\}$	**1.** $R_c: \{T_c \mid T_c > t_{\alpha,gl}\}$
2. $R_c: \{Z_c \mid Z_c < -Z_\alpha\}$	**2.** $R_c: \{T_c \mid T_c < -t_{\alpha,gl}\}$
3. $R_c: \{Z_c \mid Z_c < -Z_{\alpha/2} \vee Z_c > Z_{\alpha/2}\}$	**3.** $R_c: \{T_c \mid T_c < -t_{\alpha/2,gl} \vee T_c > t_{\alpha/2,gl}\}$

Paso 4: Valores p

1. $VP = P(Z \geq Z_c)$	**1.** $VP = P(T_{gl} \geq T_c)$
2. $VP = P(Z \leq Z_c)$	**2.** $VP = P(T_{gl} \leq T_c)$
3. $VP = P(\lvert Z \rvert \geq \lvert Z_c \rvert)$	**3.** $VP = P(\lvert T_{gl} \rvert \geq \lvert T_c \rvert)$

Se rechaza H_0 si VP<α

UNIVERSIDAD NACIONAL DE COLOMBIA

P.H para diferencia de medias $\mu_x - \mu_y$

¿Cómo determinar si las varianzas poblacionales son iguales o diferentes?

Sean S_x^2 y S_y^2 las varianzas muestrales de dos muestras aleatorias independientes de tamaños n y m de dos poblaciones normales. Los elementos de una prueba de hipótesis para comparar las varianzas de las dos poblaciones se presentan a continuación:

F de Snedecor

$qF(p, v_1, v_2)$

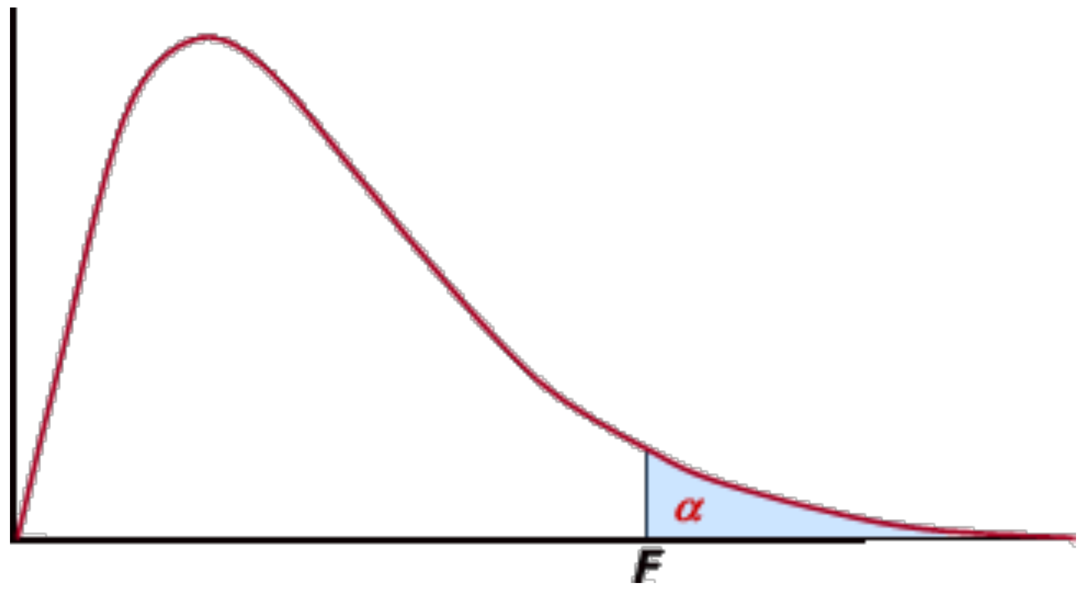
P.H para σ_x^2/σ_y^2

Paso 1: Hipótesis

1. $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 > \sigma_x^2$	$\longleftrightarrow$	**1.** $H_0 : \sigma_x^2/\sigma_y^2 = 1$ vs $H_1 : \sigma_x^2/\sigma_y^2 > 1$
2. $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 < \sigma_x^2$	$\longleftrightarrow$	**2.** $H_0 : \sigma_x^2/\sigma_y^2 = 1$ vs $H_1 : \sigma_x^2/\sigma_y^2 < 1$
3. $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 \neq \sigma_x^2$	$\longleftrightarrow$	**3.** $H_0 : \sigma_x^2/\sigma_y^2 = 1$ vs $H_1 : \sigma_x^2/\sigma_y^2 \neq 1$

Paso 2: Estadístico de prueba

$$F_c = \frac{S_x^2}{S_y^2} \sim F_{(n-1,m-1)}$$



Teorema: Si $F \sim F_{(v_1,v_2)}$, entonces

$$\frac{1}{F} \sim F_{(v_2,v_1)}$$

Paso 3: Regiones de rechazo

1. $R_c : \{F_c \mid F_c > F_{\alpha,(n-1,m-1)}\}$

2. $R_c : \{F_c \mid F_c < 1/F_{\alpha,(m-1,n-1)}\}$

3. $R_c : \{F_c \mid F_c < 1/F_{\alpha/2,(m-1,n-1)} \vee F_c > F_{\alpha/2,(n-1,m-1)}\}$

Otra Forma!!!!

σ_x^2/σ_x^2

Se calcula un I.C al $100(1-\alpha)\,\%$ de la forma:

$$\left(\frac{S_x^2}{S_y^2} \times \frac{1}{f_{\alpha/2,(n-1,m-1)}}, \frac{S_x^2}{S_y^2} \times f_{\alpha/2,(m-1,n-1)}\right)$$

Si el intervalo contiene al 1, se puede afirmar que $\sigma_x^2 = \sigma_y^2$

UNIVERSIDAD NACIONAL DE COLOMBIA

P.H para diferencia de medias $\mu_x - \mu_y$

Ejemplo:

fabricante desea saber si la tensión promedio de su hilo difiere a la de su más cercano competidor. Para ello, se observaron las tensiones de 100 hilos de su marca y 95 de la marca del competidor bajo condiciones controladas. Los resultados obtenidos se muestran en la siguiente tabla:

	Media muestral	Varianza muestral
Hilos marca propia	110.8	10.2
Hilos marca competidor	108.2	12.4

¿Existe evidencia suficiente para sugerir que hay una diferencia entre las tensiones promedio de ruptura de los dos hilos? Utilice una significancia de 0.02.

X: hilos marca propia $\quad \bar{X} = 110.8 \quad S_X^2 = 10.2 \quad n = 100$

Y: " competidor $\quad \bar{Y} = 108.2 \quad S_Y^2 = 12.4 \quad m = 95$

P.H para diferencia de medias $\mu_x - \mu_y$

1) Hipótesis

$\alpha = 0.02 \Rightarrow \alpha/2 = 0.01$.

$H_0: \mu_x - \mu_y = 0$ vs $H_1: \mu_x - \mu_y \neq 0$

2) E.P:

$$Z_c = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} = \frac{110.8 - 108.2 - 0}{\sqrt{\frac{10.2}{100} + \frac{12.4}{96}}} = 5.39$$

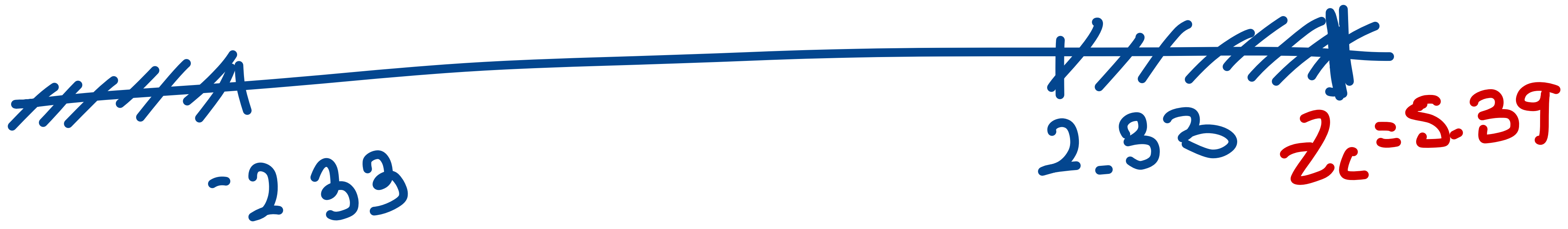
3) R.R:

$R_c = \{ Z_c \mid Z_c < -Z_{\alpha/2} \vee Z_c > Z_{\alpha/2} \}$

$R_c = \{ Z_c \mid Z_c < -Z_{0.01} \vee Z_c > Z_{0.01} \}$

$R_c = \{ Z_c \mid Z_c < -2.33 \vee Z_c > 2.33 \}$

Como $Z_c \in R_c$, se rechaza H_0 usando $\alpha = 0.02$



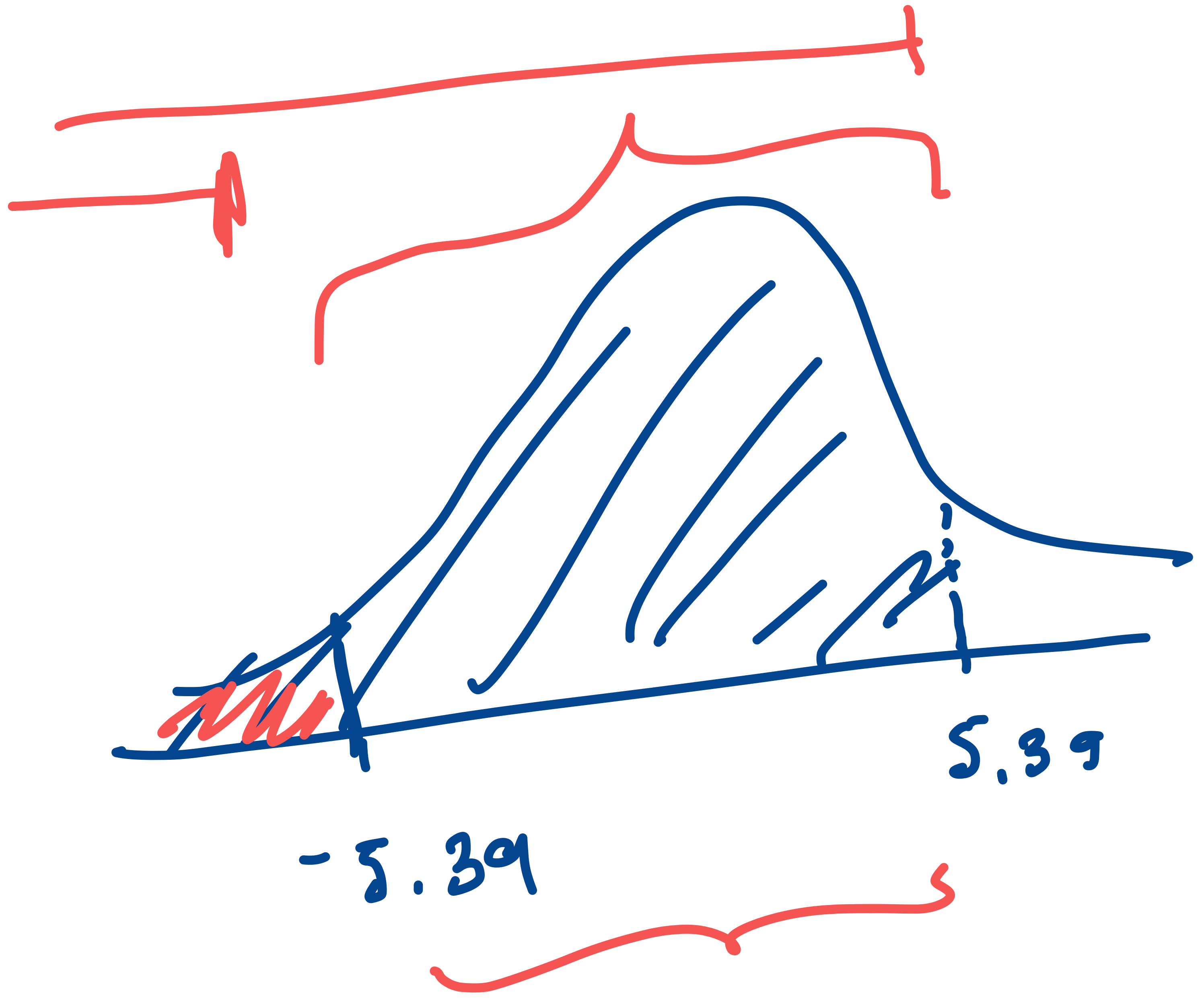
P.H para diferencia de medias $\mu_x - \mu_y$

4. Valor-P:

$Vp = P(|z| \geq |z_c|) = P(|z| > 5.39) = 1 - P(|z| < 5.39)$

$= 1 - \underbrace{P(-5.39 < z < 5.39)} = 1 - [\text{pnorm}(5.39) - \text{pnorm}(-5.39)]$

$= 0.00000007045$



Como $vp < \alpha = 0.02$, se rechaza Ho.

P.H para diferencia de medias $\mu_x - \mu_y$

Ejemplo:

Se llevó a cabo un estudio para determinar el grado en el cual el alcohol entorpece la habilidad de pensamiento para llevar a cabo determinada tarea. Para ello, se seleccionaron al azar 40 personas de distintas características y se les pidió que participaran en el experimento. Después de proporcionarles la información pertinente, cada persona llevó a cabo la tarea sin nada de alcohol en su organismo. Tras esto, cada persona consumió una cantidad suficiente de alcohol para tener un contenido en su organismo de 0.1% y la tarea volvió a llevarse a cabo. La información obtenida del estudio es la siguiente:

	Tiempo medio	**Desviación estándar**
Resultados antes del consumo de alcohol	$33.90\ min$	$10.34\ min$
Resultados después del consumo de alcohol	$47.90\ min$	$11.19\ min$

a. ¿Puede concluirse con un nivel de significancia de 0.05 que el tiempo promedio después de consumir alcohol es mayor que el tiempo promedio antes del consumo por más de 10 minutos?

b. ¿Cuál debe ser el valor del tiempo promedio después del consumo de alcohol para que no se pueda concluir la afirmación del inciso a?

P.H para diferencia de medias $\mu_x - \mu_y$

X: tiempos Antes de consumir alcohol. $\bar{X} = 33.9$ $S_x = 10.34$ $n = 40$

Y: " después " " ". $\bar{Y} = 47.9$ $S_y = 11.19$ $m = 40$

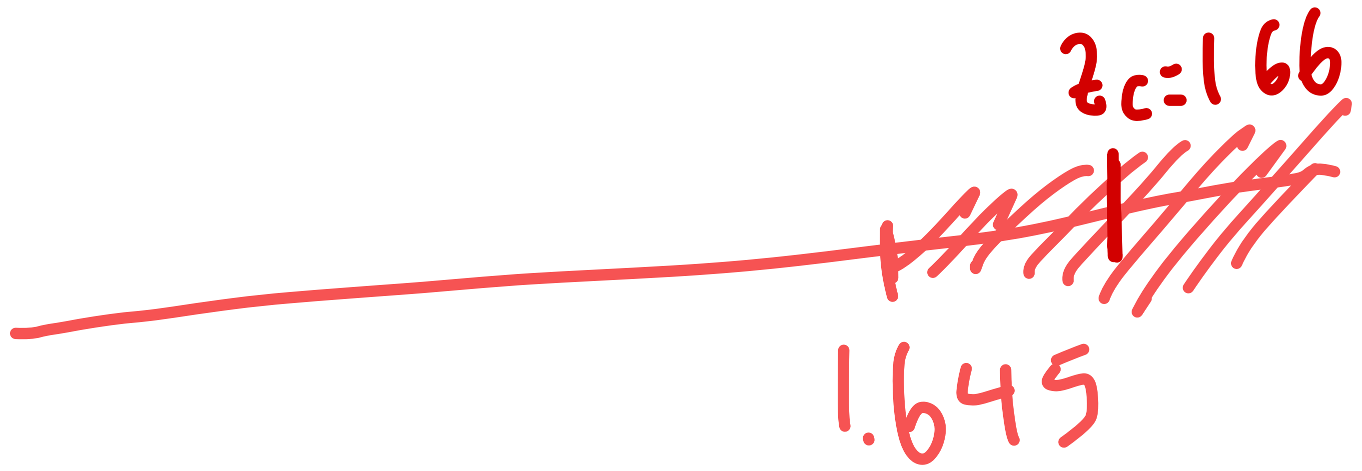
1) $H_0 : \mu_Y - \mu_X \leq 10$ vs $H_1 : \mu_Y - \mu_X > 10$ ✓

2) $$z_c = \frac{(\bar{Y} - \bar{X}) - \delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_Y^2}{m}}} = \frac{(47.9 - 33.9) - 10}{\sqrt{\frac{(10.34)^2}{40} + \frac{(11.19)^2}{40}}} = 1.66$$

P.H para diferencia de medias $\mu_x - \mu_y$

3) R.R:

$$R_c = \{ z_c \mid \overbrace{z_c > z_\alpha} \} = \{ z_c \mid z_c > z_{0.05} \} = \{ z_c \mid z_c > 1.645 \}$$



Como $z_c \in R_c$, Se rechaza Ho usando $\alpha = 5\%$

4) Valor-P

$$VP = P(z > z_c) = P(z > 1.66) = 1 - P(z \leq 1.66)$$

1- pnorm(1.66)

$$= 0.048$$

como $vp < \alpha = 0.05$ Se rechaza Ho.

Hay suficiente evidencia muestral para ....

P.H para diferencia de medias $\mu_x - \mu_y$

b) Para no concluir, no se debe rechazar Ho

$$Z_c < Z_\alpha$$

$$Z_c < 1.645$$

$$\frac{(\bar{Y} - 33.9) - 10}{\sqrt{\frac{(10.34)^2}{40} + \frac{(11.19)^2}{40}}} < 1.645$$

$$\bar{Y} < 1.645\sqrt{\frac{(10.34)^2}{40} + \frac{(11.19)^2}{40}} + 10 + 33.9$$

$$\Rightarrow \bar{Y} < \underbrace{47.86}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

P.H para diferencia de medias $\mu_x - \mu_y$

P.H para diferencia de medias $\mu_x - \mu_y$

Ejemplo:

Se desea comprar una gran cantidad de bombillas y se tiene que elegir entre las marcas A y B. Para ello, se compraron 15 focos de cada marca, y se encontró que las bombillas probadas de la marca A tuvieron un tiempo de vida medio de 1.120 horas, con una desviación estándar de 75 horas; mientras que las de la marca B tuvieron un tiempo de vida medio de 1.064 horas, con una desviación estándar de 82 horas. Por experiencia se sabe que la vida útil de cada una de las marcas se distribuye de manera normal. Con un nivel de significancia de 0.05, ¿Cuál marca de bombillas debería comprar la empresa?

X: Vida útil marca A. $n=15$ $X \sim$ normal $\bar{X} = \underbrace{1120}$ $S_x = 75$

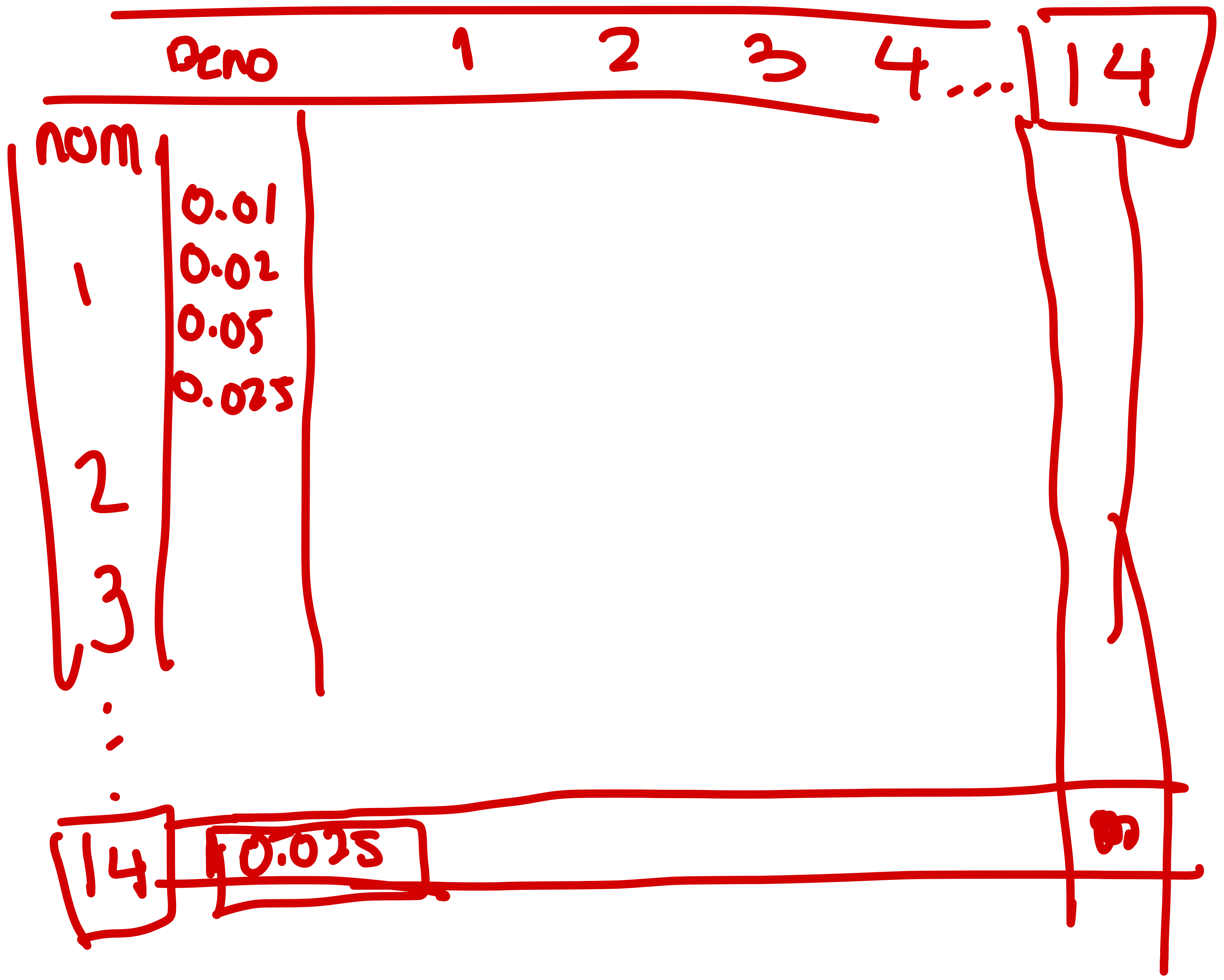
Y: " " " B. $m=15$ $Y \sim$ normal $\bar{Y} = \underbrace{1064}$ $S_y = 82$.

$\alpha = 0.05$

P.H para diferencia de medias $\mu_x - \mu_y$

1) $H_0: \mu_x - \mu_y \leq 0 \quad H_1: \mu_x - \mu_y > 0$ ✓

$qf(\overset{1-\alpha/2}{0.975}, \overset{num}{14}, \overset{den}{14})$



σ_x^2 y σ_y^2 $\rightarrow$ son desconocidas.

Se debe probar si son iguales o no:

$$\left(\frac{S_x^2}{S_y^2} * \frac{1}{f_{\alpha/2}(n-1, m-1)} , \frac{S_x^2}{S_y^2} \times f_{\alpha/2}(m-1, n-1) \right)$$

$1-\alpha/2$

$$\left(\frac{(75)^2}{(82)^2} * \frac{1}{\underbrace{f_{0.025}(14,14)}} , \frac{S_x^2}{S_y^2} * f_{0.025}(14,14) \right)$$

como el I.C contiene al 1.

$$\left(\frac{(75)^2}{(82)^2} * \frac{1}{2.98} , \frac{75^2}{82^2} \cdot 2.98 \right) \Rightarrow (0.2807, 2.49) \quad \frac{\sigma_x^2}{\sigma_y^2} = 1 \text{ es decir}$$

son iguales

P.H para diferencia de medias $\mu_x - \mu_y$

2) E.P:

$$T_c = \frac{\bar{X} - \bar{Y} - \delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1120 - 1064}{78.58\sqrt{\frac{1}{15} + \frac{1}{15}}} = 1.95$$

$$S_p = \sqrt{\frac{(15-1)\,75^2 + (15-1)\,82^2}{15+15-2}} = 78.58$$

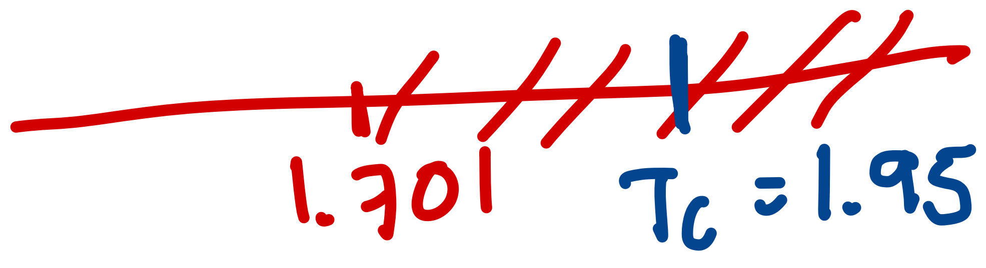
3) R.R:

$$R_c = \left\{T_c \mid T_c > t_{\alpha,(n+m-2)}\right\} = \left\{T_c \mid T_c > t_{0.05\,(28)}\right\}$$

Usando $\alpha = 0.05$

Se rechaza Ho.

$$= \left\{T_c \mid T_c > 1.701\right\}$$



P.H para diferencia de medias $\mu_x - \mu_y$

vp: $P(T_{n+m-2} > T_c) = P(T_{28} > 1.95) = 0.03062$

→ pt(1.95, 28, lower.tail = F)

$vp < \alpha = 0.05$

Se rechaza Ho.

$vp \approx 0$

$vp = 0.8$

→ $0.025 < vp < 0.05$

P.H para diferencia de medias $\mu_x - \mu_y$

Gracias

Clase preparada por: Verónica Guarín