COMPARACIÓN DE MEDIAS

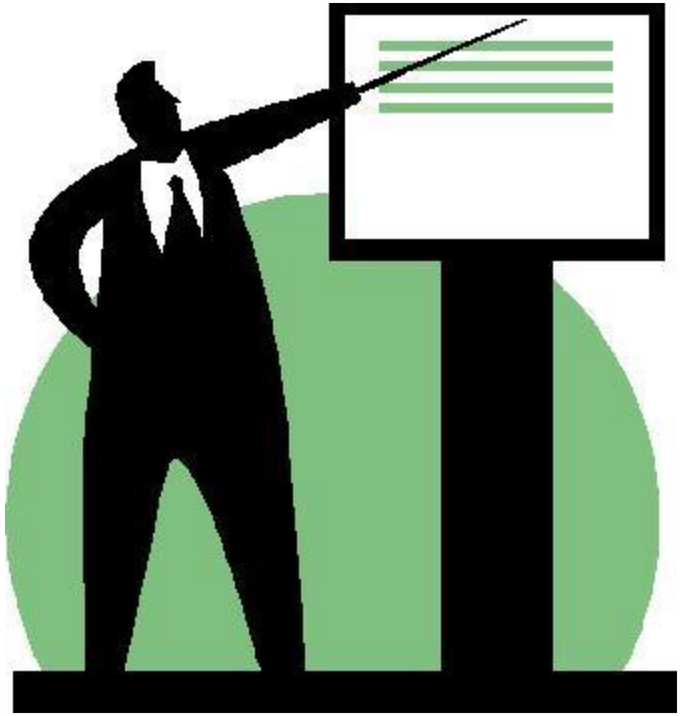
Oscar Centeno Mora

INTRODUCCIÓN

- El tema anterior presentó la prueba de hipótesis como método de inferencia estadística.

- El desarrollo de la prueba de hipótesis se utilizó bajo el estadístico de diferencia sobre una única media.

- Las formas bilaterales y unilaterales para plantear la prueba de hipótesis en la comparación de una media fueron las siguientes:



$$H_0: \mu = c_1 \qquad H_0: \mu \leq c_1 \qquad H_0: \mu \geq c_1$$

$$H_A: \mu \neq c_1 \qquad H_A: \mu > c_1 \qquad H_0: \mu < c_1$$

$$z = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}$$

INTRODUCCIÓN

- ¿Qué otras pruebas o suposiciones se podrían realizar?

INTRODUCCIÓN

- Nos podría interesaría saber si hay diferencias estadísticas entre hombres y mujeres, o entre dos tipos de clases de animales, entre la pierna derecha y la izquierda, y si el porcentaje difiere para dos tipos de poblaciones, entre otros.

- A las diferencias anteriores se les conoce como comparación de medias, y se suelen trabajar bajo el enfoque de pruebas de hipótesis, cambiando tanto la forma de plantear la hipótesis como los estadísticas de prueba.

- El capítulo de comparación de medias presentará las diferencias de medias para dos poblaciones independientes, dependientes y para la comparación cuando se trabaja con proporciones o porcentajes.

ÍNDICE

1 Medias independientes

2 Comparación por parejas

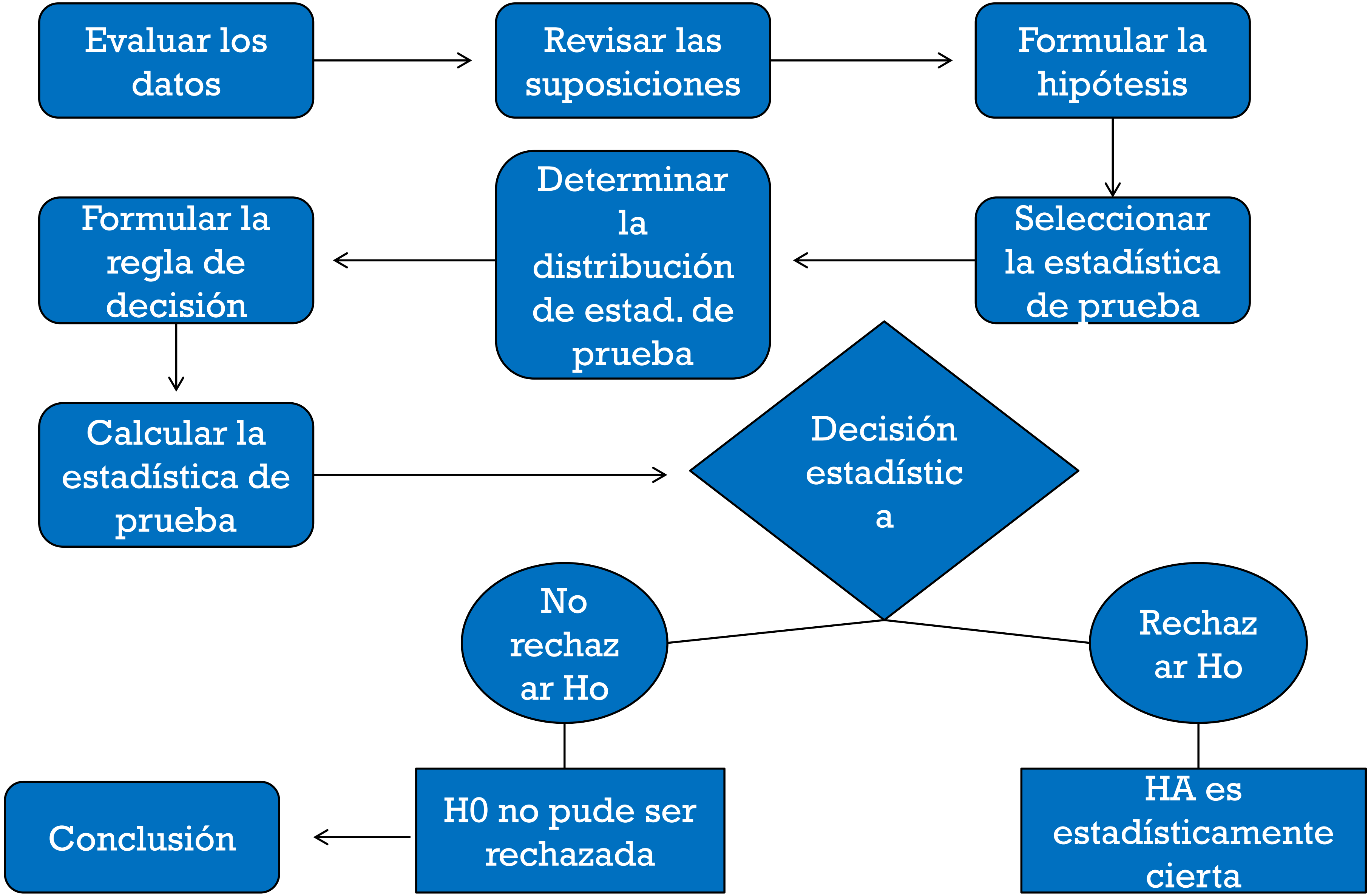
3 La comparación en la proporción

4 Comparación de proporciones

5 Resumen de las pruebas y los estadísticos asociados

RECORDATORIO….

RECORDATORIO: PASOS DE UNA PRUEBA DE HIPÓTESIS: RESUMEN



ÍNDICE

1 Medias independientes

ÍNDICE

1

Medias independientes

Variancias conocida

Variancias desconocida similares

Variancia desconocida diferentes

MEDIAS INDEPENDIENTES CON IGUAL VARIANCIA

- El primer tipo de comparación de medias se realiza bajo la suposición de dos población que son independientes.

- La prueba de hipótesis que involucra la diferencia entre las medias de 2 poblaciones se utiliza con mas frecuencia para determinar si es razonable o no concluir si dos poblaciones son distintas entre si.

- En este caso se puede encontrar diversas posibilidades:

  -Comparación de medias con variancias conocidas

  - Comparación de medias con variancias desconocidas iguales

  - Comparación de medias con variancias desconocidas diferentes

- Recordar que todas las pruebas pueden ser **bilaterales o unilaterales**.

EJEMPLOS DE ESTAS PRUEBAS

- ¿Hay diferencias en los salarios promedio entre hombres y mujeres?

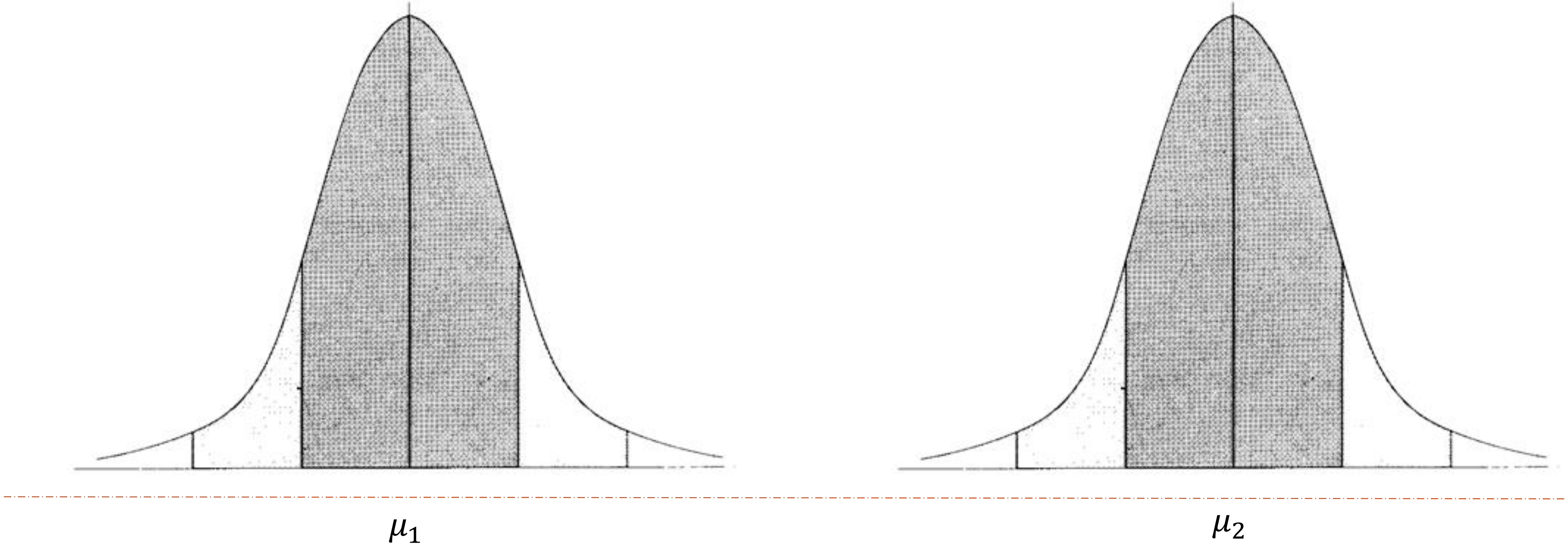
- ¿Luego de los 40 años, se mueren más hombres que mujeres?

- ¿Nacen más hombres que mujeres?

- ¿Las mujeres suelen poseer mayor problemas de estrés que los hombres?

MEDIAS INDEPENDIENTES CON IGUAL VARIANCIA

- A grosso modo, al realizar comparaciones de medias para dos poblaciones supuestas como "independientes", se busca ver si a nivel de la posición, existe alguna diferencia en μ_1 y μ_2.



¿ $\mu_1 \neq \mu_2$?

MEDIAS INDEPENDIENTES CON VARIANCIA CONOCIDA

- La prueba de hipótesis de diferencias de medias de dos poblaciones independientes establece si es posible determinar o no la diferencia de medias de las poblaciones.

- Las pruebas de hipótesis se plantean como:

$$H_0: \mu_1 - \mu_2 = 0 \qquad H_A: \mu_1 - \mu_2 \neq 0$$

$$H_0: \mu_1 - \mu_2 \geq 0 \qquad H_A: \mu_1 - \mu_2 < 0$$

$$H_0: \mu_1 - \mu_2 \leq 0 \qquad H_A: \mu_1 - \mu_2 > 0$$

- En norma, las etapas vistas en el capítulos se deben de llevar a cabo, pero para acortar camino, lo que interesa en esta prueba específica es el estadístico de prueba:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

MEDIAS INDEPENDIENTES CON VARIANCIA CONOCIDA

- Un equipo de investigadores desea saber si los datos que han recolectado proporcionan la evidencia suficiente para indicar una diferencia entre las concentraciones medias de acido úrico en el suero de individuos normales e individuos con síndrome de Down. Los datos consisten en las lecturas de acido urico en el suero de 12 individuos can sindrome de Down y 15 individuos sanos. Las medias son $\bar{x}_1 = 4.5$ mg/100 ml y $\bar{x}_2 = 3.4$ mg/100 ml.

- Solución: Se podrá decir que los datos ofrecen evidencia de que las medias poblacionales son diferentes si es posible rechazar la hipótesis nula que indica que las medias son iguales. La conclusión se obtiene mediante el procedimiento de los diez pasos de la prueba de hipótesis.

MEDIAS INDEPENDIENTES CON VARIANCIA CONOCIDA

1. **Datos**: ver el enunciado

2. **Supuestos**: los datos corresponden a dos muestras aleatorias simples e independientes, cada una extraída de una población que sigue una distribución normal, con una variancia igual a 1 para la población con síndrome de Down, y de 1.5 para la población sana.

3. **Hipótesis**:

$$H_0: \mu_1 - \mu_2 = 0 \qquad H_A: \mu_1 - \mu_2 \neq 0$$

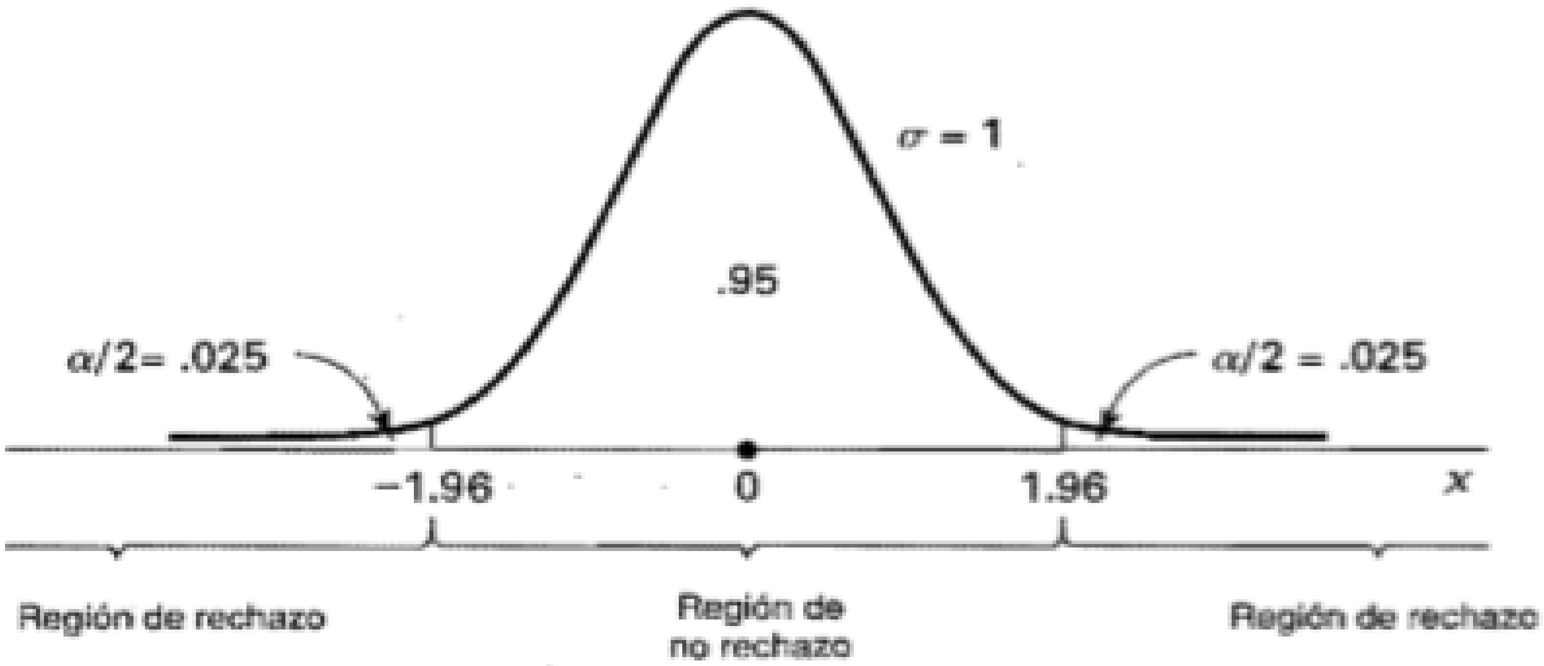
4. **Estadística de prueba**:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

5. **Distribución de la estadística de prueba**: Cuando la hipótesis nula es verdadera, la estadística de prueba sigue una distribución normal estándar.

MEDIAS INDEPENDIENTES CON VARIANCIA CONOCIDA

6. Regla de decisión: Sea α=0.05. Los valores críticos de z son $\pm$ 1.96. Se rechaza H_0 a menos que $-1.96 < z_{calculada} < 1.96$. Las regiones de rechazo y no rechazo se muestran a continuación.



MEDIAS INDEPENDIENTES CON VARIANCIA CONOCIDA

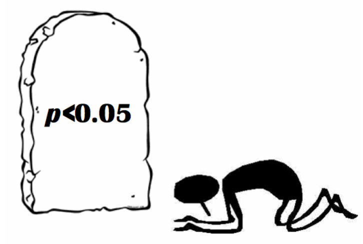
7. **Cálculo de la estadística de prueba**:

$$z = \frac{(4.5 - 3.4) - 0}{\sqrt{\frac{1}{12} + \frac{1.5}{15}}} = \frac{1.1}{0.4282} = 2.57$$

8. **Decisión estadística**: Se rechaza H_0 porque 2.57 >1.96.

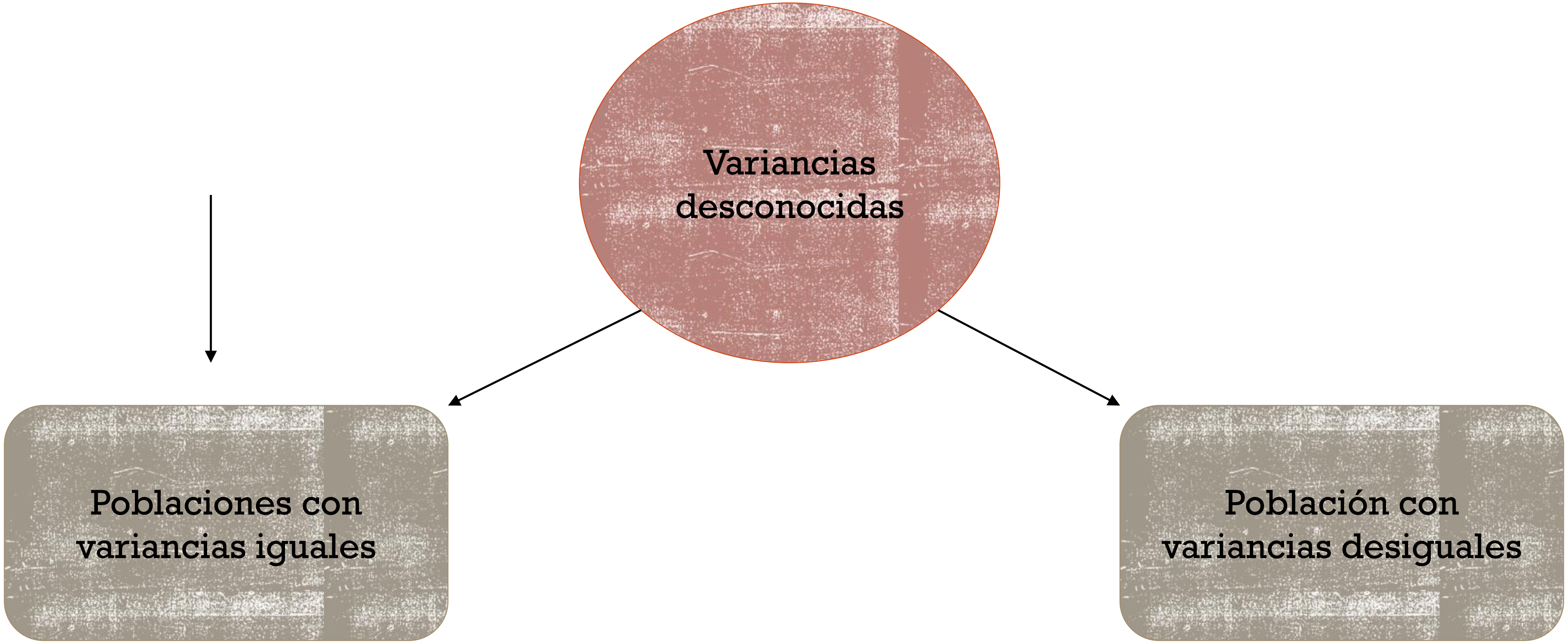
9. **Conclusión**: se concluye que, de acuerdo con estos datos, hay indicios de que las medidas de las poblaciones son diferentes.

10. **Valor de p**. Para esta prueba, p=0.0102.



MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

- Cuando las variancias poblacionales no se conocen, existen dos posibilidades. Las variancias de dos poblaciones pueden ser iguales o pueden ser diferentes. Se considera primero el caso donde se sabe, o es razonable suponer, que son iguales.

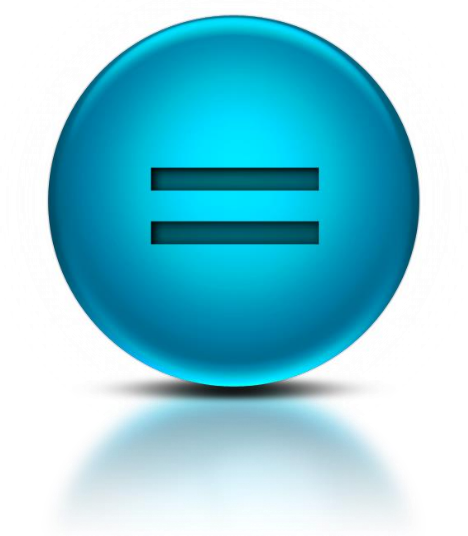


MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

¡PONDEREMOS LAS VARIANCIAS!

MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

Poblaciones con variancias iguales. Cuando se desconocen las variancias de las poblaciones, pero se supone que son iguales, es adecuado ponderar las variancias de las muestras por medio de la siguiente fórmula:



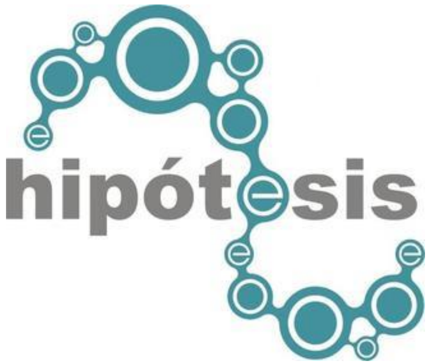
$$S^2{}_p = \frac{(n_1 - 1)S^2{}_1 + (n_2 - 1)S^2{}_2}{n_1 + n_2 - 2}$$

Se trabaja sobre el mismo planteamiento de hipótesis:

$$H_0: \mu_1 - \mu_2 = 0 \qquad H_A: \mu_1 - \mu_2 \neq 0$$

$$H_0: \mu_1 - \mu_2 \geq 0 \qquad H_A: \mu_1 - \mu_2 < 0$$

$$H_0: \mu_1 - \mu_2 \leq 0 \qquad H_A: \mu_1 - \mu_2 > 0$$

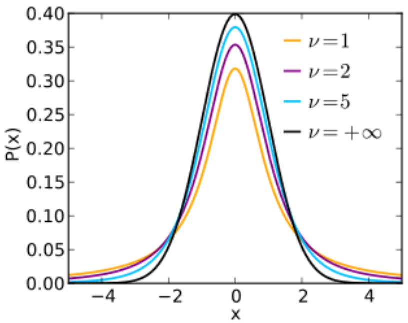


MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

- Cuando cada una de las dos muestras aleatorias simples e independientes se extrae de una población que sigue una distribución normal y las dos poblaciones tienen variancias desconocidas pero iguales, la estadística de prueba para $H_0: \mu_1 = \mu_2$ se obtiene a partir del siguiente estadístico

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

- El anterior estadísticos sigue una distribución t de Student con n1+n2 – 2, grados de libertad.



MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

- ¿Cómo utilizamos la tabla de la t de Student?

Tabla 2. Distribución t de Student

gl \ α/2	0,40	0,30	0,20	0,10	0,050	0,025	0,010	0,005	0,001	0,0005
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,192	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,853	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,256	0,530	0,853	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,259	0,851	1,303	1,648	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,256	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

- Un estudio tiene como objetivo examinar las características de destrucción pulmonar en personas que fuman cigarros antes de desarrollar un marcado enfisema pulmonar. Se practicaron mediciones de tres índices de destrucción pulmonar en los pulmones de personas longevas que no fumaban y en personas con tabaquismo que murieron repentinamente fuera del hospital por causas no respiratorias. Una calificación alta indica un mayor daño pulmonar. La siguiente tabla muestran las calificaciones producidas en una muestra de nueve personas que no fuman y 16 fumadores. Se pretende saber si es posible concluir, con base en los datos, que las personas que sí fuman, en general, tienen los pulmones mas dañados que las personas no fumadoras.

TABLA 7.3.1 Calificaciones de los índices de destrucción pulmonar para el ejemplo 7.3.2

No fumadores:	18.1,	6.0,	10.8,	11.0,	7.7,	17.9,	8.5,	13.0,	18.9
Fumadores:	16.6,	13.9,	11.3,	26.5,	17.4,	15.3,	15.8,	12.3,	18.6,
	12.0,	24.1,	16.5,	21.8,	16.3,	23.4,	18.8		

FUENTE: D. H. Eidelman H. Ghezzo, W. D. Kim y M. G. Cosio, "The Destructive Index and Early Lung Destruction in Smokers", *American Review of Respiratory Disease, 144*, 156-159.

MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

1. **Datos**: ver el enunciado

2. **Supuestos**: los datos corresponden a dos muestras aleatorias simples e independientes. No se conocen las variancias poblacionales, pero se supone que son iguales.

3. **Hipótesis**: $H_0: \mu_1 - \mu_2 = 0$ $\quad H_A: \mu_1 - \mu_2 \neq 0$

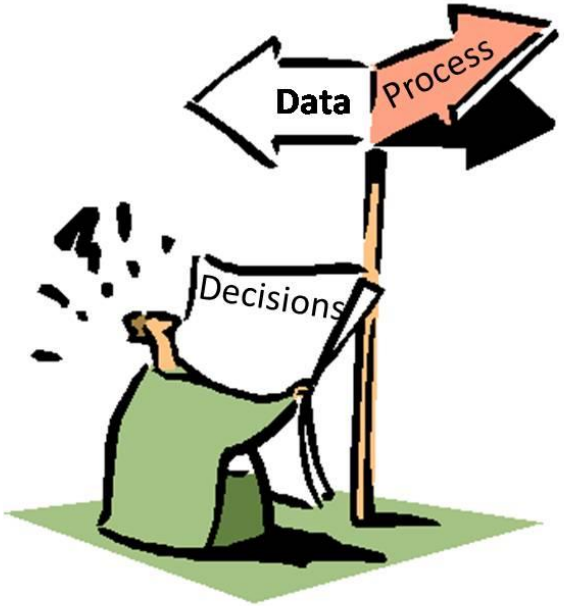
4. **Estadística de prueba**:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

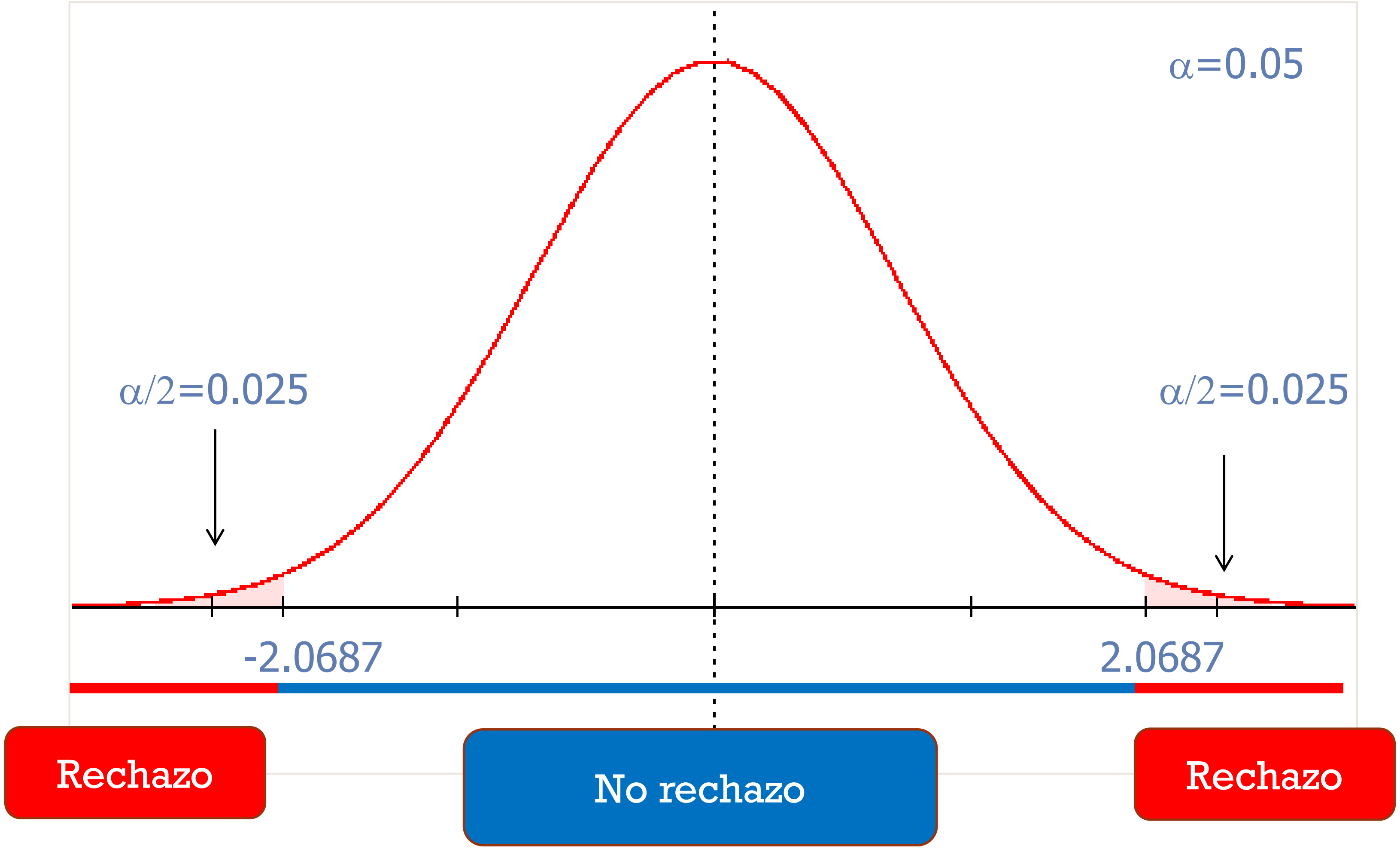
MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

5. **Distribución estadística de prueba**: Cuando la hipótesis nula es verdadera la estadística de prueba sigue una distribución t de Student con n1+n2-2 grados de libertad

6. **Regla de decisión**: Sea α=0.05. Los valores críticos de *t* son ± 2.0687. Se rechaza H_0 a menos que $-2.0687 < t_{calculada} < 2.0687$. Las regiones de rechazo y no rechazo se muestran a continuación (siguiente filmina).



MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA



MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

7. **Cálculo de la estadística de prueba**: A partir de los datos muestrales tenemos que:

$$\bar{x}_1 = 17.5 \; ; \; s_1 = 4.4711$$
$$\bar{x}_2 = 12.4 \; ; \; s_2 = 4.8992$$

Se calcula la variancia ponderada:

$$S^2{}_p = \frac{15(4.4711)^2 + 8(4.8492)^2}{15+8} = 21.2165$$

Y obtenemos el estadístico:

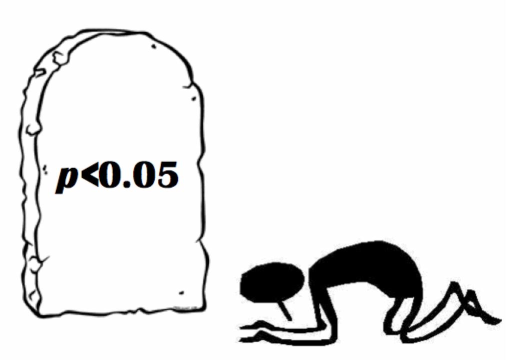
$$t = \frac{(17.5 - 12.4) - 0}{\sqrt{\frac{21.2165}{16} + \frac{21.2165}{9}}} = 2.6573$$

MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

8. **Decisión estadística**: se rechaza H_0 dado que 2.6573 $>$ 2.0687, es decir, 2.6573 cae dentro de la región de rechazo.

9. **Conclusión**: Con base en estos resultados, se puede concluir que las dos medias poblacionales son diferentes, es decir, se concluye que, según indican los índices **del** estudio, las personas que sí fuman tienen los pulmones mas dañador que las personas que no fuman.

10. **Valor de p**: para esta prueba .01> p >.005, porque 2.500 < 2.6573 < 2.8073.



MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

Variancias desconocidas

Poblaciones con variancias iguales

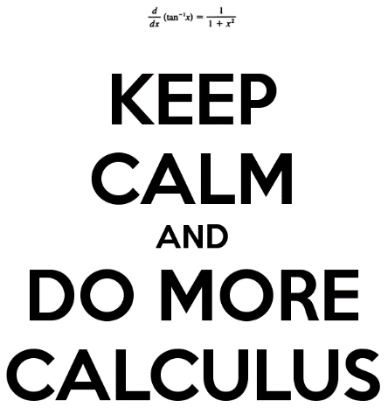
Población con variancias desiguales

MEDIAS INDEPENDIENTES CON VARIANCIA DESCONOCIDA

- Hay varias formas de proceder. En este curso se tomará el siguiente estadístico:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

- Todas las otras etapas son equivalente al caso de las variancias supuestas iguales, con la excepción de la no ponderación de las variancias.



- De igual forma, se utiliza la tabla t y los criterios de rechazo y otros no varían.

EJERCICIOS…

Evans *et al.* (A-8) realizaron un estudio para determinar si la frecuencia y las características de los problemas podiátricos en pacientes de la tercera edad enfermos de diabetes presentan diferencias con respecto a pacientes de la misma edad pero sin diabetes. Los individuos estudiados, internados en una clínica, tenían de 70 a 90 años de edad. Entre los hallazgos de los investigadores están las siguientes estadísticas con respecto a las calificaciones en las mediciones de los reflejos tendinosos profundos:

Muestra	*n*	Media	Desviación estándar
Sin diabetes	79	2.1	1.1
Con diabetes	74	1.6	1.2

EJERCICIOS…

Un investigador de enfermería desea saber si los graduados de los programas de enfermería a nivel bachillerato y los graduados en programas asociados de enfermería difieren en cuanto a las calificaciones medias obtenidas en un estudio de personalidad. Una muestra de 50 graduados de programas asociados (grupo A) y una muestra de 60 graduados de bachillerato (grupo B) proporcionaron las siguientes medias y desviaciones estándar:

Muestra	$\bar{x}$	s
A	52.5	10.5
B	49.6	11.2

Con base en estos datos, ¿qué puede concluir el investigador? Sea $\alpha = .05$.

EJERCICIOS…

Una prueba diseñada para medir las actitudes de las madres en cuanto al trabajo de parto y el periodo de expulsión se aplicó a dos grupos de madres primerizas. La muestra 1 (asistentes) acudieron a clases de instrucción prenatal impartidas en el departamento de salud. La muestra 2 (ausentes) no asistieron a clases. El tamaño de las muestras, medias y desviaciones estándar de las calificaciones para las pruebas son las siguientes:

Muestra	n	$\bar{x}$	s
1	15	4.75	1.0
2	22	3.00	1.5

¿Proporcionan estos datos la evidencia suficiente para indicar que las asistentes, en promedio, tienen calificaciones más altas que las ausentes? Sea $\alpha = .05$.

ÍNDICE

1 Medias independientes

2 Comparación por parejas

COMPARACIÓN PAREADA

- En el análisis sobre la diferencia entre las medias de dos poblaciones, antes se supuso que las muestras provenían de dos poblaciones independientes.

- El supuesto anterior no siempre es válido. Por ejemplo, un método que se utiliza con frecuencia para averiguar la eficacia de un tratamiento o procedimiento experimental es aquel que hace uso de observaciones relacionadas que resultan de muestras no independientes.

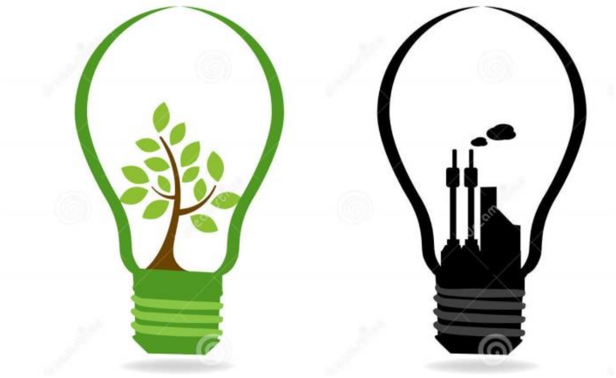
- Una prueba de hipótesis que se basa en este tipo de datos se conoce como prueba de *comparaciones par parejas* o *comparaciones pareadas.*
- *Ejemplos de medidas pareadas son:* mismos individuos pueden ser examinados antes y después, formar parejas de animales del mismo sexo de una misma camada, pares de gemelos, brazo derecho e izquierdo, etc.

COMPARACIÓN PAREADA

- En lugar de llevar a cabo el análisis con observaciones individuales, se puede utilizar d_i, la diferencia entre pares de observaciones, como variables de interés.

$$d_i = xi_1 - xi_2$$

- Cuando las n diferencias de las muestras calculadas de los n pares de mediciones forman una muestra aleatoria simple extraída de una población de diferencias que siguen una distribución normal, la prueba de hipótesis y la estadística de la prueba para la hipótesis respecto a la diferencia de la media poblacional μ_d se establece como



Hipótesis

$$H_0: \mu_d = 0 \qquad H_0: \mu_d \neq 0$$

Estadístico

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}}$$

COMPARACIÓN PAREADA

- Refiriendonos al estadístico anteriores, se tiene que:

Estadístico

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}}$$

- $\bar{d}$ es la diferencia de media muestral
- μ_d es la diferencia de la media poblacional supuesta
- $s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$, n es el número de diferencias muestrales
- s_d es la desviación estándar de las diferencias muestrales
- Cuando H_0 es verdadera, la estadística de prueba sigue una distribución t de Student con n-1 grados de libertad.

COMPARACIÓN PAREADA

- **¡CIUDADO!**

- Dependiendo como se tome los las diferencias d_i, entonces se deberá tener precaución en la elección de la zonas de rechazo y de no rechazo.

- Por ejemplo: $d_i = xi_1 - xi_2$ $\quad$ $d_i = xi_2 - xi_1$

- Los d_i generarán estadísticos similares, pero uno podrá ser negativo y el otro positivo. De ahí que se debe de tener cuidado en la elección del sustractor, dado que este definirá la zona de rechazo.

- Recordar, una PH es ante todo un razonamiento ante un supuesto, por lo que se debe de tener claro qué es lo que se quiere o no rechazar.

COMPARACIÓN PAREADA

Se condujo un estudio para determinar la perdida de peso, la composición corporal, la distribución de grasa corporal y la tasa metabólica en reposo en individuos obesos antes y después de 12 semanas de tratamiento con dieta muy baja en calorías (DMBC), y comparar la hidrodensitometría con el análisis de impedancia bioeléctrica. Los 17 individuos (nueve mujeres y ocho hombres) que participaron en el estudio eran pacientes externos de un programa de tratamiento con base hospitalaria para la obesidad. Los pesos de las mujeres antes y después del tratamiento de 12 semanas de DMBC se muestran en la tabla 7.4.1 (siguiente filmina). Se pretende saber si estos datos ofrecen suficiente evidencia que permita concluir que el tratamiento es eficaz para reducir el peso en mujeres obesas.

COMPARACIÓN PAREADA

- Solución:

- Se puede decir que existe la suficiente evidencia para concluir que el programa de dietas es eficaz si es posible rechazar la hip6tesis nula que indica que el cambio en la media de la poblacion μ_d es cero o positivo. Es posible tomar una decision por medio del procedimiento de los diez pasos de la prueba de hipotesis.

TABLA 7.4.1 Pesos (kg) de mujeres obesas antes (A) y después (D) del tratamiento de 12 semanas con DMBC

A:	117.3	111.4	98.6	104.3	105.4	100.4	81.7	89.5	78.2
D:	83.3	85.9	75.8	82.9	82.3	77.7	62.7	69.0	63.9

COMPARACIÓN PAREADA

1. **Datos**: corresponden al peso de nueve individuos antes y después del programa experimental de dieta. El análisis estadístico se aplica sobre la diferencia entre los pesos de antes y después. Los datos restantes son: $d_i =$ -34.0, -25.5, -22.8, -21.4, -23.1, -22.7, -19.0, -20.5 y -14.3.

2. **Supuestos** : las diferencias que se observan forman la muestra aleatoria simple extraída de una población de diferencias con distribución normal que podrían ser generadas bajo las mismas circunstancias.

3. **Hipótesis**: las hipótesis nula y alternativa deben establecerse de acuerdo con la manera de efectuar la resta de las mediciones para obtener las diferencias. En este caso, se pretende saber si es posible concluir que el programa es eficaz para reducir el peso. Si resulta eficaz, se esperaría que los pesos de después tendieran a ser menores que los pesos de antes. Si se restan los pesos de antes a los pesos de después (D - A), se esperaría que las diferencias tendieran a ser negativas. Además, se esperaría que la media de la población de cada diferencia fuera negativa. Bajo estas condiciones, preguntarse si es posible concluir que el programa es eficaz, es lo mismo que preguntarse si la diferencia de la media poblacional es negativa (menor que cero).

1
$$H_0: \mu_d \geq 0$$
$$H_A: \mu_d < 0$$

2
$$H_0: \mu_d \leq 0$$
$$H_A: \mu_d > 0$$

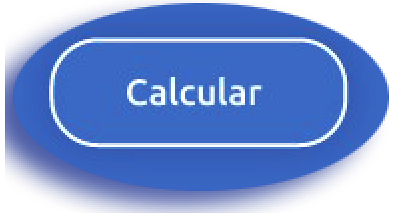
3
$$H_0: \mu_d = 0$$
$$H_A: \mu_d \neq 0$$

COMPARACIÓN PAREADA

4. **Estadística de prueba:**

$$t = \frac{\bar{d} - \mu_d}{s_{\bar{d}}}$$

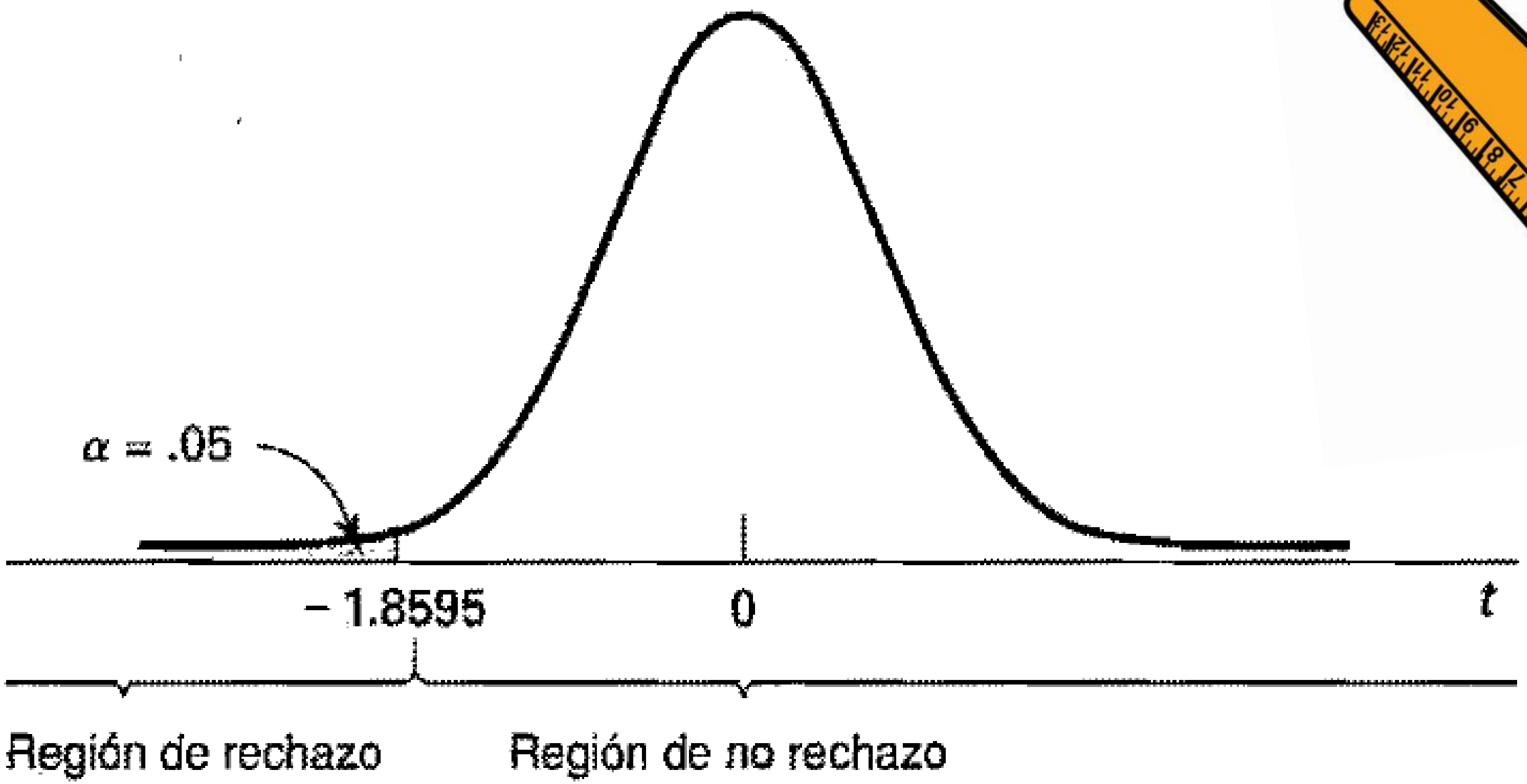
5. **Distribución de la estadística de prueba**: si la hipótesis nula es verdadera, la estadística de prueba sigue una distribución *t* de Student con *n* - 1 grados de libertad.

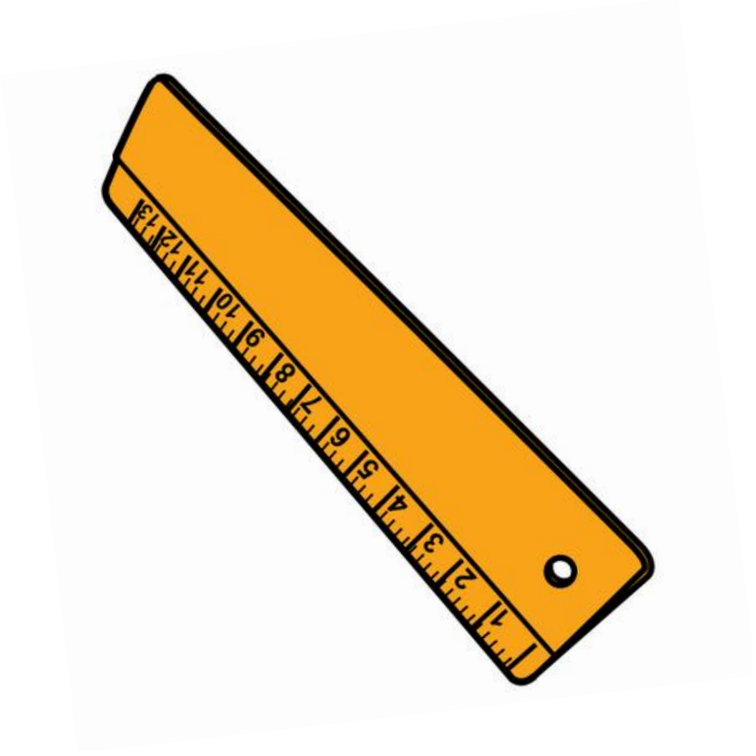


- 6. **Regla de decisión:** sea $\alpha=0.05$. El valor crítico de *t* es -1.8595. Se rechaza *Ho* si el valor calculado de *t* es menor o igual que el valor critico.

COMPARACIÓN PAREADA

6. **Regla de decisión**





COMPARACIÓN PAREADA

- 7. **Cálculo de la estadística de prueba:** a partir de las n 9 diferencias d_i, se calculan las siguientes medidas descriptivas:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{(-34.0) + (-25.5) + \cdots + (-14.3)}{9} = \frac{-203.3}{9} = -22.5889$$

$$s_d{}^2 = \frac{\sum(d_i - d)^2}{n-1} = \frac{n \sum d_i{}^2 - (\sum d_i)^2}{n(n-1)} = \frac{9(4818.69) - (-203.3)^2}{9(8)} = 28.29$$

$$t = \frac{-22.5889 - 0}{\sqrt{28.29/9}} = \frac{-22.5889}{1.77314} = -12.7395$$

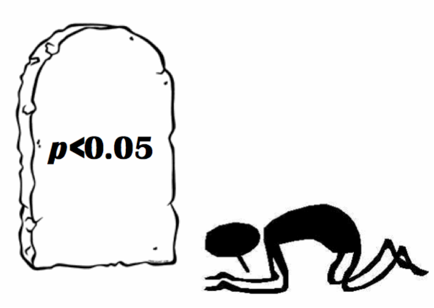
COMPARACIÓN PAREADA

8. **Decisión estadística**: se rechaza *Ho* porque -12.7395 esta en la region de rechazo.

9. **Conclusión**: se puede concluir que el programa de dieta es eficaz

10. **Valor de p**: para esta prueba, $p < .005$ porque $-12.7395 <$

-0.3554.



APLICACIÓN: MEDIAS PAREADAS

7.4.3 El propósito de una de las investigaciones realizadas por Alahuhta *et al*. (A-14) es evaluar la influencia del bloqueo extradural para la operación cesárea en diversas variables hemodinámicas maternas y fetales, simultáneamente, y determinar si el bloqueo modifica la función del miocardio fetal. Los individuos estudiados eran ocho parturientas sanas con 38 a 42 semanas de embarazo de un solo feto, sin complicaciones, que serían sometidas a operación cesárea con anestesia para bloqueo extradural. Los siguientes datos corresponden a los valores inferiores de esta variable en las dos etapas:

Etapa 1:	70	87	72	70	73	66	63	57
Etapa 2:	79	87	73	77	80	64	64	60

FUENTE: Con el permiso de Seppo Alahuhta, M. D.

¿Ofrecen suficiente evidencia estos datos, con un nivel de significación de .05, para indicar que, bajo condiciones similares y generales, la media de la presión arterial diastólica en las madres es diferente en las dos etapas?

APLICACIÓN: MEDIAS PAREADAS

7.4.1 Un artículo publicado por Kashima *et al.* (A-12) describe una investigación relacionada con los padres de niños con retraso mental, en la cual se presenta información sobre la enseñanza de autocuidados en un programa apoyado en diferentes medios de comunicación, principalmente a través de videotapes y manuales de instrucción. Como parte del estudio, participaron 17 familias en el programa de capacitación impartido por personal con amplia experiencia en proyectos de capacitación paterna. Antes y después del programa de capacitación, se aplicó una prueba de comportamiento y descripción a cada jefe de familia o padre principal. El examen evalúa el conocimiento de los principios de modificación del comportamiento. Una calificación alta indica mayor conocimiento. Las siguientes calificaciones corresponden a las pruebas de los jefes de familia, antes y después del programa de capacitación:

Antes:	7	6	10	16	8	13	8	14	16	11	12	13	9	10	17	8	5
Después:	11	14	16	17	9	15	9	17	20	12	14	15	14	15	18	15	9

FUENTE: Datos utilizados con el permiso de Bruce L. Baker, Ph. D.

¿Es posible concluir, con base en estos datos, que el programa de capacitación aumenta el conocimiento respecto a los principios de modificación del comportamiento?. Sea $\alpha = .01$.

ÍNDICE

1 Medias independientes con variancia igual

2 Medias pareadas

3 La comparación en la proporción

COMPARACIÓN EN LA PROPORCIÓN

- La prueba de hipótesis de proporciones poblacionales se realiza casi en la misma forma utilizada para las medias cuando son satisfechas las condiciones necesarias para emplear la curva normal.

- Pueden efectuarse pruebas unilaterales ó bilaterales, dependiendo de la cuestión que se plantee.

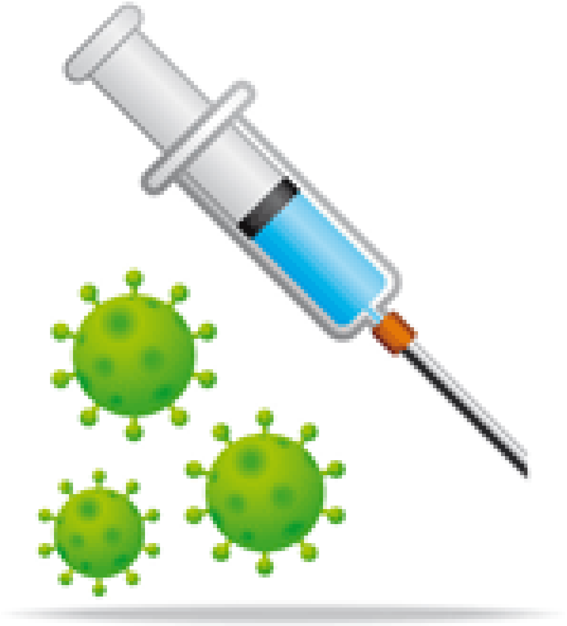
- Las hipótesis y el estadístico se plantean como sigue: ("a" entre 0 y 1)

Hipótesis		Estadístico
$H_0: P = a$	$H_A: P \neq a$	$$z = \frac{\hat{p} - P}{\sqrt{\frac{PQ}{n}}}$$
$H_0: P \geq a$	$H_A: P < a$	
$H_0: P \leq a$	$H_A: P > a$	

COMPARACIÓN EN LA PROPORCIÓN

- En una investigación de consumidores de drogas intravenosas en una ciudad, se encontró que 18 de 423 individuos con VIH positivo. Se pretende saber si es posible concluir que menos de 5 por ciento de los consumidores de drogas intravenosas en la población muestreada tienen VIH positivo.

- Solución: el problema revela la suposición de una prueba de hipótesis referida a una proporción. De igual forma, se procede con los pasos antes visto.



COMPARACIÓN EN LA PROPORCIÓN

1. **Datos**: se obtienen a partir de la respuesta de 423 individuos,

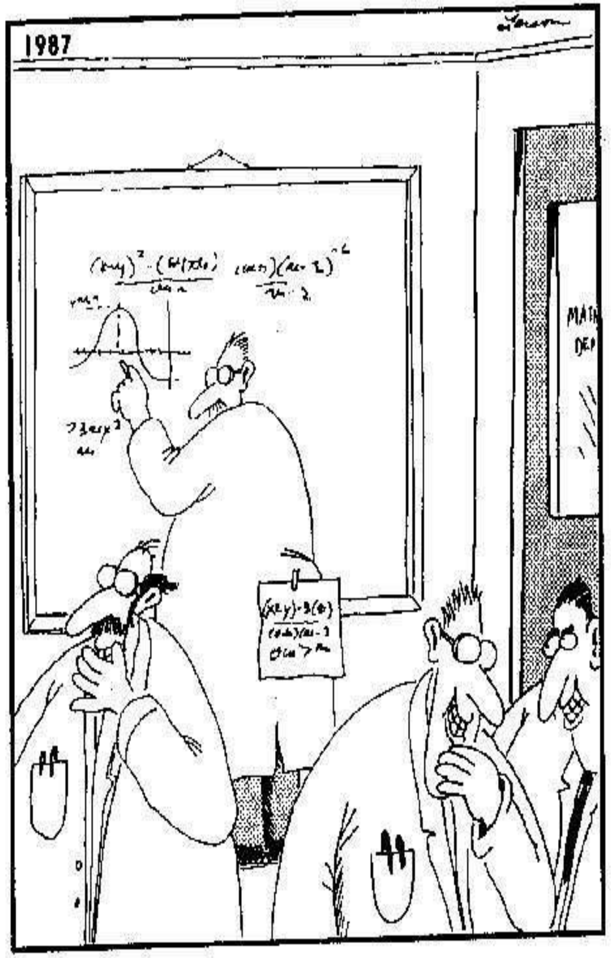
de los cuales 18 tenían la característica de interés. Esto es que el valor estimado es el siguiente:

$$\hat{p} = {}^{18}/_{423} = 0.426$$

2. **Supuestos**: la distribución muestral de $\hat{p}$ sigue una distribución aproximadamente normal… (se aplicamos el TLC…).

3. **Hipótesis**: $H_0: P \geq 0.05$ $\qquad H_A: P < 0.05$

4. **Estadística de prueba:** $z = \dfrac{\hat{p} - P}{\sqrt{\dfrac{PQ}{n}}}$



COMPARACIÓN EN LA PROPORCIÓN

5. **Distribución de la estadística de prueba:** la estadística de prueba sigue una distribución aproximadamente normal con una media de cero.

6. **Regla de decisión**: sea $\alpha=0.05$. El valor crítico de z es -1.645. Se rechaza Ho si el valor de z es $\leq$ -1.645.

7. **Cálculo de la estadística de prueba**:

$$z = \frac{0.0426 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{423}}} = -0.70$$

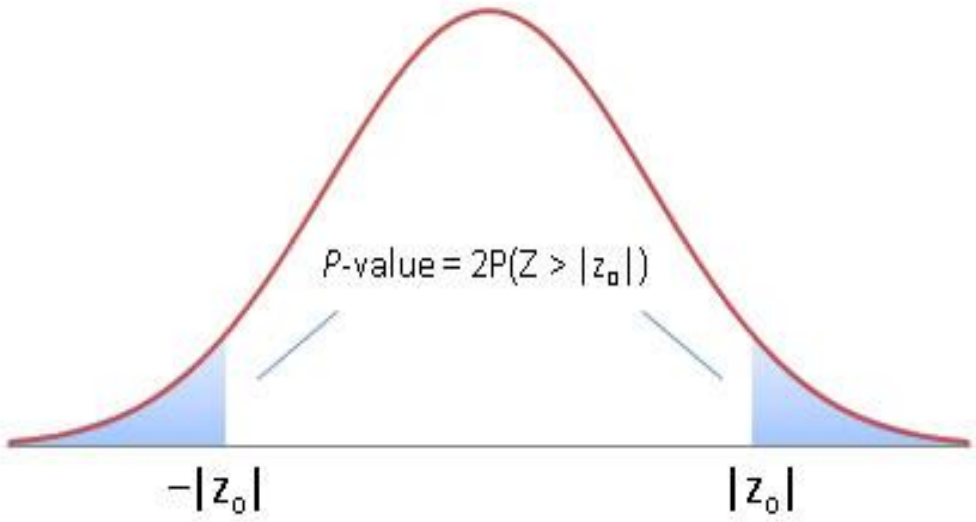
COMPARACIÓN EN LA PROPORCIÓN

8. **Decisión estadística**. No se rechaza H_0 dado que -0.70 > -1.645.

9. **Conclusión**: la proporción de la población que tiene VIH positivo probablemente sea .05 0 mas.

(***otra forma de concluir… ?***)

10. Valor de p: p=0.2420



APLICACIÓN: UNA PROPORCIÓN

7.5.2 En un artículo publicado en la revista *American Journal of Public Health*, Colsher *et al.* (A-19) describen los resultados de una encuesta de salud aplicada a 119 convictos varones de 50 años de edad y mayores residentes de las instalaciones de un centro de readaptación social del estado. Se encontró que 21.6 por ciento de los encuestados dijeron tener antecedentes de enfermedades venéreas. Con base en estos hallazgos, ¿es posible concluir que en la población muestreada más de 15 por ciento tiene antecedentes de enfermedades venéreas? Sea $\alpha = .05$.

7.5.3 Henning *et al.* (A-20) encontraron que 66 por ciento de los niños en una muestra de 670 completaron toda la serie de vacunas contra la hepatitis B. ¿Es posible concluir que, con base en estos datos, en la población muestreada, más de 60 por ciento tienen la serie completa de vacunas contra la hepatitis B? Sea $\alpha = .05$.

ÍNDICE

1 Medias independientes

2 Comparación por parejas

3 La comparación en la proporción

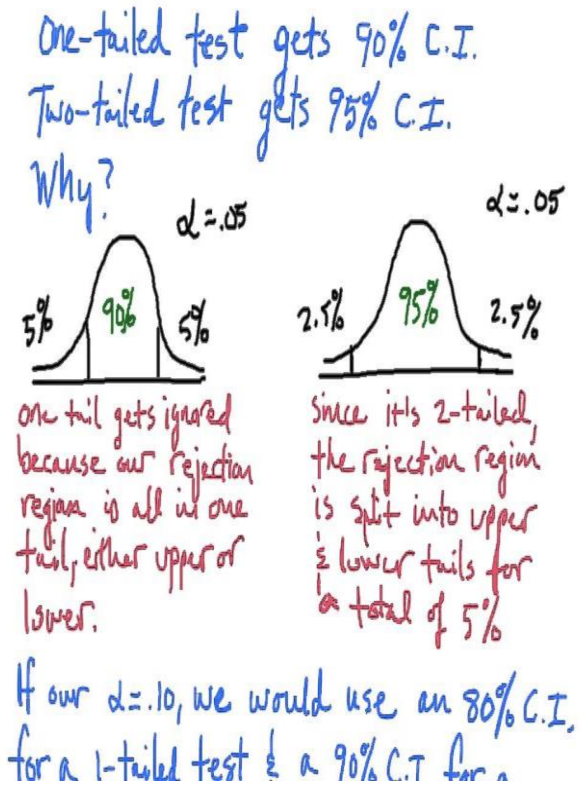
4 Comparación de proporciones

COMPARACIÓN DE PROPORCIONES

- La prueba que se utiliza con mas frecuencia con relación a la diferencia entre las proporciones de dos poblaciones es aquella en la que su diferencia es cero.

- Sin embargo, es posible probar que dicha diferencia es igual a algún otro valor. Es posible efectuar pruebas tanto unilaterales como bilaterales.

- Cuando la hipótesis nula que va a probarse es $P_1 - P_2 = 0$, se podría suponer que ambas proporciones son iguales. Aunque esto se suele utilizar como justificación para combinar los resultados de las dos muestras y obtener una estimación ponderada de la proporción, no se procederá de tal forma.



COMPARACIÓN DE PROPORCIONES

- Para el presente caso se tiene que:

Hipótesis

$$H_0: P_1 - P_2 = 0 \qquad H_A: P_1 - P_2 \neq 0$$

$$H_0: P_1 - P_2 \geq 0 \qquad H_A: P_1 - P_2 < 0$$

$$H_0: P_1 - P_2 \leq 0 \qquad H_A: P_1 - P_2 > 0$$

Estadístico

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p1-p2}}$$

- En donde la expresión σ_{p1-p2} se expresa como sigue:

$$\sigma_{p1-p2} = \sqrt{\frac{P1(1-P1)}{n1} + \frac{P2(1-P2)}{n2}}$$

COMPARACIÓN DE PROPORCIONES

- En un estudio de cuidados nutricionales en asilos para ancianos, encontraron que entre 55 pacientes con hipertensión, 24 tenían una dieta con restricción de sodio. De 149 pacientes sin hipertensión, 36 tenían una dieta sin sodio. ¿Es posible concluir que, en las poblaciones muestreadas, la proporción de pacientes con dieta restringida en sodio es mayor entre pacientes con hipertensión que entre pacientes sin hipertensión?

- Según el tipo de problema, se procede con una prueba de diferencias de proporciones.

COMPARACIÓN DE PROPORCIONES

Datos: los datos corresponden a la información del consumo de sodio en las dietas de los pacientes internados en el asilo de ancianos con y sin hipertensión, tal como se describe en la proposición del ejemplo.

Supuestos: se supone que los pacientes estudiados forman una muestra aleatoria independiente extraída de poblaciones con y sin hipertensión.

Hipótesis: $H_0: P_1 - P_2 \leq 0$ $H_A: P_1 - P_2 > 0$

Estadística de prueba:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p1-p2}} \quad \Rightarrow \quad \sigma_{p1-p2} = \sqrt{\frac{P1(1-P1)}{n1} + \frac{P2(1-P2)}{n2}}$$

COMPARACIÓN DE PROPORCIONES

Distribución de la estadística de prueba: Si la hipótesis nula es verdadera, la estadística de prueba sigue aproximadamente una distribución normal estándar.

Regla de decisión: sea α=0.05. El valor crítico de *z* es 1.645. Se rechaza Ho si el valor de z es mayor que 1.645.

Cálculo de la estadística de prueba:

$$p_1 = \frac{24}{55} = 0.4364$$

$$p_2 = \frac{36}{149} = 0.2416$$

$$n_1 = 55$$

$$n_2 = 149$$

$$\sigma_{p1-p2} = \sqrt{\frac{0.4364(1-0.4364)}{55} + \frac{0.2416(1-0.2416)}{149}}$$

$$= 0.075509$$

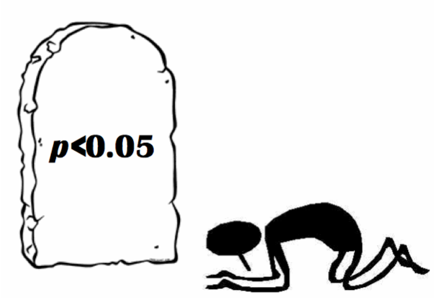
COMPARACIÓN DE PROPORCIONES

Cálculo de la estadística de prueba:

$$z = \frac{(0.4364 - 0.2416) - (0)}{\sqrt{\frac{0.4364(1 - 0.4364)}{55} + \frac{0.2416(1 - 0.2416)}{149}}} = 2.58$$

Decisión estadística: se rechaza dado que 2.58 > 1.645

Conclusión: la proporción de pacientes con dieta restringida en sodio es mayor entre los pacientes hipertensos que entre los pacientes sin hipertensión.



Valor de *p*: para la presente, p=0.0034.

APLICACIÓN: DIFERENCIA DE PROPORCIONES

7.6.1 Babaian y Camps (A-22) afirman que el antígeno específico prostático (AEP), encontrado en las células ductales epiteliales de la próstata, es específico para el tejido prostático y es detectable en el suero de hombres con próstata normal y en hombres con enfermedades benignas o malignas de esta glándula. Los investigadores determinaron los valores de AEP en una muestra de 124 hombres que se sometieron a una biopsia de la próstata. Sesenta y siete hombres tenían concentraciones elevadas de AEP (>4 ng/ml). De estos, a 46 se les diagnóstico cáncer. Diez de los 57 hombres con valores de AEP ≤ 4 ng/ml tenían cáncer. Con base en estos datos, ¿es posible concluir que, en general, los hombres con valores elevados de AEP tienen mayor probabilidad de tener cáncer de próstata? Sea $\alpha = .01$.

7.6.2 La mayoría de las personas que dejan de fumar, se quejan de que al hacerlo suben de peso. Hall *et al.* (A-23) diseñaron una nueva técnica para prevenir que la gente suba de peso, la cual compararon contra otras dos condiciones que incluían una condición control de tratamiento estándar ideada para representar la atención estándar del sobrepeso inducido por dejar de fumar. Una de las hipótesis de los investigadores era que las tasas de abstinencia de tabaco serían mayores con la nueva técnica que las registradas en las otras dos condiciones. De 53 individuos asignados a la nueva condición, 11 dejaron de fumar al final de 52 semanas. Diecinueve de los 54 individuos asignados a la condición de control se abstuvieron hasta el final del mismo periodo. ¿Ofrecen estos datos suficiente evidencia para apoyar, con un nivel de significación de .05, la hipótesis de los investigadores?

ÍNDICE

1 Medias independientes

2 Comparación por parejas

3 La comparación en la proporción

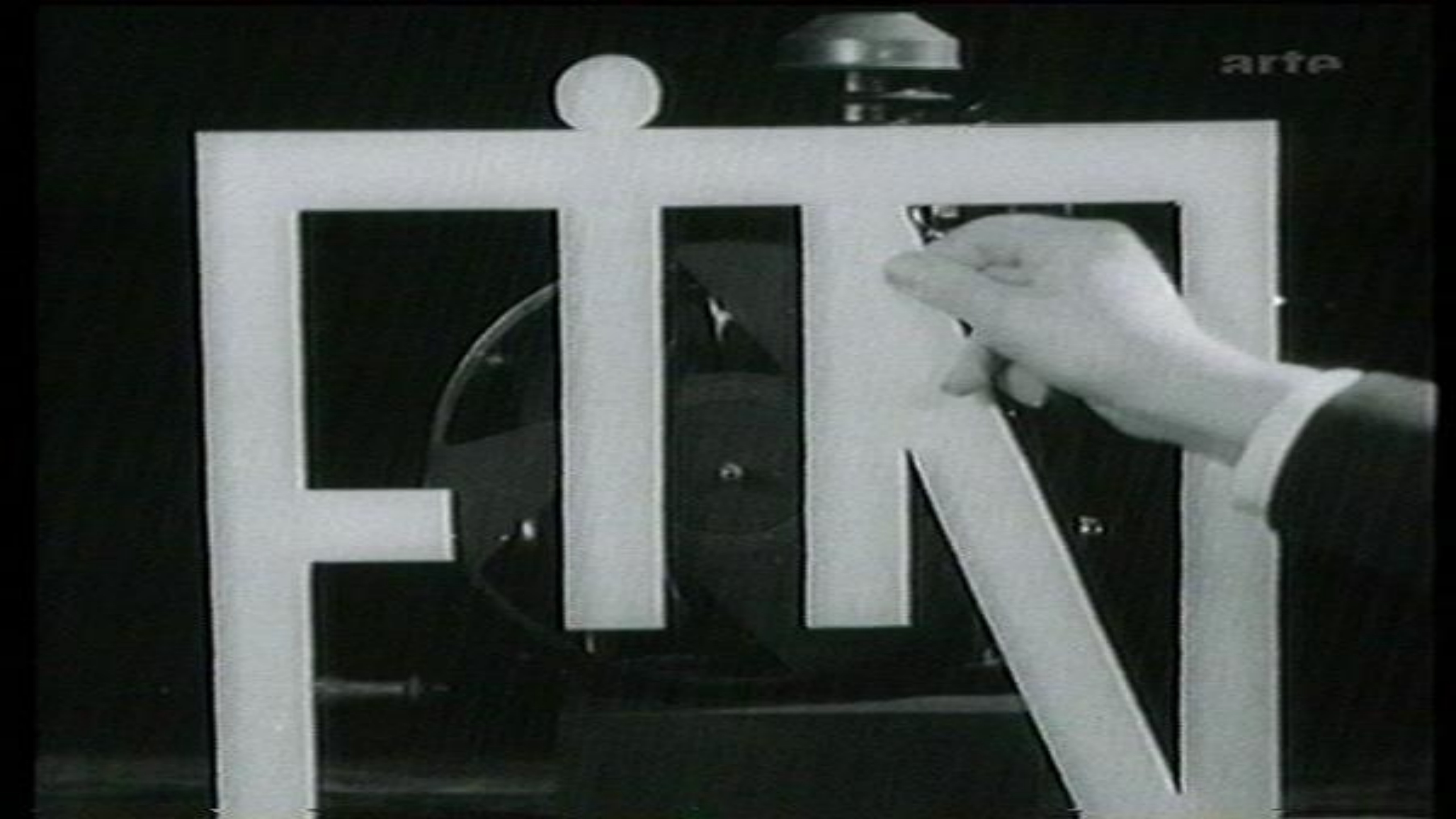
4 Comparación de proporciones

5 Resumen de las pruebas y los estadísticos asociados

Prueba	Estadístico	Distribución de probabilidad

RESEÑAS FINALES

- El presente tema continuo con el tema de prueba de hipótesis.
- Se presentaron las pruebas de diferencias de medias independientes para variancia conocida y desconocida, medias pareadas, diferencia en una proporción y diferencia en las proporciones.
- Interesa saber el caso donde en vez de comparar dos medias o poblaciones, se pueden comprar 3 o más.
- El siguiente tema , Análisis de Variancia, trata sobre la comparación de 2 o más medias.


FIN
arte