Tema 7 : Pruebas de Hipótesis

Contenido

Pruebas que involucran a dos poblaciones, para muestras apareadas y para muestras independientes.

Inferencias acerca de la diferencia entre dos medias poblacionales

Suponga que los empleados de una fábrica usan dos métodos distintos para realizar una determinada tarea. La gerencia desea identificar el método que utilice menos tiempo para realizar esta tarea, de manera tal de maximizar la producción. Sea μ_A la media poblacional del tiempo empleando el método A y μ_B la media poblacional del tiempo requerido para realizar la tarea con el método B.

Puesto que no hay ninguna indicación de cuál sea el mejor método, se empieza por suponer que con los dos métodos se obtiene la misma media poblacional del tiempo requerido para realizar la tarea.

Inferencias acerca de la diferencia entre dos medias poblacionales

Las **hipótesis** son:

H_0: $\mu_A - \mu_B = 0$;

Las medias poblacionales de los tiempos para realizar una tarea con el método A y B son iguales.

H_a: $\mu_A - \mu_B \neq 0$;

Las medias poblacionales de los tiempos para realizar una tarea con el método A y B son diferentes.

Inferencias entre dos medias poblacionales
Muestras apareadas y Muestras Independientes

En la elección del método de muestreo para obtener los datos y probar las hipótesis, se consideran dos diseños alternativos. Uno se basa en muestras independientes y el otro en **muestras por pares** o muestras apareadas.

1. *Diseño de muestras independientes:* se toma una muestra aleatoria simple de trabajadores y cada uno de ellos usa el método A. Se toma otra muestra aleatoria simple de trabajadores res y cada uno de ellos usa el método B.
2. *Diseño de muestras pareadas:* se toma una muestra aleatoria simple de trabajadores. Cada trabajador primero usa uno de los métodos y después usa el otro método. A cada trabajador se le asigna en forma aleatoria el orden en que usará los dos métodos, algunos trabajadores primero usarán el método A y otros el método B.

Pruebas que involucran dos poblaciones

- $H_o: \mu_A = \mu_B$
  $H_a: \mu_A < \mu_B$ Unilateral
- $H_o: \mu_A = \mu_B$
  $H_a: \mu_A > \mu_B$
- $H_o: \mu_A = \mu_B$ Bilateral
  $H_a: \mu_A \neq \mu_B$

$H_o: \mu_d = \mu_A - \mu_B = 0$

OBSERVACIONES APAREADAS

MUESTRAS INDEPENDIENTES

Pruebas que involucran dos poblaciones
Muestras Apareadas

Se deben calcular las Diferencias (Di) de los resultados obtenidos con ambos métodos y probar si el promedio de la diferencia $\overline{D}$ =0. Cada par sería el resultado de un empleado con el método A y el mismo empleado con el método B.

Para esto se procede calculando el promedio y el desvío estándar de la diferencia y luego se calcula Z si n>30 o T en caso contrario.

El criterio de decisión es el cálculo de P-Valor y su comparación con α.

OBSERVACIONES PAREADAS

Par			A-B
1	A	B	D1
2	B	A	D2
3	B	A	D3
4	A	B	D4
5	A	B	D5
6	A	B	D6
7	B	A	D7
8	B	A	D8
9	A	B	D9
...			
n	B	A	Dn

Estadístico de prueba:

$$Z = \frac{\overline{D} - \mu_D}{\frac{\sigma_D}{\sqrt{n}}}$$

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} ; S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n-1}}$$

Pruebas que involucran dos poblaciones

OBSERVACIONES PAREADAS

Par			A-B
1	A	B	D1
2	B	A	D2
3	B	A	D3
4	A	B	D4
5	A	B	D5
6	A	B	D6
7	B	A	D7
8	B	A	D8
9	A	B	D9
…			
n	B	A	Dn

Estadístico de prueba:

$$Z = \frac{\overline{D} - \mu_D}{\dfrac{\sigma_D}{\sqrt{n}}}$$

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n} ; S_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{n-1}}$$

Prueba de Hipótesis para Dos Muestras Pareadas

Ejemplo

Sujeto	Antes	Después
A	6,6	6,8
B	6,5	2,4
C	9,0	7,4
D	10,3	8,5
E	11,3	8,1
F	8,1	6,1
G	6,3	3,4
H	11,6	2,0

Los datos corresponden a 8 individuos seleccionados al azar: mediciones antes y después de la hipnosis en una escala de dolor en centimetros. Se quiere probar que el promedio en la escala de dolor es diferente luego de la hipnosis.

Hipótesis: $\begin{cases} H_0 : \mu_A = \mu_D \\ H_1 : \mu_A \neq \mu_D \end{cases}$

Nivel de significancia: $\alpha = 0{,}05$

Regla de Decisión: Rechazar H_0 si: Valor P < 0,05

Prueba de Hipótesis para Dos Muestras Pareadas

Muestras pareadas

Observación	Grupo 1	Grupo 2	Diferencia
1	X_{11}	X_{21}	$D_1=X_{11}-X_{21}$
2	X_{12}	X_{22}	$D_2=X_{12}-X_{22}$
...	...	...	...
n	X_{1n}	X_{2n}	$D_n=X_{1n}-X_{2n}$

Media

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$$

Desviación Estándar

$$S_D^2 = \frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum_{i=1}^{n} D_i\right)^2}{n}}{n-1}$$

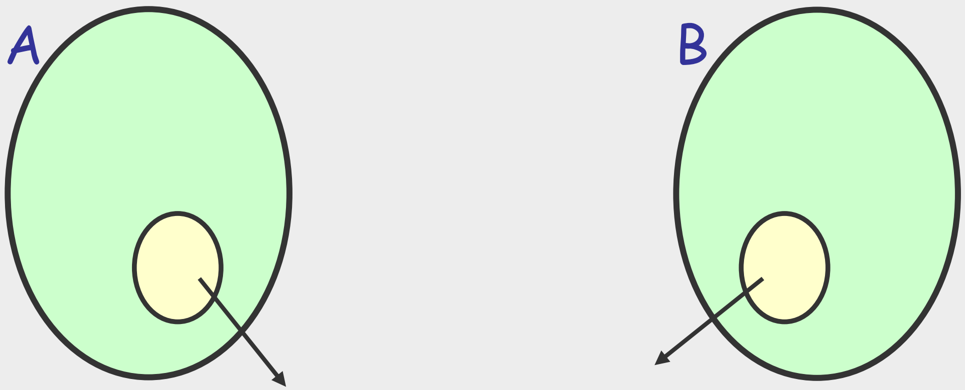
Pruebas de hipótesis para Muestras Apareadas

SUJETO	ANTES	DESPUES	DIFERENCIA
A	6,6	6,5	0,1
B	6,5	2,4	4,1
C	9	7,4	1,6
D	10,3	8,5	1,8
E	11,3	8,1	3,2
F	6,3	3,4	2,9
G	11,6	2	9,6
PROMEDIO			3,3285714
DESVIO (S)			3,0516194
n=7	T=($\bar{X}$-μ)/(S/√n)		2,8858685
	P-Valor (2.886)=	0.0139	

Se concluye que hay diferencia en la escala del dolor luego de la hipnosis.

Pruebas que involucran dos poblaciones
Muestras Independientes

Muestras Independientes:



n_A $\qquad$ n_B

$\bar{x}_A$ $\qquad$ $\bar{x}_B$

S_A^2 $\qquad$ S_B^2

Estadístico de prueba

$$tc = \frac{\bar{x}_D - \mu_D}{S_D} = \frac{\bar{x}_A - \bar{x}_B}{S_a\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

$$S_a^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

Pruebas que involucran dos poblaciones Muestras Independientes

Se deben calcular la media y la varianza de cada muestra, luego se calcula una varianza promedio ponderada S_a^2(luego se toma la raíz cuadrada ya que en el denominador de la T va S_a.

El criterio de decisión es el cálculo de P-Valor y su comparación con α.

Pruebas de diferencias de promedios - Con muestras independientes:

Se desea conocer el comportamiento de dos raciones alimentarias (A y B) para cerdos y determinar cuál produce mayor incremento de pesos. Para ello se alimentaron 2 lotes de animales durante cierto tiempo, determinándose los incrementos de pesos. A continuación se presentan los resultados. Arribar a una conclusión al nivel 5%.

Ración	Número de Cerdos	Incremento Promedio	Variancia
A	$n_A = 15$	$\bar{x}_A = 18{,}2$	$S^2_A = 40{,}4$
B	$n_B = 10$	$\bar{x}_B = 20{,}8$	$S^2_B = 26{,}4$

$$H_0 : \mu_A = \mu_B \rightarrow \mu_A - \mu_B = 0$$
$$H_a : \mu_A \neq \mu_B \rightarrow \mu_A - \mu_B \neq 0$$

$$t_c = \frac{\bar{x}_A - \bar{x}_B}{S_d} = \frac{\bar{x}_A - \bar{x}_B}{S_a\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{18{,}2 - 20{,}8}{5{,}91\sqrt{\frac{1}{15} + \frac{1}{10}}} = \frac{-2{,}6}{2{,}41} = -1{,}078$$

Siendo

$$S_a^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} = \frac{14 * 40,4 + 9 * 26,4}{15 + 10 - 2} = 34,92$$

El criterio de decisión es el cálculo de P-Valor(-1.078)=0.146 comparación con α/2=0.025, se concluye que hay diferencia significativa entre las raciones.

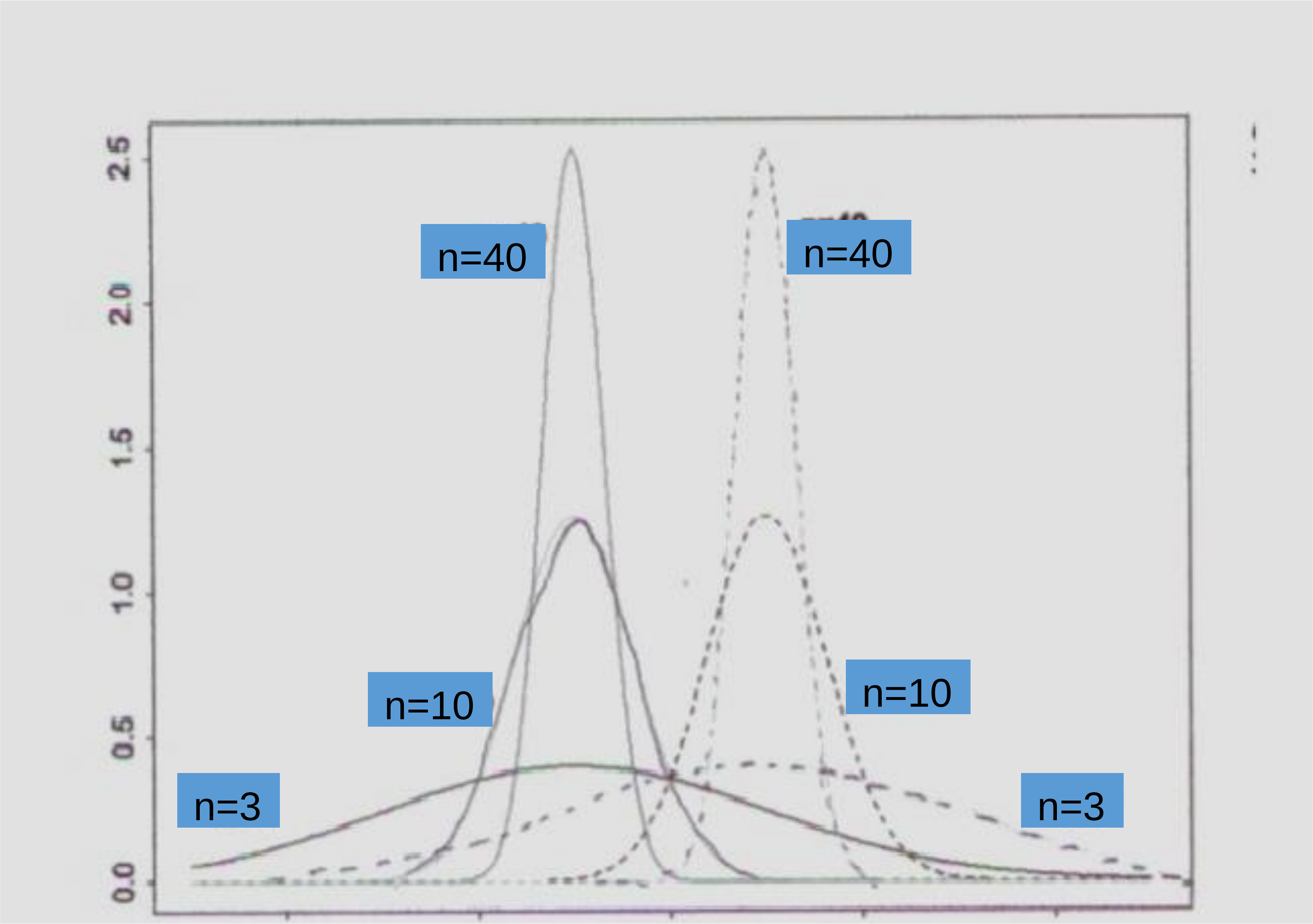
Tipos de Error

Ho

	Verdadera	Falsa
Rechazo la Ho	**Error tipo I ó α**	**Decisión Correcta** **Probabilidad de** **$1 - \alpha$= CONFIANZA**
Acepto la Ho	**Decisión Correcta** **$1 - \beta$ = POTENCIA**	**Error tipo II o β**

α = Error tipo I
β = Error tipo II

¿Qué pasa cuando el tamaño de muestra aumenta?



El área de solapamiento de la distribuciones disminuye, por lo tanto las poblaciones pueden diferenciarse mejor.