Pruebas de hipótesis para dos poblaciones

Ejemplo 5: diferencia de medias

Un gerente aplicó el mismo test de capacitación a 2 grupos. El primero de 50 empleados (n_1) obtuvo una media de 65 pts ($\bar{x}_1$) con una desviación De 10 pts (s_1). El segundo grupo de 40 empleados (n_2) arrojó una media de 62 pts ($\bar{x}_2$) con una desviación estándar de 8 pts (s_2).

¿Existe diferencia significativa entre las medias de los dos grupos a un nivel de significancia del 5% (α)?

Tenemos:

Pob. 1	Pob. 2
$n_1 = 50$	$n_2 = 40$
$\bar{x}_1 = 65$	$\bar{x}_2 = 62$
$s_1 = 10$	$s_2 = 8$
$\alpha = 0.05$	

Planteamiento de hipótesis:

$H_0: \mu_1 - \mu_2 = 0$ ($\mu_1 = \mu_2$ ∴ no hay dif.)

$H_1: \mu_1 - \mu_2 \neq 0$ ($\mu_1 \neq \mu_2$ ∴ hay dif.)

Tipo de prueba: Prueba Z de dos colas ($n_1 \geq 30 \wedge n_2 \geq 30$)

Se asumirá $\sigma_1 = s_1$, $\sigma_2 = s_2$

$\frac{\alpha}{2} = 0.025$ ← RR | RA | RR → $\frac{\alpha}{2} = 0.025$

$Z_{c_1} = -1.95966$ $Z_{c_2} = 1.95966$

$Z_{c_1} = Z_{0.025} = -1.95966$

$Z_{c_2} = Z_{0.975} = 1.95966$

Estandarizar el Z obs:

$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\bar{x}_1, \bar{x}_2$ **se distribuye normalmente,** $n_1, n_2 \geq 30$ **o** 40, **donde** σ_1, σ_2 **se conocen.**

$z_\alpha = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Se parte de un $H_o: \mu_1 - \mu_2 = d_0$ $n_1 \geq 30; n_2 \geq 30$ $d_0 = 0$

$$Z_{obs} = \frac{65 - 62}{\sqrt{\frac{10^2}{50} + \frac{8^2}{40}}} \simeq 1.5811$$

Como:

$Z_{c_1} = -1.95996 < Z_{obs} = 1.5811 < Z_{c_2} = 1.95996$

se cumple que $Z_{obs} \in RA$

∴ Se puede asumir que ambos grupos son similares en cuanto al rendimiento en el test aplicado.

NOTA: La diferencia observada en ambos casos promedios muestrales, al no poder rechazar Ho, se debe entonces al azar o al tipo de muestreo utilizado, pero, no se podría asumir que se deba a que realmente exista una diferencia significativa entre los rendimiento de ambos grupos en el test aplicado.

Usando valor P:

$\text{Valor}P = P(z < -Z_{obs}) + P(z > Z_{obs})$

$= 2 \cdot P(z > z_{obs})$

$= 2 \cdot P(z > 1.5811)$

$= 2 \cdot (0.05693) = 0.11386$

Como $\text{Valor}P = 0.11386 > \alpha = 0.05$

No se encuentra suficiente evidencia en contra para rechazar H_0.

Ejemplo 6: Diferencia de proporciones

Cierta empresa, toma una muestra al azar de 200 clavos (n_A) fabricados por la máquina A, de los cuales 19 resultaron con defectos (éxitos). Igualmente se toma una muestra de 100 (n_B) clavos fabricados por la máquina B y arrojó 5 clavos defectuosos (éxito). ¿El fabricante podrá afirmar que la máquina B produce clavos de mejor calidad a un nivel de significancia del 1% (α)?

Tenemos: $\alpha = 0.01$

Pob. A	Pob. B
$n_A = 200$	$n_B = 100$
$\hat{p}_A = \frac{19}{200} = 0.095$	$\hat{p}_B = \frac{5}{100} = 0.05$

Planteamiento de hipótesis:

Afirmación: Máquina B produce clavos de mejor calidad:

$P_A > P_B \iff P_A - P_B > 0$

$H_0: P_A - P_B = 0$ ó $P_A = P_{(B)}$

$H_1: P_A - P_B > 0$ ó $P_A > P_B$

Tipo de prueba: Prueba Z de cola derecha

$\hat{P}_A \cdot n_A = 19 \geq 5$ $\quad P_B \cdot n_B = 5 \geq 5$

RA | RR → $\alpha = 0.01$

$Z_c = 2.32635$

$Z_c = Z_{0.99} = 2.32635$

Entonces:

$RA =]-\infty, 2.32635[$

$RR =]2.32635, +\infty[$

Estandarizamos el P obs:

Estadístico de contraste o crítico	¿Cuándo?
$Z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, con $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$	La muestra es grande. $n_i p_i > 5, i = 1, 2$

$$\hat{p} = \frac{200 \cdot (0.095) + 100 \cdot (0.05)}{300} = 0.08$$

Luego:

$$Z_{obs} = \frac{0.095 - 0.05}{\sqrt{(0.08)(0.92)\left(\frac{1}{200} + \frac{1}{100}\right)}} = 1.3543$$

Como $Z_{obs} = 1.3543 < Z_c = 2.32635$

Se cumple que $Z_{obs} \in RA$

Por lo tanto, no se cuenta con suficiente evidencia en contra para rechazar Ho, de esta manera se puede asumir que los clavos producidos por ambas máquinas son se similar calidad

Ejemplo 10: comparación de varianzas

Se obtuvieron las estaturas de 21 mujeres (n_M) y 31 hombres (n_H) seleccionados aleatoriamente de una población de alumnos de cierta escuela. En la siguiente tabla se resumen los datos encontrados.

Sean σ_m^2 la varianza de las estaturas de las mujeres y σ_h^2 la varianza de las estaturas de los hombres. Asumiendo que las estaturas para ambos grupos se distribuyen de forma normal, determine, mediante una prueba de hipótesis y con nivel de significancia del 1% (α), si se debe suponer que $\sigma_m^2 = \sigma_h^2$

	Nº Alumnos	Media	Desv.Std	gl
Mujeres	21	63,8	2,18	20
Hombres	31	69,8	1,92	30

Tenemos: $\alpha = 0.01$

Planteamiento de hipótesis:

$H_0: \frac{\sigma_m^2}{\sigma_h^2} = 1$ ($r_0 = 1$) ó $\sigma_m^2 = \sigma_h^2$

$H_1: \frac{\sigma_m^2}{\sigma_h^2} \neq 1$ ó $\sigma_m^2 \neq \sigma_h^2$

Tipo de prueba: Dos colas de distribución F.

$\frac{\alpha}{2} = 0.005$ → RR | RA | RR → $\frac{\alpha}{2} = 0.005$

f_{c_1} $\quad f_{c_2}$

$f_{c_1} = f_{0.005, 20, 30} = 0.32016$

$f_{c_2} = f_{0.995, 20, 30} = 2.82304$

∴ $RA =]0.32016, 2.82304[$

$RR =]0, 0.32016[\cup]2.82304, +\infty[$

Estandarizar el valor observado a términos de distribución F:

$F = \frac{s_1^2}{r_0 s_2^2}$; $v_1 = n_1 - 1$; $v_2 = n_2 - 1$	Aplica $H_o: \frac{\sigma_1^2}{\sigma_2^2} = r_o$

Como $r_0 = 1$, entonces:

$$f_{obs} = \frac{S_m^2}{S_h^2 \cdot r_0} = \frac{(2.18)^2}{(1.92)^2 \cdot 1} = 1.2892$$

Como:

$f_{c_1} = 0.32016 < f_{obs} = 1.2892 < f_{c_2} = 2.82304$

se concluye que $f_{obs} \in RA$

Por lo tanto, no se encontró suficiente evidencia en contra para rechazar Ho, por lo que es posible asumir homocedasticidad de varianzas

Ejemplo 11: comparación de varianzas

Un vendedor de accesorios de computadoras afirma que los discos duros Modelo A tienen una vida útil promedio que supera a la vida útil promedio del Modelo B en más de un año. Para investigar se obtuvo la siguiente información:

Disco duro	n	$\bar{x}$	s
Modelo A	15	6,5	2,3
Modelo B	17	5,4	1,6

Suponga que la vida útil de los ambos modelos se comportan de forma normal.

1. Prueba la hipótesis de que las varianzas de las duraciones de ambos modelos son iguales, a un nivel de significancia del 5%.
2. ¿Es aceptable la información del vendedor?

Sugerencia; hacer una prueba de diferencia de medias, pero, como las muestras son pequeñas, se requiere distribución T, la cuál tiene dos fórmulas distintas para estandarizar que dependen de si las varianzas son iguales o diferentes, por tanto, antes de hacer la prueba de hipótesis para la diferencia de medias, se debe hacer primero una prueba de hipótesis para las varianzas que permita concluir si se pueden asumir como iguales o distintas.

CASO 2:	$T_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$ **Con:** $v = n_1 + n_2 - 2\ gl$	Poblaciones **normales o** $n_1, n_2 \geq 30$ **o** 40, σ_1, σ_2 **no se conocen**, pero se asumen $\sigma_1 = \sigma_2$ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
CASO 3:	$T_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ **Donde:** $v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$	Poblaciones **normales** o $n_1, n_2 \geq 30$ **o** 40 σ_1, σ_2 **NO se conocen y se asume** $\sigma_1 \neq \sigma_2$