CARRERA DE INGENIERÍA DE SISTEMAS

Prueba de Hipótesis

Comparación de dos poblaciones

ESTADÍSTICA APLICADA
PROF. MSC. VILMA S. ROMERO ROMERO
5TO CICLO | 06-10-2022

UNIVERSIDAD DE LIMA

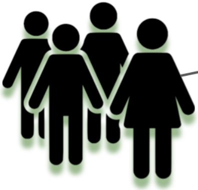
Introducción

- Expandir la idea de prueba de hipótesis a dos muestras.
  - Seleccionar muestras aleatorias de dos poblaciones diferentes para evaluar si las medias o proporciones poblaciones son iguales
- Plantear y evaluar preguntas como:
  - ¿Hay alguna diferencia en el número promedio de defectos realizados en los turnos de día y tarde en los productos de Alicorp?
  - ¿Hay alguna diferencia en el número promedio de días ausentes entre los trabajadores jóvenes (menores de 21 años) y los trabajadores mayores (más de 60 años) de tiendas departamentales?
  - ¿Hay alguna diferencia en la proporción de egresados de la Universidad del Pacífico y egresados de la Universidad ESAN que se titulan durante el primer año de egreso?

UNIVERSIDAD DE LIMA

Muestras independientes

Variable X
en la
Muestra 1



Variable Y
en la
Muestra 2

Ej. Notas de una **muestra de estudiantes de Ingeniería Industrial**

Ej. Notas de una **muestra de estudiantes de Ingeniería de Sistemas**

Muestras pareadas o dependientes

Variable X
en la
Muestra

Variable Y
en la
Muestra

Ej. **Notas del examen parcial** de una muestra de estudiantes de Ingeniería

Ej. **Notas del examen final** de la misma muestra de estudiantes de Ingeniería

Los mismos sujetos o relacionados

UNIVERSIDAD
DE LIMA

Prueba de hipótesis para la razón de varianzas

- Analizar si la variabilidad (desviación estándar o varianza) es similar para dos poblaciones, sabiendo que se distribuyen normalmente.
- Utilizar datos recopilados de **dos muestras aleatorias independientes**, una de la población 1 y otra de la población 2.
- Las dos varianzas muestrales S_1^2 y S_2^2 serán la base para hacer inferencias sobre las dos varianzas poblacionales σ_1^2 y σ_2^2.

Hipótesis	Estadístico de Prueba	Región Crítica	Regla de Decisión Rechazar H_0
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	$F_0 = \frac{S_1^2}{S_2^2}$ $F_0 \sim F_{(n_1-1;n_2-1)}$	$\left\langle 0; F_{(\alpha,n-1,n_2-1)} \right\rangle$	$F_0 < F_{(\alpha,n_1-1,n_2-1)}$
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$		$\left\langle F_{(1-\alpha,n_1-1,n_2-1)}; \infty+ \right\rangle$	$F_0 > F_{(1-\alpha,n_1-1,n_2-1)}$
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$		$\left\langle 0; F_{(\alpha/2,n_1-1,n_2-1)} \right\rangle \cup \left\langle F_{(1-\alpha/2,n_1-1,n_2-1)}; \infty+ \right\rangle$	$F_0 < F_{(\alpha/2,n_1-1,n_2-1)}$ o $F_0 > F_{(1-\alpha/2,n_1-1,n_2-1)}$

Ejemplo 1:

Un determinado estudio concluyó que, en promedio, los adultos pasan más tiempo conectados a internet que los adolescentes. Suponga que se realiza un estudio de seguimiento para el cual se toma una muestra de 26 adultos y 30 adolescentes. Las desviaciones estándar del tiempo que pasan conectados a internet durante un mes fueron 94 minutos y 58 minutos, respectivamente. ¿Los resultados de la muestra respaldan la conclusión de que la varianza del tiempo que pasan conectados a internet en el caso de los adultos es mayor que en el caso de los adolescentes? Usar $\alpha = 0.01$.

Solución:

1) Se formulan las hipótesis:

$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$

Donde:

$\sigma_1^2 =$ Varianza del tiempo que los adultos pasan conectados a internet.
$\sigma_2^2 =$ Varianza del tiempo que los adolescentes pasan conectados a internet.

2) $\alpha = 0.01$

3) El estadístico de prueba es: $F_0 = \frac{S_1^2}{S_2^2} \to F_{(26-1,30-1)}$ y los datos tienen una distribución normal.

4) Región crítica:

$$RC = \langle F_{(0.99,25,29)}; \infty+ \rangle = \langle 2.4783; \infty+ \rangle$$

UNIVERSIDAD DE LIMA

5) Valor del estadístico de prueba es:

$$F_0 = \frac{94^2}{58^2} = 2.627$$

6) **Decisión y conclusión:** Como $F_0 = 2.627 \in RC$; por lo tanto, se rechaza la hipótesis nula. Existe evidencia que sugiere que la variabilidad del tiempo que los adultos permanecen conectados a internet es superior a la de los adolescentes, al 1% de significación.

Prueba de hipótesis para la diferencia de dos medias

UNIVERSIDAD DE LIMA

Varianzas conocidas y muestras independientes

- Las dos poblaciones se distribuyen normalmente.
- Las **dos muestras son independientes**, es decir, no deben estar relacionadas.
- Las varianzas poblacionales deben ser conocidas.

Hipótesis	Estadístico de Prueba	Región Crítica	Regla de Decisión Rechazar H_0
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$	$Z_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\langle -\infty; Z_\alpha \rangle$	$Z_0 < Z_\alpha$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$		$\langle Z_{1-\alpha}; \infty + \rangle$	$Z_0 > Z_{1-\alpha}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$Z_0 \sim N(0,1)$	$\left\langle -\infty; Z_{\alpha/2} \right\rangle \cup \left\langle Z_{1-\alpha/2}; \infty + \right\rangle$	$Z_0 < Z_{\alpha/2}$ o $Z_0 > Z_{1-\alpha/2}$

Ejemplo 2:

El procedimiento Fast Lane está diseñado para reducir el tiempo que un cliente pasa en la línea de pago. Dicho sistema se instaló recientemente en un supermercado. A la gerente de la tienda le gustaría saber si el tiempo promedio de pago usando el método de pago estándar es más largo que usando Fast Lane. Para esto, ella reunió información muestral resumida en la tabla presentada. El tiempo se mide desde que el cliente ingresa a la línea hasta que sus bolsas están en el carrito. Es decir, el tiempo incluye tanto la espera en la cola como la salida. Utilizar un nivel de significación del 1%.

Tipo de cliente	$\bar{x}$	σ	**Tamaño de muestra**
Estándar	5.50 minutos	0.40 minutos	50
Fast Lane	5.30 minutos	0.30 minutos	100

Solución:

1) Se formulan las hipótesis:

$H_0 : \mu_s = \mu_f$
$H_1 : \mu_s > \mu_f$

Donde:

$\mu_s =$ Tiempo promedio de pago usando el método de pago estándar.
$\mu_f =$ Tiempo promedio de pago usando el método de pago Fast Lane.

2) $\alpha = 0.01$

3) El estadístico de prueba es: $Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

4) Región crítica:

$$RC = \langle Z_{0.99}; \infty+\rangle = \langle 2.33; \infty+\rangle$$

UNIVERSIDAD DE LIMA

5) Valor del estadístico de prueba es:

$$Z_0 = \frac{5.5 - 5.3}{\sqrt{\frac{0.40^2}{50} + \frac{0.30^2}{100}}} = 3.12$$

6) **Decisión y conclusión:** Como $Z_0 = 3.13 \in RC$; por lo tanto, se rechaza la hipótesis nula. Entonces, hay suficiente evidencia para decir que H_1 es verdadera, al 1% de significación. Esto indica que el método Fast Lane es más rápido.

Varianzas desconocidas y muestras independientes

- **Varianzas desconocidas, pero iguales** $(\sigma = \sigma_1^2 = \sigma_2^2)$

Hipótesis	Estadístico de Prueba	Región Crítica	Regla de Decisión Rechazar H_0
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$	$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $t_0 \sim t_{(n_1+n_2-2)}$	$\left\langle -\infty; t_{(\alpha, n_1+n_2-2)} \right\rangle$	$t_0 < t_{(\alpha, n_1+n_2-2)}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$		$\left\langle t_{(1-\alpha, n_1+n_2-2)}; \infty+ \right\rangle$	$t_0 > t_{(1-\alpha, n_1+n_2-2)}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$		$\left\langle -\infty; t_{(\alpha/2, n_1+n_2-2)} \right\rangle \cup$ $\left\langle t_{(1-\alpha/2, n_1+n_2-2)}; \infty+ \right\rangle$	$t_0 < t_{(\alpha/2, n_1+n_2-2)}$ o $t_0 > t_{(1-\alpha/2, n_1+n_2-2)}$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Varianzas desconocidas y muestras independientes

■ Varianzas desconocidas y diferentes $(\sigma_1^2 \neq \sigma_2^2)$

Hipótesis	Estadístico de Prueba	Región Crítica	Regla de Decisión Rechazar H_0
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$	$t_0 = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $t_0 \sim t_{(\nu)}$	$\left\langle -\infty; t_{(\alpha,\nu)} \right\rangle$	$t_0 < t_{(\alpha,\nu)}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$		$\left\langle t_{(1-\alpha,\nu)}; \infty+ \right\rangle$	$t_0 > t_{(1-\alpha,\nu)}$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$		$\left\langle -\infty; t_{(\alpha/2,\nu)} \right\rangle \cup \left\langle t_{(1-\alpha/2,\nu)}; \infty+ \right\rangle$	$t_0 < t_{(\alpha/2,\nu)}$ o $t_0 > t_{(1-\alpha/2,\nu)}$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Ejemplo 3:

Se desea comparar los promedios de los tiempos, en minutos, que dos máquinas emplean en cierta producción. Para esto, se registran los tiempos de 9 y 8 artículos elegidos al azar de la producción de cada máquina.

Máquina 1	12	28	10	25	24	19	22	33	17
Máquina 2	16	20	16	20	16	17	15	21	

Con un nivel de significancia del 5%, ¿se puede afirmar que los tiempos promedios de las dos máquinas son diferentes?

Solución:

1) Se formulan las hipótesis:
   $H_0 : \mu_1 = \mu_2$ (no existe diferencia en el tiempo promedio que emplean ambas máquinas)
   $H_1 : \mu_1 \neq \mu_2$ (existe diferencia en el tiempo promedio que emplean ambas máquinas)
2) $\alpha = 0.05$
3) Para determinar el estadístico de prueba apropiado, primero se debe evaluar si las varianzas poblaciones son iguales o diferentes.
   a) Las hipótesis que se formulan son: $H_0 : \sigma_1^2 = \sigma_2^2$
      $H_1 : \sigma_1^2 \neq \sigma_2^2$
   b) Valor del estadístico de prueba: $F_0 = \frac{S_1^2}{S_2^2} = \frac{55.11}{5.411} = 10.1848$
   c) Región crítica, considerando $\alpha = 0.05$
      $$RC = \left\langle 0; F_{(0.025,8,7)} \right\rangle \cup \left\langle F_{(0.975,8,7)}; \infty+ \right\rangle = \langle 0; 0.2208 \rangle \cup \langle 4.899; \infty+ \rangle$$
   d) **Decisión:** Se rechaza $H_0 \Rightarrow$ las **varianzas son diferentes**.

4) Región Crítica:

$$RC = \left\langle -\infty; t_{(0.025,9)} \right\rangle \cup \left\langle t_{(0.975,9)}; \infty+ \right\rangle = \langle -\infty; -2.262\rangle \cup \langle 2.262; \infty+\rangle$$

5) Valor del estadístico de prueba es:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{21.11 - 17.63}{\sqrt{\frac{55.11}{9} + \frac{5.411}{8}}} = 1.335$$

6) **Decisión y conclusión:** Como $t_0 = 1.335 \notin RC$; por lo tanto, no se rechaza H_0. Entonces, no hay suficiente evidencia para decir que H_1 es verdadera, al 5% de significación.

Esto indica que no existe diferencia en el tiempo promedio que emplean ambas máquinas en producir el artículo.

Ejemplo 4:

Owens Lawn Care Inc. fabrica y ensambla cortadoras de césped que se envían a distribuidores en todo Estados Unidos y Canadá. Se han propuesto dos procedimientos diferentes para montar el motor en el cortacésped. Para evaluar los dos métodos, se decidió realizar un estudio de tiempo y movimiento. Una muestra de cinco empleados fue cronometrada usando el primer método y seis usando el segundo método. Los resultados, en minutos, se muestran a continuación.

Método 1	2	4	9	3	2	
Método 2	3	7	5	8	4	3

¿Hay alguna diferencia en los tiempos promedio de montaje? Use un nivel de significancia de 0.10.

Solución:

1) Se formulan las hipótesis:
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
2) $\alpha = 0.10$
3) Para determinar el estadístico de prueba apropiado, primero se debe evaluar si las varianzas poblaciones son iguales o diferentes.
   a) Las hipótesis que se formulan son: $H_0 : \sigma_1^2 = \sigma_2^2$
   $H_1 : \sigma_1^2 \neq \sigma_2^2$
   b) Valor del estadístico de prueba: $F_0 = \frac{S_1^2}{S_2^2} = \frac{8.5}{4.4} = 1.932$
   c) Región crítica, considerando $\alpha = 0.10$
   $$RC = \left\langle 0; F_{(0.05,4,5)} \right\rangle \cup \left\langle F_{(0.95,4,5)}; \infty+ \right\rangle = \langle 0; 0.1598 \rangle \cup \langle 5.1922; \infty+ \rangle$$
   d) **Decisión:** No se rechaza $H_0 \Rightarrow$ las **varianzas son iguales**.

UNIVERSIDAD DE LIMA

4) Región Crítica:

$$RC = \left\langle -\infty; t_{(0.05,9)} \right\rangle \cup \left\langle t_{(0.95,9)}; \infty+ \right\rangle = \langle -\infty; -1.833 \rangle \cup \langle 1.833; \infty+ \rangle$$

5) Valor del estadístico de prueba es:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{4 - 5}{\sqrt{6.2222(\frac{1}{5} + \frac{1}{6})}} = -0.662$$

Donde: $s_p^2 = \frac{(5-1)8.5+(6-1)4.4}{5+6-2} = \frac{56}{9} = 6.2222$

6) **Decisión y conclusión:** Como $t_0 = -0.662 \notin RC$; por lo tanto, no se rechaza H_0. Entonces, no hay suficiente evidencia para decir que H_1 es verdadera, al 10% de significación.

Esto indica que no hay diferencia en los tiempos promedios para el montaje del motor utilizando los dos métodos.

Muestras pareadas o dependientes

- Comparar el comportamiento de una variable o factor en dos instantes y/o contextos, con el objetivo de evaluar los cambios ocurridos.
- Generalmente, las muestras de datos pareados son pequeñas $\Rightarrow$ la distribución t es la distribución de referencia para la prueba.

Hipótesis	**Estadístico de Prueba**	**Región Crítica**	**Regla de Decisión** **Rechazar H_0**
$H_0 : \mu_d = 0$ $H_1 : \mu_d < 0$	$t_0 = \dfrac{\bar{d}}{S_d/\sqrt{n}}$ $t_0 \sim t_{(n-1)}$	$\left\langle -\infty; t_{(\alpha,n-1)} \right\rangle$	$t_0 < t_{(\alpha,n-1)}$
$H_0 : \mu_d = 0$ $H_1 : \mu_d > 0$		$\left\langle t_{(1-\alpha,n-1)}; \infty+ \right\rangle$	$t_0 > t_{(1-\alpha,n-1)}$
$H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$		$\left\langle -\infty; t_{(\alpha/2,n-1)} \right\rangle \cup \left\langle t_{(1-\alpha/2,n-1)}; \infty+ \right\rangle$	$t_0 < t_{(\alpha/2,n-1)}$ o $t_0 > t_{(1-\alpha/2,n-1)}$
$\bar{d}$ es el promedio de la diferencia entre las observaciones pareadas o relacionadas. s_d es la desviación estándar de las diferencias entre las observaciones pareadas o relacionadas.			

- **Supuesto:** La población de las diferencias sigue una distribución normal.

Ejemplo 5:

Algunos investigadores afirman que la música puede ser relajante y, por lo tanto, reducir el estrés. Doce pacientes que afirman estar sufriendo estrés relacionado con el trabajo fueron seleccionados al azar. Se obtuvo una frecuencia de pulso en reposo inicial (en latidos por minuto, lpm), y cada persona participó en una terapia de relajación de un mes de duración que incluía escuchar música. Se tomó una frecuencia de pulso en reposo final al final del experimento.

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Inicio	67	71	67	83	70	75	71	68	72	88	78	70
Final	61	72	70	76	58	61	74	59	61	64	71	77

¿Existe alguna evidencia que sugiera que la terapia de relajación que incluía escuchar música redujo la frecuencia del pulso promedio y, por lo tanto, el nivel de estrés? Suponer que las distribuciones de la frecuencias del pulso inicial y final son normales, y usar un $\alpha = 0.05$.

Solución:

De los datos tenemos:

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Inicio	67	71	67	83	70	75	71	68	72	88	78	70
Final	61	72	70	76	58	61	74	59	61	64	71	77
Diferencia	6	-1	-3	7	12	14	-3	9	11	24	7	-7

$$\mu_d = \mu_{Inicio} - \mu_{Final} \quad , \quad \bar{d} = 6.333 \quad , \quad s_d = 8.732$$

1) Se formulan las hipótesis:
   $H_0 : \mu_d = 0$ (El programa no tiene efecto)
   $H_1 : \mu_d > 0$ (El programa está diseñado para reducir el estrés)

2) $\alpha = 0.05$

3) El estadístico de prueba es: $t_0 = \frac{\bar{d}}{S_d/\sqrt{n}} \sim t_{(n-1)}$

UNIVERSIDAD DE LIMA

4) Región Crítica:

$$RC = \left\langle t_{(0.95,11)}; \infty+ \right\rangle = \langle 1.796; \infty+ \rangle$$

5) Valor del estadístico de prueba es:

$$t_0 = \frac{6.333}{\frac{8.732}{\sqrt{12}}} = 2.513$$

6) **Decisión y conclusión:** Como $t_0 = 2.513 \in RC$; por lo tanto, se rechaza H_0. Entonces, hay suficiente evidencia para decir que H_1 es verdadera, al 5% de significación.

Hay evidencia que sugiere que el programa de terapia de relajación que incluye escuchar música reduce la frecuencia del pulso en reposo de una persona (por ende, el nivel de estrés).

UNIVERSIDAD DE LIMA

PH para la diferencia de dos proporciones

- **Comparar dos proporciones poblacionales.**
- Una agencia de publicidad podría estar interesada en comparar las verdaderas proporciones de niños que vieron un determinado comercial de televisión en dos regiones diferentes del país.

Hipótesis	**Estadístico de Prueba**	**Región Crítica**	**Regla de Decisión** **Rechazar H_0**
$H_0: p_1 = p_2$ $H_1: p_1 < p_2$	$Z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $Z_0 \dot{\sim} N(0,1)$	$\langle -\infty; Z_\alpha \rangle$	$Z_0 < Z_\alpha$
$H_0: p_1 = p_2$ $H_1: p_1 > p_2$		$\langle Z_{1-\alpha}; \infty+ \rangle$	$Z_0 > Z_{1-\alpha}$
$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$		$\langle -\infty; Z_{\alpha/2} \rangle \cup \langle Z_{1-\alpha/2}; \infty+ \rangle$	$Z_0 < Z_{\alpha/2}$ o $Z_0 > Z_{1-\alpha/2}$

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$$

UNIVERSIDAD DE LIMA

Ejemplo 6:

El Departamento de Ventas de una compañía de perfumes está particularmente interesado en saber si hay una diferencia en las proporciones de mujeres jóvenes y mayores que comprarían la nueva fragancia "Spring" si se comercializara. Una muestra aleatoria de 100 mujeres jóvenes reveló que a 19 les gustó la fragancia "Spring" lo suficientemente bien como para comprarla. Del mismo modo, una muestra de 200 mujeres mayores reveló que a 62 les gustó la fragancia lo suficientemente bien como para hacer una compra.

¿Existe evidencia que permita afirmar que no existe diferencia en la proporción de mujeres que compraría la nueva fragancia "Spring" entre ambos grupos? Usar $\alpha = 0.05$.

UNIVERSIDAD DE LIMA

Solución:

De los datos tenemos:

Mujeres jóvenes	$n_1 = 100$	$x_1 = 19$	$\hat{p}_1 = 19/100 = 0.19$	$\hat{p} = \frac{19+62}{100+200} = 0.27$
Mujeres mayores	$n_2 = 200$	$x_2 = 62$	$\hat{p}_2 = 62/200 = 0.31$	

1) Se formulan las hipótesis:
   $H_0 : p_1 = p_2$ (No existe diferencia en la proporción de mujeres que compraría "Spring")
   $H_1 : p_1 \neq p_2$ (Existe diferencia en la proporción de mujeres que compraría "Spring")

2) $\alpha = 0.05$

3) El estadístico de prueba es: $Z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ y los datos tienen una distribución binomial.

4) Región Crítica:

$$RC = \langle -\infty; Z_{0.025} \rangle \cup \langle Z_{0.975}; \infty+ \rangle = \langle -\infty; -1.96 \rangle \cup \langle 1.96; \infty+ \rangle$$

5) Valor del estadístico de prueba es:

$$Z_0 = \frac{0.19 - 0.31}{\sqrt{0.27(1-0.27)\left(\frac{1}{100}+\frac{1}{200}\right)}} = -2.21$$

6) **Decisión y conclusión:** Como $Z_0 = -2.21 \in RC$; por lo tanto, se rechaza H_0. Entonces, hay suficiente evidencia para decir que H_1 es verdadera, al 5% de significación.

Es decir, hay una diferencia en la proporción de mujeres jóvenes y mayores que comprarían "Spring".

Recursos Adicionales I

Devore, J. (2019). *Introducción a la probabilidad y estadística para ingeniería y ciencias*. Cengage, 1 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

Johnson, R. A. (2012). *Probabilidad y estadística para ingenieros*. Pearson Educación, 8 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.

Mendenhall, W., Beaver, R. J., and Beaver, B. M. (2015). *Introducción a la probabilidad y estadística*. Cengage, 14 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

UNIVERSIDAD DE LIMA

Recursos Adicionales II

Millones, R., Barreno, E., Vásquez, F., and Castillo, C. (2017). *Estadística Descriptiva y Probabilidades: Aplicaciones en la ingeniería y los negocios*. Lima: Fondo Editorial de la Universidad de Lima, 1 edition. Código Biblioteca U.Lima: 519.53 E.

Triola, M. (2018). *Estadística*. Pearson Educación, 12 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.

UNIVERSIDAD DE LIMA