Estadística sin espinas

(6) Contraste de hipótesis

v0.1 24_06_05

Aprende sin espinas
con @carlos_cactus

Sócrates se equivocaba. El conocimiento no es lo único que crece al compartirse:
La alegría también.

A la inspiración del bucle_infinito,
al Cibergrupo y al tHash_A, por su amistad,
y sobre todo, a quienes dicen "pero quiero"
cuando sienten "no puedo".

¡Un saludo sin espinas!
@carlos_cactus :D

Y si quieres saber más:
¡Encuéntrame en Telegram como @carlos_cactus o habla con Espinito, el bot Sin Espinas, en @GestionSinEspinasBot.

Únete a la comunidad de Telegram Sin Espinas y no te pierdas nada!

Deja de preocuparte por aprobar y ¡Aprende sin Espinas!

VI. CONTRASTE DE HIPÓTESIS

6.1. Conceptos previos

6.1.1. HIPÓTESIS

Una hipótesis es una afirmación sobre cierta propiedad de una o más poblaciones.

6.1.2. CONTRASTE DE HIPÓTESIS

El contraste o TEST de hipótesis es un procedimiento mediante el cual se toma una decisión sobre si cierta propiedad de una POBLACIÓN asumida como cierta es COMPATIBLE con lo OBSERVADO en una MUESTRA de esa población.

El contraste de hipótesis permite por tanto INFERIR ciertas propiedades de la POBLACIÓN (cuyas propiedades NO se conocen) a partir de una realización muestral (cuyas propiedades SÍ se conocen).

Esto se hace mediante la formulación de una hipótesis y su posterior contraste para su ACEPTACIÓN o RECHAZO.

6.1.3. TIPOS DE CONTRASTE DE HIPÓTESIS

→ Contrastes no paramétricos: No se tratan en esta materia.

La hipótesis concierne a propiedades de la distribución (dependencia o independencia de sucesos, homogeneidad).

→ Contrastes paramétricos: Sí se tratan en esta materia.

Si la hipótesis concierne a parámetros poblacionales (como la media o la varianza). Se destacan:

SÍ SE TRATAN en este curso

- Contrastes de hipótesis de 1 población
    - Contraste para la media μ:
        - **Con varianza σ^2 conocida**
        - **Con varianza σ^2 DESCONOCIDA**
    - **Contraste para la proporción p.**

- Contrastes de hipótesis de 2 poblaciones
    - Contraste para la diferencia de medias μ_1-μ_2
        - **Con varianzas σ^2 conocidas**
        - **Con varianzas σ^2 DESCONOCIDAS E IGUALES $\sigma_1^2 = \sigma_2^2$**

    - **Contraste para datos apareados**

NO SE TRATAN en este curso

- Contrastes de hipótesis de 1 población
  - Contraste para la varianza σ^2
    - Con media μ conocida
    - Con media μ DESCONOCIDA

- Contrastes de hipótesis de 2 poblaciones
  - Contraste para la diferencia de medias μ_1-μ_2 con σ^2 DESCONOCIDAS Y DISTINTAS $\sigma_1^2 \neq \sigma_2^2$
  - Contraste para la diferencia de proporciones p_1-p_2

6.1.4. HIPÓTESIS NULA E HIPÓTESIS ALTERNATIVA

Un contraste de hipótesis se resuelve mediante la aceptación de una afirmación o bien su refutación.

→ Se llama hipótesis NULA a la hipótesis que se desea verificar. Se denota como H_0.

→ Se llama hipótesis ALTERNATIVA a la negación de la hipótesis nula. Se denota por H_1.

Aceptar H_0 comporta descartar H_1. Es decir, son MUTUAMENTE EXCLUYENTES.

Mediante el contraste de hipótesis, la validez de las hipótesis nula y alternativa se verifican en un conjunto dado.

Las hipótesis se formulan en función de parámetros de la distribución (como condición de la media, de la proporción, de la diferencia de medias…)

6.1.5. REGIÓN DE ACEPTACIÓN Y REGIÓN DE RECHAZO

El conjunto probabilístico en el que habitan las hipótesis nula y alternativa, como parámetros, está dividido en 2 REGIONES:

→ REGIÓN DE ACEPTACIÓN (de la hipótesis nula):

Conjunto de realizaciones muestrales para las cuales NO SE RECHAZA H_0 (y SÍ SE RECHAZA H_1).

→ REGIÓN DE RECHAZO (de la hipótesis nula o región CRÍTICA):

Conjunto de realizaciones muestrales para las cuales se RECHAZA H_0 (y se ACEPTA H_1).

6.1.6. NIVEL DE SIGNIFICACIÓN EN UN CONTRASTE DE HIPÓTESIS

El nivel de significación α en un contraste de hipótesis coincide con el error máximo que se fija como tolerable.
En concreto, es el máximo error admisible de tipo I, es decir, que se comete si se RECHACE la hipótesis nula aun siendo CIERTA.

El nivel de significación admisible en cada contraste depende de la situación y no es trivial elegir qué es asumible y qué no.

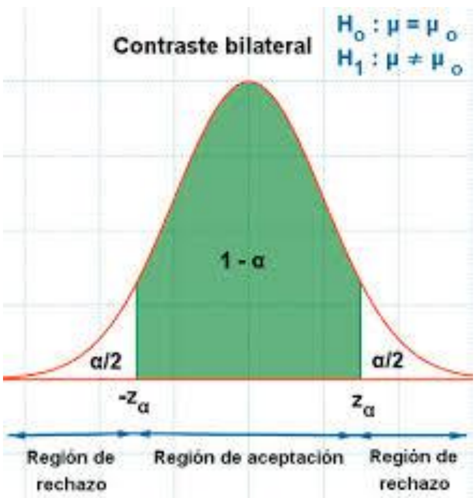
6.2. Tipos de contrastes paramétricos

Según la formulación de las hipótesis nula y alternativa, para un parámetro θ contrastado (siendo α el nivel de significación y $1-\alpha$ el nivel de confianza), se distinguen 3 tipos de contraste paramétrico:

a) Contraste bilateral (con la región de rechazo a AMBOS lados):

$$\begin{cases} H_0: & \theta = \theta_0 \\ H_1: & \theta \neq \theta_0 \end{cases}$$

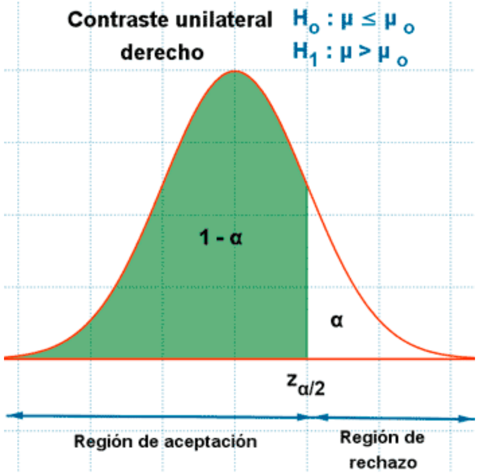
Es decir, que el parámetro adopte un valor dado o que por contrario no lo haga.



b) Contraste unilateral DERECHO (con la región de rechazo a la DERECHA):

$$\begin{cases} H_0: & \theta & \leq & \theta_0 \\ H_1: & \theta & > & \theta_0 \end{cases}$$

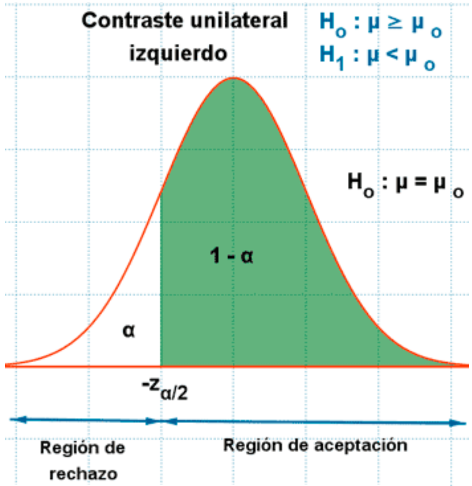
Es decir, que el parámetro adopte un valor MENOR O IGUAL que otro dado o que, por contrario, no lo haga.



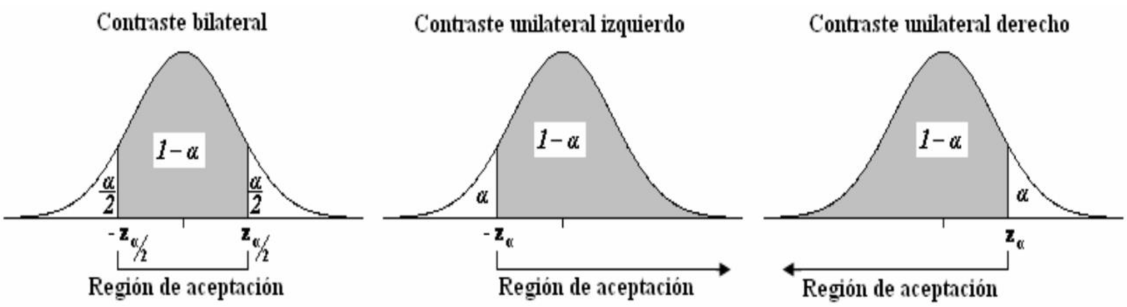
c) Contraste unilateral IZQUIERDO (por la región de rechazo a la IZQUIERDA):

$$\begin{cases} H_0: & \theta & \geq & \theta_0 \\ H_1: & \theta & < & \theta_0 \end{cases}$$

Es decir, que el parámetro adopte un valor MAYOR O IGUAL que otro dado o que, por contrario, no lo haga.



Se tiene:



Nótese que en el caso del contraste bilateral la probabilidad α/2 se distribuye en 2 colas, una a cada lado de la región de aceptación, mientras que, en los contrastes unilaterales, toda la probabilidad de la región de rechazo está concentrada en 1 sola cola.

Nótese que el NOMBRE del tipo de contraste (bilateral, izquierdo o derecho) hace referencia a la REGIÓN DE RECHAZO.

6.3. Estadístico de contraste

Un estadístico de contraste es una FUNCIÓN de la muestra (depende de alguno de sus parámetros) cuya distribución es conocida, asumiendo cierta la hipótesis nula.

El estadístico de contraste sirve para obtener VALOR CRÍTICO, que después se compara con el valor que adopta distribución de la muestra observada.

Esa comparación permite decidir si se RECHAZA o NO SE RECHAZA la hipótesis nula.

En cada contraste, se aplica un estadístico concreto que depende del comportamiento de la distribución observada, así como la condición de rechazo o no rechazo de H_0.

<table>
<tr><th># Poblaciones</th><th colspan="2">Parámetro contrastado</th><th>Estadístico de contraste</th></tr>
<tr><td rowspan="3">Contrastes de 1 población</td><td rowspan="2">Media μ</td><td>Con varianza σ^2 conocida</td><td>$z=\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1)$</td></tr>
<tr><td>Con varianza σ^2 DESCONOCIDA</td><td>$t=\frac{\bar{x}-\mu_0}{\frac{\hat{s}}{\sqrt{n}}}\sim t_{n-1}$</td></tr>
<tr><td colspan="2">Proporción p</td><td>$z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\sim N(0,1)$</td></tr>
<tr><td rowspan="3">Contrastes de 2 poblaciones</td><td rowspan="2">Diferencia de medias $\mu_1-\mu_2$</td><td>Con varianzas σ^2 conocidas</td><td>$z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$</td></tr>
<tr><td>Con varianzas σ^2 DESCONOCIDAS E IGUALES $\sigma_1^2=\sigma_2^2$</td><td>$t=\frac{\bar{x}_1-\bar{x}_2}{s_t\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}$</td></tr>
<tr><td colspan="2">Datos aparejados con diferencia d</td><td>$t=\frac{\bar{d}}{\frac{s_{\bar{d}}}{\sqrt{n}}}\sim t_{n-1}$</td></tr>
</table>

6.4. p-valor o nivel crítico

6.4.1. CONCEPTO DE P-VALOR

De igual modo que el área que se encuentra más allá del valor crítico corresponde con el nivel de significación α, el p-valor es el área excluida por el estadístico de contraste más allá del centro de la distribución, como cola de probabilidad.

Es decir, el p-valor es al estadístico de constraste lo que el nivel de significación al valor crítico.

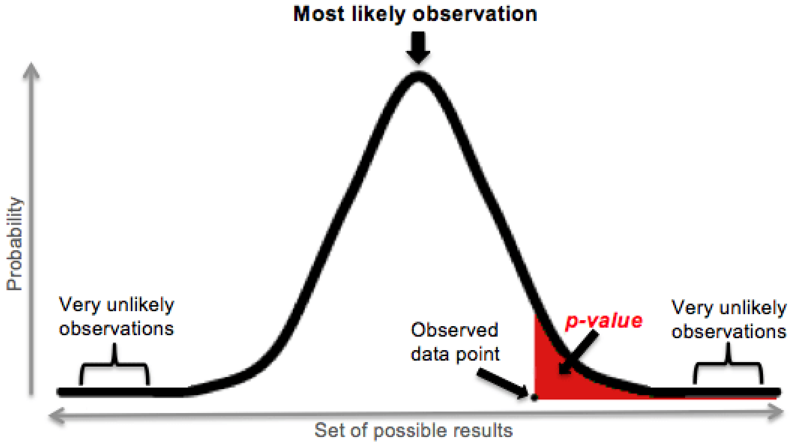
El p-valor permite resolver contrastes de hipótesis ofreciendo un método alternativo al valor crítico.

En lugar de comparar un estadístico de contraste con un valor crítico que fija la frontera entre las regiones de aceptación y de rechazo, la resolución mediante p-valor se basa en la comparación de 2 áreas de probabilidad:

- El p-valor: área excluida más allá del estadístico de contraste
- El nivel de significación α:áres más allá del valor crítico.

El p-valor corresponde con la PROBABILIDAD de observar valores TAN probables o MENOS probables que los que devuelve el estadístico de contraste a partir de la muestra (es decir, más alejados del centro 0 en una distribución simétrica como la t o la Z).

Formalmente, se define como la probabilidad de observar valores tan alejados o más alejados de la media como los que se observan, asumiendo que la hipótesis nula es CIERTA.



A *p-value* (shaded red area) is the probability of an observed (or more extreme) result arising by chance

Por tanto, es un estimador de cuán probable es que el valor observado se deba a la CASUALIDAD (y no responda a una desviación significativa):

- Cuanto más próximo sea el p-valor a 1 (mayor área roja) MAYOR CONFIANZA se tiene al ACEPTAR H_0, ya que es MÁS PROBABLE que la muestra observada se deba a la CASUALIDAD.

- Cuanto más próximo sea el p-valor a 0 (menor área roja), MAYOR CONFIANZA se tiene al RECHAZAR H_0, ya que es MENOS PROBABLE que la desviación observada se deba a la CASUALIDAD.

Se puede entender como el nivel de significación MÍNIMO que permite RECHAZAR H_0.

6.4.2. CÁLCULO DEL P-VALOR

El método usado para el cálculo del p-valor depende del tipo de contraste:

- Para contrastes BILATERALES:

$$p - valor = P\left(\frac{X > |x_0|}{H_0}\right) + P\left(\frac{X < -|x_0|}{H_0}\right) = P(\mathrm{Z} > |\mathrm{z}|) + P(\mathrm{Z} < -|\mathrm{z}|)$$

Lo cual, por simetría y complementariedad es:

$$p - valor = 2 \cdot \big(1 - P(Z < z)\big)$$

- Para contrastes UNILATERALES por la DERECHA:

$$p - valor = P\left(\frac{X \geq x_0}{H_0}\right) = P(Z > z_0) = 1 - \mathrm{P}(\mathrm{Z} < \mathrm{z}_0)$$

Nótese que en este caso siempre se tendrá $z_0 > 0$ y, por tanto, se podrá consultar en la tabla $\mathrm{P}(\mathrm{Z} < \mathrm{z}_0)$ y calcularlo por complementariedad.

- Para contrastes UNILATERALES por la IZQUIERDA:

$$p - valor = P\left(\frac{X \leq x_0}{H_0}\right) = P(Z \leq z_0)$$

Nótese que en este caso siempre se tendrá $z_0 < 0$, pero por simetría, se puede usar el valor en positivo.

Donde:
z es el estadístico de contraste
z_0 es el valor crítico.

Es decir, el p-valor es la probabilidad con que el estadístico de contraste adopta el valor observado o uno más alejado del centro asumiéndose cierta H_0 y que no hay diferencia conceptual en el cálculo de los contrastes unilaterales.

6.4.3. INTERPRETACIÓN DEL P-VALOR Y TOMA DE DECISIÓN

El p-valor corresponde con el área excluida más allá del estadístico de contraste observado, es decir, la probabilidad de observar valores más alejados del centro.
El p-valor se compara con el NIVEL DE SIGNIFICACIÓN para resolver el contraste.

Ambas regiones de probabilidad, el nivel de significación α y el p-valor, aumentan desde los extremos de la distribución hacia el centro.
La relación entre las áreas que representan α y o-valor definen si H0 se puede rechazar.

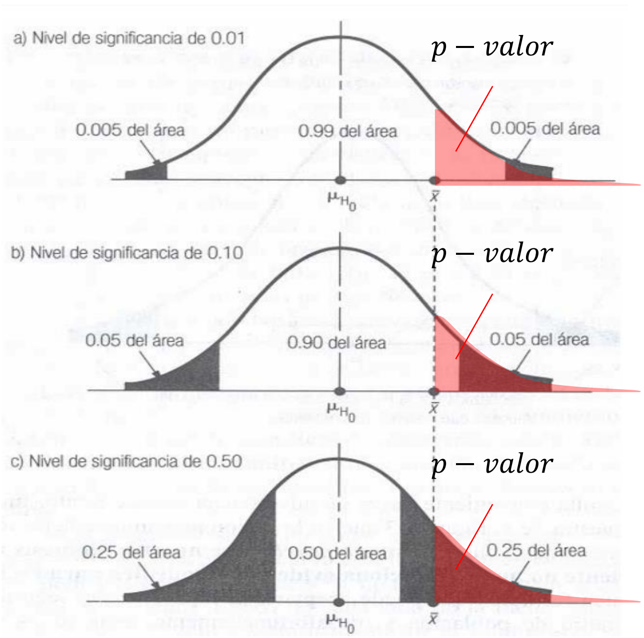


Figura a) $p_{valor} >> \alpha$

H0 NO se rechaza.
La probabilidad de encontrar valores como el observado o más alejados, EXCEDE (es mayor) al nivel de confianza.

El estadístico observado está en la región de confianza, alejado del valor crítico que corresponde a la abscisa que excluye el área α.

Figura b) $p_{valor} > \alpha$

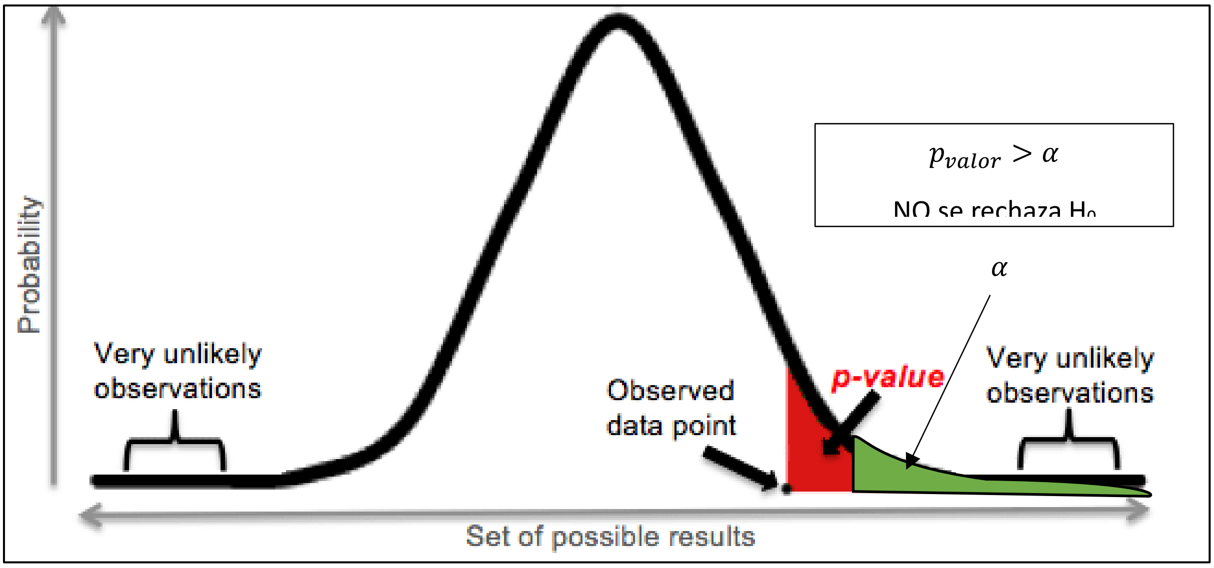
H0 NO se rechaza.
A medida que el nivel de signficiación α crece, el estadístico observado es más próximo a la región de rechazo.

Figura c) $p_{valor} < \alpha$

H0 SÍ se rechaza.
En el momento en que el área que corresponde con el p-valor queda incluida por el área α el estadístico observado queda incluido en la región de rechazo.

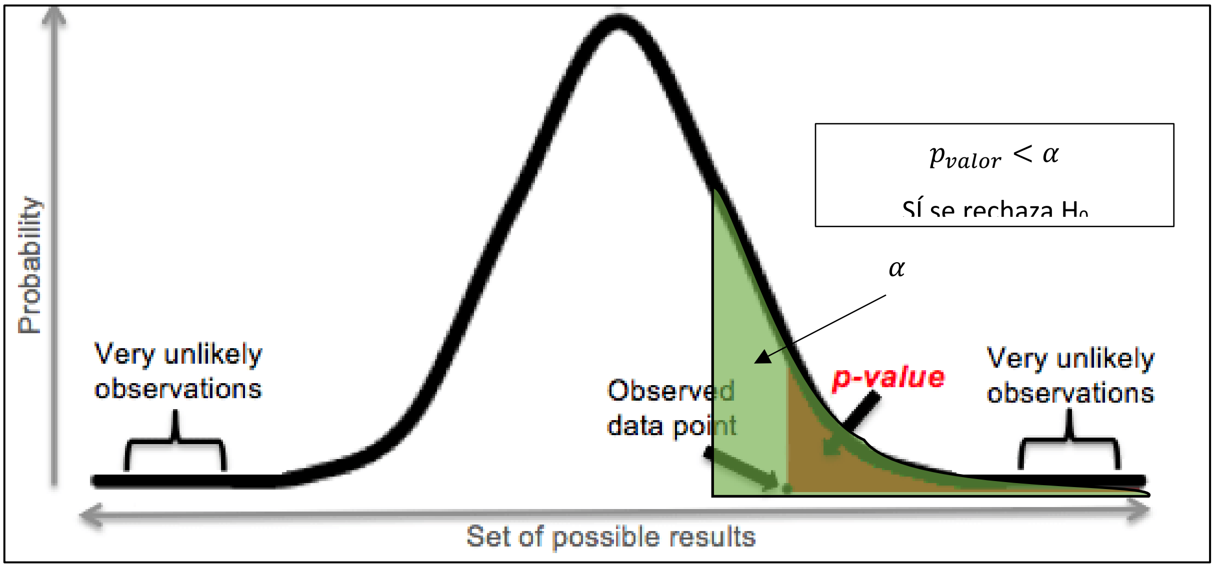
- $Si\ p_{valor} > \alpha$ NO se rechaza H0

  Probablemente la desviación ES CAUSAL.
  El estadístico de contraste NO ingresa en la región de rechazo, el p-valor lo mantiene en la región de ACEPTACIÓN, fuera del área α.



- $Si\ p_{valor} < \alpha$ SÍ se rechaza H0

  Probablemente la desviación NO es casual.
  El estadístico de contraste SÍ ingresa en la región de rechazo, dentro del área α.



Como p-valor (área roja) es MENOR que α (área VERDE), el estadístico observado queda en la región de RECHAZO.

6.5. Método para la resolución de un contraste de hipótesis

La resolución de un contraste de hipótesis se basa en el cálculo de las regiones de aceptación y rechazo y varía según el tipo de parámetro θ estudiado, pero implica:

1. FORMULACIÓN DE HIPÓTESIS nula H_0 y la alternativa H_1 y elección del tipo de contraste según el contexto (bilateral, unilateral derecho o izquierdo).

2. Imposición de un NIVEL DE SIGNIFICACIÓN α.

3. Determinación del ESTADÍSTICO DE CONTRASTE adecuado al test que se realiza. Su expresión depende de:

    - La ley de distribución que sigue (solo se usa normal Z y T-Student en este curso).

    - El tipo de contraste.

4. DECIDIR si se RECHAZA O NO SE RECHAZA H_0. Para ello, hay que determinar la REGIÓN DE ACEPTACIÓN.
   Hay 2 opciones:

    a) Valor crítico:

    - $Si\ estadístico\ de\ contraste < valor\ crítico$

      NO SE RECHAZA H_0
      La muestra NO ingresa en la región de rechazo.

    - $Si\ estadístico\ de\ contraste > valor\ crítico$

      SÍ SE RECHAZA H_0
      La muestra SÍ ingresa en la región de rechazo.

    b) p-valor:

    - $Si\ p_{valor} > \alpha$ NO se rechaza H_0
      Probablemente la desviación es casual.

    - $Si\ p_{valor} \leq \alpha$ Se rechaza H_0
      Probablemente la desviación es significativa.

6.6. Error en el contraste de hipótesis

La inferencia conlleva 2 tipos de errores potenciales:

		Propiedades de la POBLACIÓN (REALES)	
		H_0 es CIERTA	H_0 es FALSA
Resultado de la inferencia (plausibles)	Aceptar H_0	Inferencia CORRECTA	ERROR de tipo II (β)
	Rechazar H_0	ERROR de tipo I (α)	Inferencia CORRECTA

El error de tipo I sucede si se RECHAZA H_0 cuando H_0 es CIERTA y su probabilidad α es:

$$p(Error\ de\ tipo\ I\ ) = \alpha = p\left(\frac{Rechazo\ de\ H_0}{H_0\ es\ CIERTA}\right)$$

Nótese que α coincide con el nivel de significación.

El error de tipo I sucede si se ACEPTA H_0 cuando H_0 es FALSA y su probabilidad β es:

$$p(Error\ de\ tipo\ II\ ) = \beta = p\left(\frac{Aceptar\ H_0}{H_0\ es\ FALSA}\right)$$

La mayoría de métodos para la minimización del error de tipo I provocan que el error de tipo II aumente, y viceversa. Es decir, existe una relación de PROPORCIONALIDAD INVERSA entre ambos tipos de error:

$$\alpha \propto \frac{1}{\beta}$$

De modo que, COMO NORMA GENERAL, se prefiere MINIMIZAR α, el error de tipo I, aunque eso suponga la maximización de β, el error de tipo II (esta estrategia fue descrita por Neyman y Pearson en el s. XX).

Se dice que esta estrategia general es CONSERVADORA, ya que tiende a la ACEPTACIÓN de H_0, es decir, la hipótesis nula de PARTIDA, en detrimento de H_1 que la MODIFICA. Operando de este modo, se fija α, pero no se fija β. Por eso, podemos afirmar que:

- El RECHAZO de H_0 **se hace con la seguridad fijada por el nivel de confianza**, lo cual acota el error de tipo I (que se RECHACE H_0 aún siendo CIERTA).
- La ACEPTACIÓN de H_0 **NO ES INFALIBLE**, puesto que no se controla el error de tipo II cometido (que se acepte H_0 aún siendo falsa).

Debido a esa asimetría, se dice "RECHAZO de H_0" pero se evita el uso de "ACEPTACIÓN de H_0". En su lugar, se usa "NO SE RECHAZA H_0".

EJEMPLO

La presunción de inocencia en el sistema judicial se basa en tomar como hipótesis de partida la INOCENCIA del acusado.
Un juicio se resolverá de forma CORRECTA si:
- El acusado es INOCENTE y se le declara INOCENTE.
- El acusado es CULPABLE y se le declara CULPABLE.

Se cometerá un error de tipo I (con probabilidad α) si:
- El acusado es INOCENTE y se le declara CULPABLE. Entonces:

$$\alpha = probabilidad\ de\ CONDENADO\ siendo\ INOCENTE = p\left(\frac{Rechazo\ de\ H_0}{H_0\ es\ CIERTA}\right)$$

Se cometerá un error de tipo II (con probabilidad β) si:

El acusado es CULPABLE y se le declara INOCENTE.

$$\beta = probabilidad\ de\ ABSUELTO\ siendo\ CULPABLE = p\left(\frac{Rechazo\ de\ H_0}{H_0\ es\ CIERTA}\right)$$

6.7. Contrastes de hipótesis de 1 población

6.7.1. CONTRASTE PARA LA MEDIA μ CON VARIANZA σ CONOCIDA

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$z = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	$\lvert z\rvert > \lvert z_{\frac{\alpha}{2}}\rvert$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\mu \leq \mu_0$	$\mu > \mu_0$		$z > z_\alpha$	$(-\infty, z_\alpha)$
Unilateral por la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$		$z < -z_\alpha$	$(-z_\alpha, \infty)$

NO SE RECHAZA H_0 si:

$$Bilateral(H_1: \mu \neq \mu_0) \rightarrow \bar{x} \in \left(\mu_0 - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$DERECHA\ (H_1: \mu > \mu_0) \rightarrow\ \bar{x} < \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

$$IZQUIERDA\ (H_1: \mu < \mu_0) \rightarrow\ \bar{x} > \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Aplicando el p-valor, NO SE RECHAZA H_0 si:

$$BILATERAL\ (H_1: \mu \neq \mu_0) \rightarrow p_{valor} = P(|Z| > |z|) = P(Z > |z|) + P(Z < -|z|) = 2(1 - P(Z < z)) > \alpha$$

$$DERECHA\ (H_1: \mu > \mu_0) \rightarrow\ p_{valor} =\ P(Z > z) > \alpha$$

$$IZQUIERDA\ (H_1: \mu < \mu_0) \rightarrow p_{valor} =\ P(Z < z) > \alpha$$

6.7.2. CONTRASTE PARA LA MEDIA μ CON varianza σ DESCONOCIDA

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$t = \frac{\bar{x} - \mu_0}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$	$\lvert t\rvert > t_{n-1,\frac{\alpha}{2}}$	$\left(-t_{n-1,\frac{\alpha}{2}}, t_{n-1,\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\mu \leq \mu_0$	$\mu > \mu_0$		$t > t_{n-1,\alpha}$	$(-\infty, t_{n-1,\alpha})$
Unilateral por la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$		$t < -t_{n-1,\alpha}$	$(-t_{n-1,\alpha}, \infty)$

Nótese el uso de la T-Student para el cálculo de la probabilidad en lugar de la Normal.

$$Bilateral(H_1: \mu \neq \mu_0) \rightarrow \bar{x} \in \left(\mu_0 - t_{n-1,\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}, \mu_0 + t_{n-1,\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}\right)$$

$$DERECHA\ (H_1: \mu > \mu_0) \rightarrow\ \bar{x} < \mu_0 + t_{n-1,\alpha} \cdot \frac{\hat{s}}{\sqrt{n}}$$

$$IZQUIERDA\ (H_1: \mu < \mu_0) \rightarrow\ \bar{x} > \mu_0 - t_{n-1,\alpha} \cdot \frac{\hat{s}}{\sqrt{n}}$$

Aplicando el p-valor, NO SE RECHAZA H_0 si:

$$BILATERAL\ (H_1: \mu \neq \mu_0) \rightarrow p_{valor} = P(\lvert t_{n-1}\rvert > \lvert t\rvert) = P(t_{n-1} > \lvert t\rvert) + P(t_{n-1} < -\lvert t\rvert)$$

$$p_{valor} = 2 \cdot P(t_{n-1} > \lvert t\rvert) > \alpha$$

$$DERECHA\ (H_1: \mu > \mu_0) \rightarrow\ p_{valor} =\ P(t_{n-1} > t) > \alpha$$

$$IZQUIERDA\ (H_1: \mu < \mu_0) \rightarrow p_{valor} =\ P(t_{n-1}\ < t) > \alpha$$

6.7.3. CONTRASTE PARA LA PROPORCIÓN p DE 1 MUESTRA

ES IMPRESCINDIBLE PODER APLICAR TLC: SOLO SE CUMPLE SI n ≥ 30

Contraste	H0	H1	Estadístico de contraste	Condición de RECHAZO de H0	Región de ACEPTACIÓN
Bilateral	$p = p_0$	$p \neq p_0$	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$	$z > z_{\frac{\alpha}{2}}$ ó $z < -z_{\frac{\alpha}{2}}$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$p \leq p_0$	$p > p_0$		$z > z_\alpha$	$(-\infty, z_\alpha)$
Unilateral por la izquierda	$p \geq p_0$	$p < p_0$		$z < -z_\alpha$	$(-z_\alpha, \infty)$

NO SE RECHAZA H0 si:

$$Bilateral(H_1: p \neq p_0) \rightarrow \hat{p} \in \left(p_0 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

$$DERECHA\ (H_1: p > p_0) \rightarrow\ \hat{p} < p_0 - z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$IZQUIERDA\ (H_1: p < p_0) \rightarrow\ \hat{p} > p_0 + z_\alpha \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

Aplicando el p-valor, NO SE RECHAZA H0 si:

$$BILATERAL\ (H_1: p \neq p_0) \rightarrow p_{valor} = P(|Z| > |z|) = P(Z > |z|) + P(Z < -|z|) = 2(1 - P(Z < z)) > \alpha$$

$$DERECHA\ (H_1: p > p_0) \rightarrow\ p_{valor} =\ P(Z > z) > \alpha$$

$$IZQUIERDA\ (H_1: p < p_0) \rightarrow p_{valor} =\ P(Z < z) > \alpha$$

6.7.4. CONTRASTE PARA LA VARIANZA σ^2 CON MEDIA μ CONOCIDA

NO SE USA EN ESTE CURSO

Contraste	H$_0$	H$_1$	Estadístico de contraste	Condición de RECHAZO de H$_0$	Región de ACEPTACIÓN
Bilateral	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$x = \frac{\sum(x_i - \mu)^2}{\sigma_0^2} \sim \chi_n^2$	$x > \chi^2_{n,\frac{\alpha}{2}}$ ó $x < \chi^2_{n,1-\frac{\alpha}{2}}$	$\left(\chi^2_{n,\frac{\alpha}{2}}, \chi^2_{n,1-\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\sigma \leq \sigma_0$	$\sigma > \sigma_0$		$x > \chi^2_{n,\alpha}$	$(-\infty, \chi^2_{n,\alpha})$
Unilateral por la izquierda	$\sigma \geq \sigma_0$	$\sigma < \sigma_0$		$x < \chi^2_{n,\alpha}$	$(\chi^2_{n,\alpha}, \infty)$

Nótese el uso de la χ^2 para el cálculo de la probabilidad en lugar de la Normal.

6.7.5. CONTRASTE PARA LA VARIANZA σ^2 CON MEDIA μ DESCONOCIDA

NO SE USA EN ESTE CURSO

Contraste	H$_0$	H$_1$	Estadístico de contraste	Condición de RECHAZO de H$_0$	Región de ACEPTACIÓN
Bilateral	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$x = \frac{\sum(x_i - \bar{x})^2}{\sigma_0^2} \sim \chi_{n-1}^2$	$x > \chi^2_{n-1,\frac{\alpha}{2}}$ ó $x < \chi^2_{n-1,1-\frac{\alpha}{2}}$	$\left(\chi^2_{n-1,\frac{\alpha}{2}}, \chi^2_{n-1,1-\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\sigma \leq \sigma_0$	$\sigma > \sigma_0$		$x > \chi^2_{n-1,\alpha}$	$(-\infty, \chi^2_{n-1,\alpha})$
Unilateral por la izquierda	$\sigma \geq \sigma_0$	$\sigma < \sigma_0$		$x < \chi^2_{n-1,\alpha}$	$(\chi^2_{n-1,\alpha}, \infty)$

Nótese el uso de n – 1 grados de libertad por usar el estimador MUESTRAL $\bar{x}$ a falta del parámetro poblacional μ.

6.7.6. CONTRASTE PARA DATOS APAREJADOS

LOS DATOS APAREJADOS PERTENECEN A 1 SOLA POBLACIÓN.
NO CONFUNDIR CON CONTRASTE DE MEDIAS DE 2 POBLACIONES.

Los datos aparejados resultan de la observación de 2 variables en los mismos individuos de forma que de cada individuo hay 2 observaciones (como, por ejemplo, 2 observaciones de la misma muestra en distintos momentos del tiempo).

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu_d = 0$	$\mu_d \neq 0$	$t = \frac{\bar{d}}{\frac{s_{\bar{d}}}{\sqrt{n}}} \sim t_{n-1}$	$\lvert t \rvert > t_{n-1,\frac{\alpha}{2}}$	$\left(-t_{n-1,\frac{\alpha}{2}}, t_{n-1,\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\mu_d \leq 0$	$\mu_1 > 0$		$t > t_{n-1,\alpha}$	$(-\infty, t_{n-1,\alpha})$
Unilateral por la izquierda	$\mu_1 \geq 0$	$\mu_1 < 0$		$t < -t_{n-1,\alpha}$	$(-t_{n-1,\alpha}, \infty)$

Donde $s_{\bar{d}}$ es la cuasidesviación típica de la diferencia, calculada a partir de:

$$s_{\bar{d}}^2 = \frac{\left(\sum_{i=1}^{n} x_i^2\right) - n \cdot \bar{d}^2}{n-1}$$

6.8. Intervalos de confianza en la resolución de contraste de hipótesis

El criterio para la decisión de un contraste mediante el p-valor aplicado a un intervalo de confianza dado es:

- $0 \in Intervalo\ de\ confianza \leftrightarrow NO\ SE\ RECHAZA\ H_0$

- $0 \notin Intervalo\ de\ confianza \leftrightarrow Se\ RECHAZA\ H_0$

Para una proporción:

- $Si\ p_0 \notin Intervalo\ de\ confianza \rightarrow Se\ RECHAZA\ H_0$

- $Si\ p_0 \in Intervalo\ de\ confianza \rightarrow Se\ ACEPTA\ H_0$

6.9. Contrastes de hipótesis de 2 poblaciones

6.9.1. CONTRASTE PARA LA DIFERENCIA DE MEDIAS $\mu_1 - \mu_2$ CON VARIANZAS POBLACIONALES CONOCIDAS σ_1 Y σ_2

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$z > z_{\frac{\alpha}{2}}$ ó $z < -z_{\frac{\alpha}{2}}$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$		$z > z_\alpha$	$(-\infty, z_\alpha)$
Unilateral por la izquierda	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$		$z < -z_\alpha$	$(-z_\alpha, \infty)$

Aplicando el p-valor, NO SE RECHAZA H_0 si:

$$BILATERAL\ (H_1: \mu_1 \neq \mu_2) \rightarrow p_{valor} = 2P(Z > |z|) > \alpha$$

$$DERECHA\ (H_1: \mu_1 > \mu_2) \rightarrow p_{valor} = P(Z > z) > \alpha$$

$$IZQUIERDA\ (H_1: \mu_1 < \mu_2) \rightarrow p_{valor} = P(Z < z) > \alpha$$

Lo anterior se demuestra así:
A partir de 2 poblaciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, se toma una muestra de cada una, de tamaño n_1 y n_2. Las medias muestrales $\bar{X}_1$ y $\bar{X}_2$ se distribuyen como:

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$$

$$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

De acuerdo con la propiedad de la suma de variables normales:

$$\mathrm{Z}_1 + \mathrm{Z}_2 \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

Se cumple que la diferencia de medias muestrales $\bar{X}_1 - \bar{X}_2$ verifica:

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Tipificando esta variable diferencia $\bar{X}_1 - \bar{X}_2$ se obtiene:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Nótese que $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ es el error estándar.

Ahora, bajo el supuesto de hipótesis nula CIERTA, se verifica que $\mu_1 = \mu_2$, entonces el término $\mu_1 - \mu_2$ = 0. De lo cual:

$$\frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Y este es el estadístico de contraste bajo la premisa de H_0 CIERTA.

6.9.2. CONTRASTE PARA LA DIFERENCIA DE MEDIAS $\mu_1 - \mu_2$ CON VARIANZAS POBLACIONALES DESCONOCIDAS PERO IGUALES $\sigma_1 = \sigma_2$

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$t = \frac{\bar{x}_1 - \bar{x}_2}{s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$	$\lvert t\rvert > t_{n_1+n_2-2,\frac{\alpha}{2}}$	$\left(-t_{n_1+n_2-2,\frac{\alpha}{2}}, t_{n_1+n_2-2,\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$		$t > t_{n_1+n_2-2,\alpha}$	$(-\infty, t_{n_1+n_2-2,\alpha})$
Unilateral por la izquierda	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$		$t < -t_{n_1+n_2-2,\alpha}$	$(-t_{n_1+n_2-2,\alpha}, \infty)$

Nótese que el error estándar es $s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ donde s_c es la DESVIACIÓN TÍPICA COMÚN, que se define como:

$$s_c^2 = \frac{(n_1-1)\cdot\hat{s}_1^2 + (n_2-1)\cdot\hat{s}_2^2}{n_1+n_2-2}$$

Donde $\hat{s}_1^2$ y $\hat{s}_2^2$ son las CUASIVARIANZAS MUESTRALES de cada muestra.

Aplicando el p-valor, NO SE RECHAZA H_0 si:

$$BILATERAL\ (H_1: \mu_1 \neq \mu_2) \rightarrow p_{valor} = 2P\left(t_{n_1+n_2-2,\frac{\alpha}{2}} > \lvert t\rvert\right) > \alpha$$

$$DERECHA\ (H_1: \mu_1 > \mu_2) \rightarrow p_{valor} = P\left(t_{n_1+n_2-2,\alpha} > t\right) > \alpha$$

$$IZQUIERDA\ (H_1: \mu_1 < \mu_2) \rightarrow p_{valor} = P\left(t_{n_1+n_2-2,\alpha} < t\right) > \alpha$$

6.9.3. CONTRASTE PARA LA DIFERENCIA DE PROPORCIONES $p_1 - p_2$ DE 2 POBLACIONES

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$p_1 = p_2$	$p_1 \neq p_2$	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{n_1\hat{p}_1 + n_1\hat{p}_1}{n_1 + n_2}(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$	$z > z_{\frac{\alpha}{2}}$ ó $z < -z_{\frac{\alpha}{2}}$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$
Unilateral por la derecha	$p_1 \leq p_2$	$p_1 > p_2$		$z > z_\alpha$	$(-\infty, z_\alpha)$
Unilateral por la izquierda	$p_1 \geq p_2$	$p_1 < p_2$		$z < -z_\alpha$	$(-z_\alpha, \infty)$

Donde el error estándar es:

$$\sqrt{p_c(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Y p_c es:

$$p_c = \frac{n_1\hat{p}_1 + n_1\hat{p}_1}{n_1 + n_2}$$

Que es la estimación muestral de la proporción (necesaria a falta de la proporción poblacional desconocida).

6.10. Ejemplos de contrastes de hipótesis para la media

6.10.1. EJEMPLO 1: CONTRASTE PARA LA MEDIA CON VARIANZA CONOCIDA BILATERAL

Se supone que la distribución de la estatura media de las personas de un país se ajusta a una normal $N(164{,}9)$ en cm. En una región del país se toma una muestra de 64 personas, cuya media es de 166 cm. ¿Se puede afirmar, con un nivel de significación de 0.05, que la estatura media de las personas de esa región es DIFERENTE de la de las personas del país en su conjunto?

Se tiene:

- Una distribución normal $N(164{,}9)$, es decir, con media μ = 164 y σ = 9.
- Una muestra de 64 observaciones cuya media muestral es $\bar{x} = 166$.
- Un nivel de significación α = 0.05.

Entonces:

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

Se desea realizar un contraste de hipótesis para la MEDIA de 1 población con VARIANZA CONOCIDA.

En este caso, es BILATERAL, ya que se desea saber si le estatura "es DIFERENTE de" y no "mayor o menor que". Es decir:

$$\begin{cases} H_0: & \mu & = & \mu_0 = 164 \\ H_1: & \mu & \neq & \mu_0 = 164 \end{cases}$$

Entonces:

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	$Z > z_{\frac{\alpha}{2}}$ ó $Z < -z_{\frac{\alpha}{2}}$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$

2. Como la distribución es Normal, el ESTADÍSTICO de contraste es $Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$. En este caso:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow Z = \frac{166 - 164}{\frac{9}{\sqrt{64}}} = \frac{2}{\frac{9}{8}} = \frac{16}{9} = 1.\hat{7}$$

3. A partir del nivel de significación α se calcula la región de REGIÓN DE ACEPTACIÓN, que es este contraste es $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$.

$$\alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

Entonces, por definición de $z_{\frac{\alpha}{2}}$ se cumple:

$$p\left(Z < -z_{\frac{\alpha}{2}}\right) = \underbrace{p\left(Z > z_{\frac{\alpha}{2}}\right)}_{\substack{No\ está\ en \\ la\ tabla\ Z}} = 0.025$$

Por complementariedad:

$$\underbrace{p\left(Z \leq z_{\frac{\alpha}{2}}\right)}_{\substack{Sí\ está\ en \\ la\ tabla\ Z}} = 1 - 0.025 = 0.975$$

De lo cual, el valor crítico $z_{\frac{\alpha}{2}}$ es:

$$\underbrace{p\left(Z \leq z_{\frac{\alpha}{2}}\right)}_{\substack{Sí\ está\ en \\ la\ tabla\ Z}} = 0.975 \rightarrow z_{\frac{\alpha}{2}} = \underbrace{1.96}_{\substack{fila\ 1.9 \\ col\ 0.06}}$$

Entonces, la región de aceptación es $(-1.96, +1.96)$.

4. Se verifica si el estadístico Z calculado pertenece a la región de aceptación. En este caso:

$$Z = 1.\hat{7} \in (-1.96, +1.96)$$

5. Se decide sobre la hipótesis: el estadístico de contraste pertenece al intervalo cerrado por el valor crítico, es decir, está en la región de aceptación, por tanto, NO SE RECHAZA H_0.

   Se afirma con un nivel de confianza del 95%: la media de estatura del país es 164 cm.

   Alternativamente, se puede resolver mediante un intervalo de confianza para la media:

$$I_{1-\alpha}(\mu) = \left(\mu_0 \mp z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) \rightarrow I_{1-0.05}(\mu) = \left(164 \mp 1.96 \cdot \frac{9}{\sqrt{64}}\right)$$
$$= (161.795, 166.205)$$

   Y evaluando si la media MUESTRAL está dentro del intervalo calculado:

$$\bar{x} = 166 \in (161.795, 166.205)$$

   Como 166 se encuentra en el intervalo, NO SE RECHAZA H_0.

6.10.2. EJEMPLO 2: CONTRASTE PARA LA MEDIA CON VARIANZA CONOCIDA UNILAT. DERECHA

Se supone que la distribución de la estatura media de las personas de un país se ajusta a una normal $N(164{,}9)$ en cm. en una región del país se toma una muestra de 64 personas, cuya media es de 166 cm. ¿Se puede afirmar, con un nivel de significación de 0.05, que la estatura media de las personas de esa región es MAYOR que la media de estatura de las personas del país en su conjunto?

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

Se desea realizar un contraste de hipótesis para la MEDIA de 1 población con VARIANZA CONOCIDA.

En este caso, UNILATERAL POR LA DERECHA, ya que se desea saber si "en esa región es MAYOR que". Es decir:

$$\begin{cases} H_0: & \mu \leq \mu_0 = 164 \\ H_1: & \mu > \mu_0 = 164 \end{cases}$$

La formulación de H_0 y H_1 es:

"¿Son las personas de esa región realmente más altas que en resto del país?" (en cuyo caso se aceptaría H_1).

"¿O bien la altura media en esta muestra (MAYOR que μ) responde a cómo se hizo el muestreo?" (y se acepta H_0)".

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la derecha	$\mu \leq \mu_0$	$\mu > \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	$Z > z_\alpha$	$(-\infty, z_\alpha)$

2. El ESTADÍSTICO de contraste es $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$, pues la distribución es Normal. En este caso:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow Z = \frac{166 - 164}{\frac{9}{\sqrt{64}}} = \frac{2}{\frac{9}{8}} = \frac{16}{9} = 1.\hat{7}$$

3. La REGIÓN DE ACEPTACIÓN en este tipo de contraste es $(-\infty, z_\alpha)$. El VALOR CRÍTICO z_α se deduce a partir del nivel de significación:

$$\alpha = 0.05 \rightarrow p(Z > z_\alpha) = 0.05$$

Entonces, por complementariedad:

$$1 - p(Z \leq z_\alpha) = 0.05 \rightarrow p(Z \leq z_\alpha) = 0.95$$

Que se observa en la tabla como:

$$z_\alpha = 1.645\ (fila\ 1.6\ entre\ col\ 0.04\ y\ 0.05)$$

Entonces, la región de aceptación es $(-\infty, 1.645)$.

4. Se verifica si el estadístico Z calculado pertenece a la región de aceptación. En este caso:

$$Z = 1.\hat{7} \notin (-\infty, +1.645)$$

5. El estadístico de contraste NO pertenece a la región de aceptación, entonces se RECHAZA H_0. Se afirma con un nivel de confianza del 95%: la media de estatura en esa región es MAYOR que en el resto del país.

   Alternativamente, se puede resolver mediante un intervalo de confianza para la media:

$$I_{1-\alpha}(\mu) = \left(-\infty, \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) \rightarrow I_{0.95}(\mu) = \left(-\infty, 164 + 1.645 \cdot \frac{9}{\sqrt{64}}\right)$$

   O sea:

$$I_{1-\alpha}(\mu) = (-\infty, 165.850)$$

   Y evaluando si la media MUESTRAL está dentro del intervalo calculado:

$$\bar{x} = 166 \notin (-\infty, 165.850)$$

   Se alcanza la misma conclusión: se puede rechazar H_0.

6.10.3. EJEMPLO 3: CONTRASTE PARA LA MEDIA CON VARIANZA CONOCIDA UNILAT. IZQUIERDA

Se supone que la distribución de la estatura media de las personas de un país se ajusta a una normal $N(164{,}9)$ en cm. en una región del país se toma una muestra de 36 personas, cuya media es de 162 cm. ¿Se puede afirmar, con un nivel de significación de 0.05, que la estatura media de las personas de esa región es MENOR que la media de estatura de las personas del país en su conjunto?

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

Se desea realizar un contraste de hipótesis para la MEDIA de 1 población con VARIANZA CONOCIDA. En este caso, UNILATERAL POR LA IZQUIERDA, ya que:

$$\begin{cases} H_0: & \mu & \geq & \mu_0 = 164 \\ H_1: & \mu & < & \mu_0 = 164 \end{cases}$$

La formulación de H_0 y H_1 es: "¿son las personas de esa región realmente más BAJAS que en resto del país? (en cuyo caso se aceptaría H_1) ¿O bien la altura media en esta muestra (MENOR que μ) responde a cómo se hizo el muestreo? (y se acepta H_0)".

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0{,}1)$	$Z < -z_\alpha$	$(-z_\alpha, +\infty)$

2. El ESTADÍSTICO de contraste es $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$, pues la distribución es Normal. En este caso:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \to Z = \frac{162 - 164}{\frac{9}{\sqrt{36}}} = \frac{-2}{\frac{9}{6}} = -\frac{4}{3} = -1.\hat{3}$$

3. En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $(-z_\alpha, +\infty)$. El VALOR CRÍTICO se deduce a partir del nivel de significancia:

$$\alpha = 0.05 \to p(Z > z_\alpha) = 0.05 \to 1 - p(Z \leq z_\alpha) = 0.05 \to p(Z \leq z_\alpha) = 0.95$$

$$z_{0.95} = 1.645 \ (fila \ 1.6 \ entre \ col \ 0.04 \ y \ 0.05)$$

Entonces, la región de aceptación es $(-1.645, +\infty)$.

4. Se verifica si el estadístico Z calculado pertenece a la región de aceptación. En este caso:

$$Z = -1.\hat{3} \in (-1.645, +\infty)$$

5. El estadístico de contraste SÍ pertenece a la región de aceptación, entonces se ACEPTA H_0. Se afirma con un nivel de confianza del 95%: la media de estatura en esa región es NO ES MENOR que en el resto del país.

   Alternativamente, se puede resolver mediante un intervalo de confianza para la media:

$$I_{1-\alpha}(\mu) = \left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty\right) \rightarrow I_{0.95}(\mu) = \left(164 - 1.645 \cdot \frac{9}{\sqrt{36}}, \infty\right) = (161.532, \infty)$$

   Y evaluando si la media MUESTRAL está dentro del intervalo calculado:

$$\bar{x} = 162 \in (161.532, \infty)$$

Se alcanza la misma conclusión: no se puede rechazar H_0.

6.11. Ejemplos de contrastes de hipótesis para la proporción

6.11.1. EJEMPLO 1: CONTRASTE PARA LA PROPORCIÓN BILATERAL

La proporción de un atributo de una población es p = 0.3. 3 encuestadores diferentes (A, B y C) hacen muestreos distintos. El encuestador A ha obtenido: para n = 36 una proporción muestral $\hat{p} = 0.38$ y afirma que la proporción HA CAMBIADO con un nivel de significación de 0.05. ¿Se puede afirmar lo que el encuestador A propone?

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

   Se desea realizar un contraste de hipótesis para la PROPORCIÓN de 1 población. En este caso, BILATERAL, ya que se plantea si "la proporción HA CAMBIADO" o, por contrario, es la previamente conocida (no ha cambiado). Es decir:

$$\begin{cases} H_0: & p & = & p_0 = 0.3 \\ H_1: & p & \neq & p_0 = 0.3 \end{cases}$$

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$	$Z > z_{\frac{\alpha}{2}}$ ó $z < -z_{\frac{\alpha}{2}}$	$\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$

2. El ESTADÍSTICO de contraste es $z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, pues la distribución es Normal. Se tiene:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow z = \frac{0.38 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{36}}} = 1.047$$

3. En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $\left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)$. El VALOR CRÍTICO se deduce a partir del nivel de significación:

$$\alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow P\left(z < -z_{\frac{\alpha}{2}}\right) = P\left(z > z_{\frac{\alpha}{2}}\right) = 0.025$$

   Por complementariedad:

$$P\left(Z \leq z_{\frac{\alpha}{2}}\right) = 1 - 0.025 = 0.975$$

Entonces, en cada cola se aloja un 2,5% de probabilidad, es decir, $z_{\frac{\alpha}{2}}$:

$$z_{\frac{0.05}{2}} = 1.96\ (fila\ 1.9\ col\ 0.06)$$

Entonces, la región de aceptación es $(-1.96, +1.96)$.

4. Se verifica si el estadístico Z calculado pertenece a la región de aceptación. En este caso:

$$Z = 1.047 \in (-1.96, +1.96)$$

5. El estadístico de contraste pertenece a la región de aceptación, entonces se ACEPTA H_0. Se afirma con 95% de confianza: la proporción del atributo en la población es de 0.3.
   Alternativamente, se puede resolver mediante un intervalo de confianza para la media:

$$I_{1-\alpha}(p) = \left(p_0 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}}, p_0 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}}\right)$$

$$I_{1-0.05}(p) = \left(0.3 - 1.96 \cdot \sqrt{\frac{0.3 \cdot (1-0.3)}{36}}, 0.3 + 1.96 \cdot \sqrt{\frac{0.3 \cdot (1-0.3)}{36}}\right) = (0.150, 0.449)$$

Evaluando si la proporción MUESTRAL está dentro del intervalo calculado:

$$\hat{p} = 0.38 \in (0.150, 0.449)$$

Como 0.38 se encuentra en el intervalo, se ACEPTA H_0 y NO se puede confirmar el cambio de proporción propuesto por el encuestador A.

6.11.2. EJEMPLO 2: CONTRASTE PARA LA PROPORCIÓN UNILATERAL HACIA LA IZQUIERDA

La proporción de un tributo de una población es p = 0.3. 3 encuestadores diferentes (A, B y C) hacen muestreos distintos. El encuestador B ha obtenido: para n = 100 una proporción muestral $\hat{p} = 0.28$ y afirma que la proporción HA DISMUNIDO con un nivel de significación de 0.1. ¿Se puede afirmar lo que el encuestador B propone?

Se tiene $p_0 = 0.3\ y\ n = 100\ y\ \hat{p} = 0.28\ y\ \alpha = 0.1$

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

Se desea realizar un contraste de hipótesis para la PROPORCIÓN de 1 población. En este caso, UNILATERAL HACIA LA IZQUIERDA, ya que la hipótesis alternativa dice "ha disminuido" (si es MENOR):

$$\begin{cases} H_0: & p & \geq & p_0 = 0.3 \\ H_1: & p & < & p_0 = 0.3 \end{cases}$$

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la izquierda	$p \geq p_0$	$p < p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$Z < -z_\alpha$	$(-z_\alpha, \infty)$

2. El ESTADÍSTICO de contraste es $Z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, pues la distribución es Normal. Se tiene:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow Z = \frac{0.28 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{100}}} = -0.436$$

3. En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $(-z_\alpha, +\infty)$. El VALOR CRÍTICO es:

$$\alpha = 0.1 \rightarrow p(z > z_\alpha) = 0.1 \rightarrow 1 - p(z \leq z_\alpha) = 0.1 \rightarrow p(z \leq z_\alpha) = 0.9$$

Entonces:

$$p(z \leq z_{0.1}) = 0.9 \rightarrow z_{0.1} = \underbrace{1.28}_{\substack{fila\ 1.2 \\ col\ 0.08}} \quad (NO\ ES\ EXACTO: probabilidad = 0.8997)$$

Entonces, la región de aceptación es $(-1.28, +\infty)$.

4. Se verifica si el estadístico z calculado pertenece a la región de aceptación. En este caso:

$$Z = -0.436 \in (-1.28, +\infty)$$

5. El estadístico de contraste SÍ pertenece a la región de aceptación, entonces se ACEPTA H_0. Se afirma con un nivel de confianza del 90%: la proporción NO HA DISMINUIDO en la población.

   Alternativamente, se resuelve mediante un intervalo de confianza para la proporción:

$$I_{1-\alpha}(p) = \left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}}, \infty\right)$$

$$I_{1-0.1}(p) = \left(0.3 - 1.28 \cdot \sqrt{\frac{0.3 \cdot (1-0.3)}{100}}, +\infty\right) = (0.2413, +\infty)$$

   Y evaluando si la media MUESTRAL está dentro del intervalo calculado:

$$\hat{p} = 0.28 \in (0.2413, +\infty)$$

   Se alcanza la misma conclusión: no se puede rechazar H_0 y por tanto, no hay evidencia suficiente a favor de la afirmación del encuestador B.

6.11.3. EJEMPLO 3: CONTRASTE PARA LA PROPORCIÓN UNILATERAL HACIA LA DERECHA

La proporción de un atributo de una población es p = 0.3. 3 encuestadores diferentes (A, B y C) hacen muestreos distintos. El encuestador C ha obtenido: Para n = 900 una proporción muestral $\hat{p} = 0.33$ y afirma que la proporción HA AUMENTADO con un nivel de significación de 0.05. ¿Se puede afirmar lo que el encuestador C propone?

Se tiene $p_0 = 0.3\ y\ n = 900\ y\ \hat{p} = 0.33\ y\ \alpha = 0.05$

1. Se formula la hipótesis nula H_0 y la alternativa H_1 y se elige el tipo de contraste:

Se desea realizar un contraste de hipótesis para la PROPORCIÓN de 1 población. En este caso, UNILATERAL HACIA LA DERECHA, ya que:

$$\begin{cases} H_0: & p \leq p_0 = 0.3 \\ H_1: & p > p_0 = 0.3 \end{cases}$$

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la derecha	$p \leq p_0$	$p > p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$Z > z_\alpha$	$(-\infty, z_\alpha)$

2. El ESTADÍSTICO de contraste es $Z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, pues la distribución es Normal. Se tiene:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow z = \frac{0.33 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{900}}} = 1.963$$

3. En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $(-\infty, z_\alpha)$. El VALOR CRÍTICO es:

$$\alpha = 0.05 \rightarrow p(z > z_\alpha) = 0.05 \rightarrow 1 - p(z \leq z_\alpha) = 0.05 \rightarrow p(z \leq z_\alpha) = 0.95$$

$$z_\alpha = 1.645\ (fila\ 1.6\ entre\ col\ 0.04\ y\ 0.05)$$

Entonces, la región de aceptación es $(-\infty, 1.645)$.

4. Se verifica si el estadístico z calculado pertenece a la región de aceptación. En este caso:

$$z = 1.963 \notin (-\infty, 1.645)$$

5. El estadístico de contraste NO pertenece a la región de aceptación, entonces se RECHAZA H_0. Se afirma con un nivel de confianza del 95%: la proporción SÍ HA AUMENTADO en la población.

   Alternativamente, se resuelve mediante un intervalo de confianza para la proporción:

$$I_{1-\alpha}(p) = \left(-\infty, p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot (1-p_0)}{n}},\right)$$

$$I_{1-0.05}(p) = \left(-\infty, 0.3 - 1.645 \cdot \sqrt{\frac{0.3 \cdot (1-0.3)}{900}}\right) = (-\infty, 0.274)$$

   Y evaluando si la media MUESTRAL está dentro del intervalo calculado:

$$\hat{p} = 0.28 \notin (-\infty, 0.274)$$

   Se alcanza la misma conclusión: Se puede rechazar H_0 y por tanto, SÍ se puede apoyar la afirmación del encuestador C: la proporción ha AUMENTADO en la población.

PEC5 – Cuestionario

PREGUNTA 1

El contraste de hipótesis implica una decisión sobre si se debe aceptar o rechazar la hipótesis nula en base a la información muestral. En la toma de esta decisión:

Si queremos contrastar si la media del tiempo de ejecución de un programa que sigue una distribución normal es diferente de 22 segundos sabiendo que la desviación poblacional es 22 segundos, entonces: Hacemos un contraste de medias z bilateral.

Si queremos contrastar si la nota media de dos asignaturas del mismo grupo de estudiantes de la UOC es la misma: Hacemos un contraste de diferencia de medias apareadas t bilateral.

Si queremos contrastar si la nota media de estudiantes de la UOC es mayor que la de los estudiantes de la UPC: Hacemos un contraste de diferencia de medias t con hipótesis alternativa MedUOC > MedUPC.

PREGUNTA 2

El número de accesos a un servidor, X, sigue una distribución normal de media μ y desviación estándar 43. Se quiere contrastar H_0: μ=260 frente a H_1: μ<260:

Se tiene:
- $X \sim (\mu, 43)$
- Un contraste de hipótesis UNILATERAL POR LA IZQUIERDA (H_1: $\mu < \mu_0$).

a) Calcula el error estándar si la muestra fuera de n = 100 observaciones:

El error estándar de la media es: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \rightarrow \sigma_{\bar{x}} = \frac{43}{\sqrt{100}} = 4.3$

b) Calcular el estadístico de contraste si la media muestral μ = 250 y la muestra es de n = 100 observaciones:

En un contraste para la MEDIA con σ CONOCIDA, el estadístico es:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \rightarrow Z = \frac{250 - 260}{\frac{43}{\sqrt{100}}} = -2.3255$$

c) Calcula el valor crítico si el nivel de significación es del 5%:

Se desea calcular el valor crítico para un contraste UNILATERAL POR LA IZQUIERDA, es decir, el valor $-z_\alpha$. A partir del nivel de signficacón α:

$$P(Z < -z_\alpha) = 0.05 = \underbrace{P(Z > z_\alpha)}_{Por\ simetría} = \underbrace{1 - P(Z \leq z_\alpha)}_{Por\ complementariedad} = 0.05$$

Es decir:

$$P(Z \leq z_{0.05}) = 0.95$$

Entonces, el VALOR CRÍTICO $-z_{0.05}$ es:

$$z_{0.05} = 1.645\ (fila\ 1.6\ entre\ col\ 0.04\ y\ 0.05) \rightarrow -z_{0.05} = -1.645$$

PREGUNTA 3

El tiempo de comunicación entre dispositivos es crítico en una red y no debe estar por debajo de un cierto valor de 3 milisegundos (ms) para no comprometer su funcionamiento. Por ello el administrador de la red realizó 10 mediciones aleatorias del tiempo de comunicación entre dispositivos para decidir si pedía ampliar el ancho de banda. Los datos obtenidos en ms fueron: {3, 4, 4, 0, 5, 2, 4, 4, 0, 0}

Suponiendo que el tiempo de comunicación de la red en ms sigue una **distribución normal**, calculad (con un nivel de confianza del 90%) lo siguiente:

SUPONIENDO NORMALIDAD en X:

- Con varianza poblacional σ^2 CONOCIDA:
  → La media de X se ajusta Z Normal
- Con varianza poblacional σ^2 DESCONOCIDA:
  → La media de X se ajusta a T-Student si $n \leq 30$
  → La media de X se ajusta a Z normal si $n > 30$

En este caso, se DESCONOCE σ y n < 30.
Por tanto, se recurre a la T-Student.

a) Determinar el error estándar de la media $s_{\bar{x}}$

Para el cálculo de $s_{\bar{x}}$ no se dispone de σ y se deberá calcular la CUASIVARIANZA MUESTRAL s.

1. Se calcula la media muestral:

$$\bar{x} = \frac{3 + 4 \cdot 4 + 0 \cdot 3 + 5 + 2}{10} = 2.6$$

2. Se calcula la cuasidesviación típica MUESTRAL $\hat{s}$:

$$\hat{s}^2 = \frac{\left(\sum_{i=1}^{n} x_i^2\right) - n \cdot \bar{x}^2}{n-1} = \frac{(3^2 + 4^2 \cdot 4 + 0^2 \cdot 3 + 5^2 + 2^2) - 10 \cdot 2.6^2}{10-1} \rightarrow \hat{s} = \sqrt{3.8\hat{2}} = 1.955$$

3. Entonces, el error estándar es:

$$s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}} = \frac{1.955}{\sqrt{10}} = 0.61824$$

b) Valor crítico positivo

Como se cumplen 3 requisitos:

1. Se supone normalidad de X.
2. Se tiene n < 30
3. Se DESCONOCE σ

Se recurre a T-Student y no a Normal Z.

Con 90% de confianza y n – 1 = 10 – 1 = 9 grados de libertad se tiene en la tabla T:

$$1-\alpha=0.9 \rightarrow \alpha=0.1 \rightarrow P\left(t>t_{n-1,\alpha}\right)=0.1 \rightarrow t_{10-1,0.1}=1.383$$

$$(fila\ 9\ grado\ de\ libertad\ y\ columna\ 0.1)$$

El valor crítico es $t_{9,\ 0.1}$ = 1.383.

c) Estadístico de contraste

No se conoce la varianza poblacional σ^2, entonces el estadístico de contraste para la media es t (y no z de una normal estándar):

$$t=\frac{\bar{x}-\mu_0}{\frac{\hat{s}}{\sqrt{n}}} \rightarrow t=\frac{2.6-3}{\frac{1.955}{\sqrt{10}}}=-0.647$$

d) H = 0: Aceptamos la hipótesis nula; H = 1: Aceptamos la alternativa

Se trata de un contraste de medias UNILATERAL POR LA IZQUIERDA, ya que se dice "el tiempo no debe estar por debajo". Es decir:

$$\begin{cases} H_0: & \mu & \geq & \mu_0=3 \\ H_1: & \mu & < & \mu_0=3 \end{cases}$$

Por tanto, la región de aceptación es:

$$\left(-t_{n-1,\alpha},\infty\right) \rightarrow (-1.383,\infty)$$

Se observa que el estadístico de contraste t encontrado está en la región de aceptación:

$t=-0.647 \in (-1.383,\infty) \rightarrow Se\ acepta\ H_0$

PREGUNTA 4
Un fabricante de cables quiere comparar si la resistencia de su producto (A) es igual o no al de la competencia (C). Para ello realizó 10 mediciones aleatorias independientes en su producto y en el de la competencia. Los datos obtenidos fueron:

A: {2, 3, 1, 6, 3, 4, 5, 3, 2, 4}

C: {7, 2, 7, 4, 7, 5, 4, 7, 3, 6}

Se supone normalidad e igualdad de varianza poblacional. Para un nivel de confianza del 90%:

Se tiene un contraste de LA DIFERENCIA DE MEDIAS de **2 poblaciones INDEPENDIENTES** con VARIANZA DESCONOCIDA PERO EQUIVALENTE (supuesta en el enunciado) con α = 0.1 (90% de confianza):

$$\begin{cases} H_0: & \mu_1 & = & \mu_2 \\ H_1: & \mu_1 & \neq & \mu_2 \end{cases}$$

Se trata de un contraste BILATERAL, cuyo estadístico de contraste es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Antes de empezar:

1. Se calcula la media muestral de A y C:

$$\bar{x}_A = \frac{2+3+1+6+3+4+5+3+2+4}{10} = 3.3$$
$$\bar{x}_C = \frac{7+2+7+4+7+5+4+7+3+6}{10} = 5.2$$

2. Se calcula la cuasivarianza muestral $\hat{s}^2$ (corregida con n – 1) de ambas muestras A y C:

$$\hat{s}^2 = \frac{\left(\sum_{i=1}^{n} x_i^2\right) - n \cdot \bar{x}^2}{n-1}$$

$$\hat{s}_A^2 = \frac{(2 \cdot 2^2 + 3 \cdot 3^2 + 1^2 + 6^2 + 2 \cdot 4^2 + 5^2) - 10 \cdot 3.3^2}{10-1} = 2.2\hat{3}$$

$$\hat{s}_C^2 = \frac{(4 \cdot 7^2 + 2^2 + 2 \cdot 4^2 + 5^2 + 3^2 + 6^2) - 10 \cdot 5.2^2}{10-1} = 3.5\hat{1}$$

3. Se calcula al VARIANZA COMÚN s_t^2 y la desviación típica COMÚN s_t:

$$s_t^2 = \frac{(n_1-1)\cdot \hat{s}_1^2 + (n_2-1)\cdot \hat{s}_2^2}{n_1+n_2-2} = \frac{(10-1)\cdot 2.2\hat{3} + (10-1)\cdot 3.5\hat{1}}{10+10-2} = 2.87$$

Por tanto, DESVIACIÓN TÍPICA COMÚN s_t es:

$$s_t = \sqrt{2.87} = 1.694$$

a) Se desea $e_{\bar{x}_1-\bar{x}_2}$ error estándar de la diferencia de medias:

El error estándar de la diferencia de medias se observa en el estadístico de contraste:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\underbrace{s_t\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}_{Error\ estándar}}$$

Entonces:

$$e_{\bar{x}_1-\bar{x}_2} = s_t\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 1.694 \cdot \sqrt{\frac{1}{10}+\frac{1}{10}} = 0.7575$$

b) Valor crítico positivo

Se trata de un contraste BILATERAL con α = 0.1 y $n_1 + n_2 - 2 = 18$ grados de libertad. Entonces:

$$\alpha = 0.1 \rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05 \rightarrow t_{\underbrace{18}_{g.l.},\underbrace{0.05}_{\frac{\alpha}{2}}} = 1.7341$$

c) Estadístico de contraste es:

Se tiene:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_t\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightarrow t = \frac{3.3 - 5.2}{1.694\sqrt{\frac{1}{10}+\frac{1}{10}}} = -2.507$$

El estadístico de contraste POSITIVO es 2.507.

d) H = 0: Aceptamos la hipótesis nula; H = 1: Aceptamos la alternativa

Para este tipo de contraste, la condición de RECHAZO es:

$$|t| > t_{n_1+n_2-2,\frac{\alpha}{2}}$$

En este caso se tiene |t| = 2.507 y $t_{18,0.05} = 1.7341$:

$$2.507 > 1.73 \rightarrow Se\ RECHAZA\ H_0$$

$$Se\ introduce\ en\ Moodle\ H = 1\ (se\ acetpa\ H_1)$$

PREGUNTA 5

Un estudio de usabilidad pretende comparar la facilidad y velocidad de acceso de una aplicación. Se eligen 9 usuarios de forma aleatoria e independiente para probar dos interfaces diferentes y ver si sus rendimientos de usabilidad son iguales o no. Los datos obtenidos fueron:

Opción 1: {6, 4, 3, 4, 4, 5, 3, 1, 4}
Opción 2: {6, 4, 4, 7, 4, 7, 2, 4, 7}

Al tratarse de muestras DEPENDIENTES, ya que **el mismo usuario probó las dos interfaces (2 observaciones de cada individuo, es decir DATOS APAREJADOS NO INDEPENDIENTES**) podemos hacer un contraste de la diferencia de medias a partir de los datos.

Diferencia de medias: {0, 0, -1, -3, 0, -2, 1, -3, -3}

Suponiendo la normalidad de los datos y calculad para un nivel de confianza del 90%.

Se trata de un contraste BILATERAL para la diferencia de medias de **DATOS APAREJADOS** (NO INDEPENDIENTES):

$$\begin{cases} H_0: & \mu_d & = & 0 \\ H_1: & \mu_d & \neq & 0 \end{cases}$$

Cuyo estadístico es:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

Donde $\bar{d}$ es la media de las diferencias y s_d su CUASIDESVIACIÓN TÍPICA de la diferencia.

a) Calcular el error estándar de la diferencia de medias $e_{\bar{d}}$:

El error estándar para la diferencia de medias es:

$$e_{\bar{d}} = \frac{s_{\bar{d}}}{\sqrt{n}}$$

→ Se calcula la media:

$$\bar{d} = \frac{0+0-1-3+0-2+1-3-3}{9} = -1.\widehat{2}$$

→ Se calcula la CUASIDESVIACIÓN TÍPICA de la diferencia:

$$s_{\bar{d}}^2 = \frac{(\sum_{i=1}^{n} x_i^2)}{n-1} - \frac{n \cdot \bar{d}^2}{n-1} = \frac{(-1)^2 + 3 \cdot (-3)^2 + (-2)^2 + 1^2}{9-1} - \frac{9 \cdot \left(-1.\hat{2}\right)^2}{9-1} = 2.\hat{4}$$

$$s_{\bar{d}} = \sqrt{2.\hat{4}} = 1.563$$

Por tanto, el error es:

$$e_{\bar{d}} = \frac{s_{\bar{d}}}{\sqrt{n}} \to \frac{1.563}{\sqrt{9}} = 0.5211$$

b) Valor crítico positivo

Se trata de un contraste BILATERAL con α = 0.1 y n − 1 = 8 grados de libertad. Entonces:

$$\alpha = 0.1 \to \frac{\alpha}{2} = \frac{0.1}{2} = 0.05 \to t_{8,0.05} = 1.8595$$

c) Estadístico de contraste:

Se tiene:

$$t = \frac{\bar{d}}{\frac{s_{\bar{d}}}{\sqrt{n}}} \to t = \frac{-1.\hat{2}}{\frac{1.563}{\sqrt{9}}} = -2.3459$$

En positivo, el estadístico de contraste es 2.3459

d) H = 0: Aceptamos la hipótesis nula; H = 1: Aceptamos la alternativa

En este tipo de contraste de hipótesis, el criterio para RECHAZAR H_0 es:

$$|t| > t_{n-1,\frac{\alpha}{2}}$$

Se tiene:

$$|t| = 2.3459 > t_{n-1,\frac{\alpha}{2}} = 1.8595$$

Entonces:

$$Se\ RECHAZA\ H_0 \to Se\ introduce\ H = 1\ (aceptar\ H_1)$$

PREGUNTA 6
Una consultora de márquetin toma una muestra de 44 encuestas al azar para contrastar la proporción de consumidores que prefieren conectarse por móvil a internet. El año pasado la proporción era del 27.9%. En la muestra de este año un 36.27% encuestados manifestaron preferencia por el móvil. Se quiere determinar si dicha proporción se puede considerar que ha aumentado y así las hipótesis del contraste son H_0: p = 0.279, H_1: p > 0.279

Suponiendo la normalidad de los datos y calculad para un nivel de confianza del 99%.

Se desea realizar un contraste de hipótesis para la PROPORCIÓN de 1 población. En este caso, UNILATERAL HACIA LA DERECHA (se desea verificar "si ha aumentado", es decir, $\hat{p}$ si es MAYOR que la previa p_0), con $p_0 = 0.279$ y $\hat{p} = 0.362$. Es decir:

$$\begin{cases} H_0: & p \leq p_0 = 0.279 \\ H_1: & p > p_0 = 0.279 \end{cases}$$

Contraste	H_0	H_1	Estadístico de contraste	RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la derecha	$p \leq p_0$	$p > p_0$	$Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z > z_\alpha$	$(-\infty, z_\alpha)$

a) $\sigma_{\hat{p}}$ error estándar de la proporción:

El error estándar de la proporción es:

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} \to \sigma_{\hat{p}} = \sqrt{\frac{0.279 \cdot (1-0.279)}{44}} = 0.0676$$

Nótese que coincide con el denominador del estadístico, ya que define la variabilidad de la distribución.

b) Valor crítico positivo

En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $(-\infty, z_\alpha)$. El VALOR CRÍTICO es:

$$1 - \alpha = 0.99 \to \alpha = 0.01 \to p(Z > z_\alpha) = 0.01 \to 1 - p(Z \leq z_\alpha) = 0.01$$

Entonces:

$$1 - p(Z \leq z_\alpha) = 0.01 \to p(Z \leq z_\alpha) = \underbrace{0.99}_{\substack{fila\ 2.3 \\ ENTRE\ col\ 0.02 \\ y\ col\ 0.03}} \to z_\alpha = 2.3263$$

Nótese que el valor exacto no se puede obtener de la tabla: se consulta vía software y es 2.326348.

Entonces, la región de aceptación es $(-\infty, 2.326)$.

c) Estadístico de contraste:

El ESTADÍSTICO de contraste es $z = \frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$. Se tiene:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to z = \frac{0.3627 - 0.279}{\sqrt{\frac{0.279(1-0.279)}{44}}} = 1.2378$$

d) p-valor:

Para contrastes UNILATERALES por la DERECHA el p-valor es:

$$p - valor = P\left(\frac{X > x_0}{H_0}\right) = P(Z > z_0) = 1 - P(Z < z_0)$$

Es decir:

$$p - valor = 1 - P\left(Z < \underbrace{1.2378}_{\substack{No\ aparece \\ Se\ toma\ 1.24}} \right)$$

$$p - valor = 1 - \underbrace{0.8925}_{probabilidad\ de\ la\ abscisa\ \mathrm{z}_0=1.24} = 0.1075$$

e) Resolución del contraste:
H = 0: NO SE RECHAZA H_0.
H = 1: SE RECHAZA H_0.

Se verifica si el p-valor excede al nivel de significancia:

$$Si\ p - valor > \alpha \to NO\ SE\ RECHAZA\ H_0$$

En este caso: 0.1075 > 0.01

Por tanto, NO SE RECHAZA H_0 (se introduce H = 0).

PREGUNTA 7

Una empresa de antivirus sufrió 505 ataques diarios de hackers el año pasado. Este año quiere saber si el número de ataques ha disminuido. Con este propósito seleccionó los registros de 10 días al azar. Los números de ataques registrados fueron los siguientes:

{487, 477, 482, 476, 497, 497, 480, 476, 500, 487}

Suponiendo normalidad y que la desviación típica de la población es σ = 24, calcular lo siguiente y resolver el test tomando en cuenta un nivel de confianza del 90%:

Se desea realizar un contraste de hipótesis para la MEDIA de 1 población con VARIANZA CONOCIDA. En este caso, UNILATERAL POR LA IZQUIERDA, ya que se desea saber si "los ataques han disminuido":

$$\begin{cases} H_0: & \mu & \geq & \mu_0 = 505 \\ H_1: & \mu & < & \mu_0 = 505 \end{cases}$$

Contraste	H0	H1	Estadístico de contraste	Condición de RECHAZO de H0	Región de ACEPTACIÓN
Unilateral por la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	$z < -z_\alpha$	$(-z_\alpha, +\infty)$

a) $\sigma_{\bar{x}}$ error estándar de media:

El error estándar de la media es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \to \sigma_{\bar{x}} = \frac{24}{\sqrt{10}} = \frac{12}{5}\sqrt{10}$$

b) Valor crítico positivo

Para un 90% de confianza:

$$1 - \alpha = 0.9 \to \alpha = 0.1 \to p(Z < -z_\alpha) = 0.1 \to p(Z > z_\alpha) = 0.1$$

De lo cual:

$$p(Z \leq z_\alpha) = 1 - 0.1 = 0.9 \to z_{0.9} = 1.2816$$

Nótese que el valor exacto no se puede obtener de la tabla: se consulta que es 1.2816 exactamente.

c) Estadístico de contraste:

→ Se calcula la media de la muestra:

$$\bar{x} = \frac{2 \cdot 487 + 477 + 482 + 2 \cdot 476 + 2 \cdot 497 + 480 + 500}{10} = 485.9$$

El estadístico z es:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \to z = \frac{485.9 - 505}{\frac{24}{\sqrt{10}}} = -2.5166$$

d) p-valor:

El p-valor para un contraste UNILATERAL POR LA IZQUIERDA es:

$$p - valor = P\left(\frac{X \leq x_0}{H_0}\right) = P(Z \leq z_0) = P\,\underbrace{(Z \leq -2.5166)}_{no\ está\ en\ la\ tabla}$$

Por simetría:

$$p - valor = P\,\underbrace{(Z \leq -2.5166)}_{no\ está\ en\ la\ tabla} = P\,\underbrace{(Z > 2.5166)}_{no\ está\ en\ tabla}$$

Por complementariedad:

$$p - valor = P\,\underbrace{(Z > 2.5166)}_{no\ está\ en\ tabla} = \underbrace{1 - P(Z \leq 2.5166)}_{por\ complementariedad}$$

De lo cual:

$$p - valor = 1 - \underbrace{P(Z \leq 2.5166)}_{no\ está\ exactamente} = 1 - \underbrace{0.99405}_{Se\ toma\ entre\ 0.01\ y\ 0.02} = 0.00595$$

e) H = 0: Aceptamos la hipótesis nula; H = 1: Aceptamos la alternativa

Si p-valor > α → Se ACEPTA H_0.
En este caso, se observa:
p-valor = 0.00595 ≪ 0.1 = α → Se RECHAZA H_0

PREGUNTA 8
El coste medio, diario, de producción de placas madre es de 97 euros. La empresa fabricante mejora su línea de producción y piensa que ha reducido su coste de producción. Para comprobarlo escoge una muestra de 108 días al azar para los que calcula sus costes de producción obteniendo una media muestral de 90,09 euros y una desviación muestral de 42,57. Sabiendo que **no tenemos información sobre el tipo de distribución** que tiene el coste medio de producción diaria y tomando en cuenta un nivel de confianza del 99% calculad los siguientes datos y resolved el test.

Se desea realizar un contraste de hipótesis para la MEDIA de 1 población con VARIANZA DESCONOCIDA.

A priori, se podría pensar en usar t-Student, pero NO ES POSIBLE, YA QUE SE DESCONOCE EL COMPORTAMIENTO NORMAL DE LA MUESTRA.
Se desconoce si X sigue una distribución normal, pero como n > 30, en virtud de TLC, converge hacia una normal estándar. Es decir:

$$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \underbrace{\rightarrow}_{TLC} N(0,1)$$

Además, es UNILATERAL POR LA IZQUIERDA, ya que se desea saber si "su coste se ha reducido":

$$\begin{cases} H_0: & \mu \geq \mu_0 = 97 \\ H_1: & \mu < \mu_0 = 97 \end{cases}$$

Se tiene $\bar{x} = 90.09$ y $s = 42.57$ y α = 0.01 (99% de confianza)

Contraste	H_0	H_1	Estadístico de contraste	Condición de RECHAZO de H_0	Región de ACEPTACIÓN
Unilateral por la izquierda	$\mu \geq \mu_0$	$\mu < \mu_0$	$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	$Z < -z_\alpha$	$(-z_\alpha, \infty)$

a) $s_{\bar{x}}$ error estándar de media:

No se dispone de σ. El error estándar de la media es entonces:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \rightarrow s_{\bar{x}} = \frac{42.57}{\sqrt{108}} = 4.0963$$

b) Valor crítico positivo

En este tipo de contraste, la REGIÓN DE ACEPTACIÓN es $(-\infty, t_\alpha)$. El VALOR CRÍTICO es:

$$1-\alpha = 0.99 \to \alpha = 0.01 \to p(z < -z_\alpha) = 0.01$$

Entonces:

$$p(Z > z_\alpha) = 1 - \mathrm{p}(z \leq -z_\alpha) = 1 - 0.01 = 0.99$$

De lo cual:

$$z_\alpha = 2.3263\ (fila\ 2.3\ entre\ 0.02\ y\ 0.03)$$

Nótese que el valor exacto no se puede obtener de la tabla: se consulta que es 2.326348

c) Estadístico de contraste:

El estadístico es:

$$z = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \to z = \frac{90.09-97}{\frac{42.57}{\sqrt{108}}} = -1.6868$$

d) p-valor:

El p-valor para un contraste UNILATERAL POR LA IZQUIERDA es:

$$p-valor = P\,\underbrace{(Z \leq -1.6868)}_{no\ está\ en\ la\ tabla} = \underbrace{P\left(\underbrace{Z > 1.6868}_{no\ está\ en\ tabla}\right)}_{por\ simetría} = \underbrace{1-P(Z \leq 1.6868)}_{por\ complementariedad}$$

De lo cual:

$$p-valor = 1 - \underbrace{P(Z \leq 1.6868)}_{no\ está\ exactamente} = 1 - \underbrace{0.9545}_{\substack{Se\ toma\ 0.09\\ por\ ser\\ más\ próximo}} = 0.0455$$

e) H = 0: Aceptamos la hipótesis nula; H = 1: Aceptamos la alternativa

Si p-valor > α → Se ACEPTA H0.

En este caso, se observa:

p-valor = 0.0455 > 0.01 = α → Se ACEPTA H0

ENTREGABLE 5 – Práctica de R

Importación de los datos
Se recurre a la instrucción:

```
dades<-read.table("C:/Users/Tete/Downloads/AP.csv.csv", header=TRUE,
sep=";",na.strings="NA",
fileEncoding = "UTF-8", quote = "\"",
colClasses=c(rep("character",4),rep("numeric",2), rep("character",2)))
```

Para verificar la correcta importación de los datos, se ejecuta head(dades,3), que devuelve:

```
Cod      Country Citycode                City
1 AFG Afghanistan    40003               Herat
2 AFG Afghanistan    40001               KABUL
3 AFG Afghanistan    40002 Kandahar (Quandahar)
  Population2000 PM10Concentration1999      Region
1         323741                    46 South Asia
2        2457496                    46 South Asia
3         411752                    51 South Asia
  IncomeGroup
1  Low income
2  Low income
3  Low income
```

PREGUNTA 1
Contrastad con un nivel de significación del 10% si la concentración media de PM10 del año 1999 en las ciudades de España es superior a 40. Indicad las hipótesis nula y alternativa. A partir de la salida de R indicad el valor del estadístico de contraste, el p-valor y la conclusión a la que llegáis. Suponed que las observaciones corresponden a una muestra y que la variable considerada es normal.

Para realizar este contraste de hipótesis para la media de UNA SOLA MUESTRA se recurre a la instrucción t.test():

t.test(dades$PM10Concentration1999[datos$Country=="Spain"],h0media= 40,alternative="greater",conf.level=0.9)

Nótese que:
- Se ha denotado como "h0media" el valor de la media asumido como hipótesis nula, es decir: **que la media sea NO SEA superior a 40**.
- Se ha denotado como "alternative" el valor de la media asumido como hipótesis alternativa. En este caso: **que la media SEA superior a 40** ("greater").

Es decir:

$$\begin{cases} H_0: & \mu & \leq & 40 \\ H_1: & \mu & > & 40 \end{cases}$$

Se obtiene:

One Sample t-test
data: dades$PM10Concentration1999[datos$Country=="Spain"]
***t = 0.28709**, df = 54, p-value = 0.3876*
alternative hypothesis: true mean is greater than 40
***90 percent** confidence interval:*
38.78425 Inf
sample estimates:
mean of x
40.34545

Por tanto:
- El estadístico de contraste es: 0.28709
- El p-valor es: 0.3876
- El intervalo de confianza es: $(38.78425, \infty)$

La conclusión que se alcanza a partir del p-valor obtenido es que **SE ACEPTA la hipótesis NULA**, ya que se verifica:

$$p - valor > \alpha \leftrightarrow 0.3876 > 0.1$$

Por tanto, se puede afirmar que **la media NO ES SUPERIOR a 40** (se rechaza que sea superior a 40, que es la hipótesis alternativa) con este nivel de significación.

PREGUNTA 2

Contrastad con un nivel de significación del 5% si la concentración media de PM10 del año 1999 es más alta en las ciudades de España que en las de Francia. Indicad las hipótesis nula y alternativa. A partir de la salida de R indicad el valor del estadístico de contraste, el p-valor y la conclusión a la que llegáis. Interpretad el intervalo de confianza que os proporciona R sobre la diferencia de medias. Suponed que las observaciones corresponden a muestras y que las variables consideradas son normales y con varianzas iguales.

Se trata de un contraste de hipótesis para la diferencia de medias en 2 MUESTRAS. Se toma:

- Como hipótesis nula:
  Que LA DIFERENCIA DE MEDIAS ES 0.
- Como hipótesis alternativa:
  Que LA DIFERENCIA DE MEDIAS ES DISTINTA DE 0.

$$\begin{cases} H_0: & \mu_{España} - \mu_{Francia} & \leq & 0 \\ H_1: & \mu_{España} - \mu_{Francia} & > & 0 \end{cases}$$

Se recurre a la instrucción t.test() pero en este caso para 2 muestras:

t.test(dades$PM10Concentration1999[datos$Country=="Spain"],datos$PM10Concentration1999[datos$Country=="France"],alternative="greater",var.equal=TRUE,conf.level=0.95)

Nótese que la instrucción considera la igualdad de varianzas en ambas muestras (en el parámetro var.equal=TRUE). Se obtiene:

```
Two Sample t-test
data: dades$PM10Concentration1999[datos$Country=="Spain"] and
dades$PM10Concentration1999[datos$Country=="France"]
t = 16.124, df = 93, p-value < 2.2e-16
alternative hypothesis: true difference in mean is greater than 0.
95 percent confidence interval:
   22.48728    Inf
sample estimates:
mean of x       mean of y
    40.34545       15.27500
```

Por tanto:

- El estadístico de contraste es: 16.124
- El p-valor es: $<2.2 \cdot 10^{-16}$
- El intervalo de confianza es: $(22.48728, \infty)$

La conclusión que se alcanza a partir del p-valor obtenido es que **SE RECHAZA la hipótesis nula**, ya que se verifica:

$$p - valor < \alpha \leftrightarrow 2.2 \cdot 10^{-16} < 0.05$$

Por tanto, se **ACEPTA LA HIPÓTESIS ALTERNATIVA** y se puede afirmar que la diferencia de medias es mayor que 0 con este nivel de significación.

PREGUNTA 3

Contrastad con un nivel de significación del 1% si la proporción de ciudades correspondientes a países de ingresos bajos es menor que el 4%. Indicad las hipótesis nula y alternativa. A partir de la salida de R indicad el p-valor y la conclusión a la que llegáis. Suponed que las observaciones corresponden a una muestra.

En este caso se pide un CONTRASTE DE PROPORCIÓN.
Para realizarlo, se recurre a la instrucción prop.test().
Esta instrucción demanda como parámetros:

- El número de casos favorables (denotado arbitrariamente como f).
- El número de casos totales (denotado arbitrariamente como T).
- La proporción con la cual contrastar.

En este caso:

- f = Número de ciudades de países de ingresos bajos.
- T = Ciudades totales registradas.
- La proporción con la cual contrastar es del 4%.

En primer lugar, se almacenan los datos en las variables f y T:

f<-length(dades$City[datos$IncomeGroup=="Low income"])
T<-length(dades$City)

En segundo lugar, se escribe la instrucción del test de proporciones:

prop.test(f,T,0.04,"less",conf.level=0.99)

Se obtiene:

1-sample proportions test with continuity correction
data: f out of T, null probability 0.04
X-squared = 6.9094, df = 1, ***p-value = 0.004287***
alternative hypothesis: true p is less than 0.04
99 percent *confidence interval:*
0.00000000 0.03884373
sample estimates:
p
0.03076445

Nótese que:

- La hipótesis nula es que la proporción SEA IGUAL O MAYOR del 4%.
- La hipótesis alternativa es que la proporción sea MENOR del 4%.

Es decir:

$$\begin{cases} H_0: & p & \geq & 0.04 \\ H_1: & p & < & 0.04 \end{cases}$$

Se observa un p-valor de 0.004287
Se concluye que se **RECHAZA la hipótesis NULA** ya que se cumple:

$$p - valor < \alpha \leftrightarrow 0.004287 < 0.01$$

NO SE RECHAZA la hipótesis ALTERNATIVA según la cual la proporción de ciudades de ingresos bajos es menor del 4%, con este nivel de significación.

PREGUNTA 4

Contrastad con un nivel de significación del 5% si la proporción de ciudades correspondientes a países de ingresos bajos es diferente en los países del Sur de Asia que en los de África Sub Sahariana. Indicad las hipótesis nula y alternativa. A partir de la salida de R indicad el p-valor y la conclusión a la que llegáis. Suponed que las observaciones corresponden a una muestra.

En este caso, se pide realizar un contraste BILATERAL de proporciones, ya que se pretende conocer si ambas proporciones SON DIFERENTES.

La hipótesis nula es: Ambas proporciones son IGUALES.
La hipótesis alternativa es: Ambas proporciones son DIFERENTES.

Es decir:

$$\begin{cases} H_0: & p_{Sur\ Asia} - p_{Africa\ Subsahariana} & \leq & 0 \\ H_1: & p_{Sur\ Asia} - p_{Africa\ Subsahariana} & > & 0 \end{cases}$$

En primer lugar, como en el caso anterior, se almacenan en las variables necesarias:

- El número de ciudades del sur de Asia con ingresos bajos (f1).
- El número de ciudades del Sur de Asia (T1).
- El número de ciudades de África Sub Sahariana con ingresos bajos (f2).
- El número de ciudades de África Sub Sahariana (T2).

Se escribe:

```
f1<- length(dades$City[dades$Region=="South
Asia"&dades$IncomeGroup=="Low income"])
T1<-length(dades$City[dades$Region=="South Asia"])
f2<- length(dades$City[dades$Region=="Sub-Saharan
Africa"&datos$IncomeGroup=="Low income"])
T2<-length(dades$City[dades$Region=="Sub-Saharan Africa"])
```

Y a continuación, se realiza el test de contraste de proporciones, proporcionado los casos favorables y los totales como vectores f y T:

```
prop.test(x=c(f1,f2),n=c(T1,T2),alternative="two.sided",conf.level=0.95)
```

Se obtiene:

```
2-sample test for equality of proportions with continuity correction
data: c(f1, f2) out of c(T1, T2)
X-squared = 136.64, df = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
95% percent confidence interval:
   -0.4171941    -0.2717051
sample estimates:
    prop 1          prop 2
0.01492537       0.35937500
```

La conclusión que se alcanza a partir del p-valor obtenido es que **SE RECHAZA la hipótesis nula**, ya que se verifica:

$$p - valor < \alpha \leftrightarrow 2.2 \cdot 10^{-16} < 0.05$$

Por tanto, **SE ACEPTA LA HIPÓTESIS ALTERNATIVA** y se puede afirmar que EXISTE diferencia mayor que cero entre ambas proporciones.