TEMA 6

CONTRASTE DE HIPÓTESIS

Contenidos

1. Intervalos de confianza para una proporción.
2. Contraste de hipótesis

Tiempo estimado

5 sesiones

Criterios de Evaluación

1. Conocer el concepto de contraste de hipótesis y de nivel de significación de un contraste.
2. Saber determinar las regiones de aceptación y de rechazo de la hipótesis nula, bien para una proporción bien sobre una media y decidir, a partir de una muestra adecuada, si se rechaza o se acepta la hipótesis nula a un nivel de significación dado.

1.Hipótesis sobre la media

❑ Hipótesis sobre la media.

Vamos a partir de un ejemplo típico:

La duración de las bombillas de bajo consumo del fabricante 'Lumen' sigue una ley normal cuya desviación típica es $\sigma = 6$ meses. El fabricante afirma que la duración media de sus bombillas es de 36 meses. Ante las reiteradas quejas, una asociación de consumidores investigó la duración de las bombillas Lumen.

La aseveración que contrastamos se denomina hipótesis nula y se escribe así:

$$H_0 \ : \ \mu = 36$$

La negación la hipótesis se denomina hipótesis alternativa y se escribe así:

$$H_1 \ : \ \mu \neq 36$$

Cuando la hipótesis nula es del tipo $H_0 : \mu = \mu_0$ se dice que se trata de una hipótesis bilateral sobre la media.

Para contrastar la hipótesis nula se fijó un nivel de confianza, de la misma forma que se procede con los intervalos de confianza, $1 - \alpha = 0.99$ y se tomó una muestra aleatoria significativa de 100 bombillas, procediéndose a medir la duración de ellas.

En el contraste de hipótesis, es usual decir que el nivel de significación es α = 0.1 en lugar de el nivel de confianza es 1- α = 0.99

A la espera de la verificación, se construye un intervalo (parecido al de confianza, pero no idéntico) denominado intervalo de aceptación:

$$\left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \, , \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = (34.455 \, , 37.545)$$

La media de la muestra se denomina el estadístico de contraste: si está en el intervalo de aceptación se aceptará la hipótesis nula, y si está fuera se rechazará, aceptando entonces la hipótesis alternativa.

Bien, la media muestral resultó ser de 32 meses que, como vemos, está fuera del intervalo de aceptación (en la llamada zona de rechazo o crítica)

El hecho de que sea menor puede deberse dos hechos: o que la supuesta media es falsa o que es un sesgo estadístico de los que siempre se produce.

Por ello, a la vista de los datos se rechazó que la duración media de las bombillas fuese de 36 meses.

Aquí una tabla con todos los posibles casos que pueden presentarse en un contraste de hipótesis sobre la media de una población:

H_0	TIPO	NIVEL SIGNIFICACIÓN	INTERVALO DE ACEPTACIÓN
$\mu = \mu_0$	Bilateral	$\alpha(\%) \to z_{\alpha/2}$	$\left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \, , \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$
$\mu \geq \mu_0$	Unilateral	$\alpha(\%) \to z_\alpha$	$\left(\mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \, , +\infty\right)$
$\mu \leq \mu_0$	Unilateral	$\alpha(\%) \to z_\alpha$	$\left(-\infty \, , \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$

2.Hipótesis sobre la proporción.

Aquí los casos que podemos encontrarnos:

H_0	TIPO	NIVEL SIGNIFICACIÓN	INTERVALO DE ACEPTACIÓN
$p = p_0$	Bilateral	$\alpha(\%) \to z_{\alpha/2}$	$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \, , p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\right)$
$p \leq p_0$	Unilateral	$\alpha(\%) \to z_\alpha$	$\left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \, , +\infty\right)$
$p \geq p_0$	Unilateral	$\alpha(\%) \to z_\alpha$	$\left(-\infty \, , p_0 + z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\right)$

☞ Ejemplo: Un estudio sociológico afirma que al menos 70% de las familias cena viendo la televisión. Para contrastar la veracidad de esta afirmación se toma una muestra de 500 familias, en la que se observa que 340 ven la televisión mientras cenan. Con un nivel de significación de 0,01 vamos a decidir, mediante un contraste de hipótesis, si la afirmación es cierta.

Hipótesis (unilateral sobre la proporción):

$$H_0 \,:\, p \geq 0.70 = p_0 \quad \to \quad H_1 \,:\, p < 0.70$$

Es una hipótesis unilateral sobre la proporción.

Muestra y estadístico:

Tamaño muestral: $n = 500$

Proporción muestral: $\widetilde{p} = \frac{340}{500} = 0.48$

Significación y nivel crítico:

Nivel de significación: $\alpha = 0.01$ ➡ Valor crítico: $z_\alpha = 2.33$

$$\mathrm{p}\,[z > z_\alpha] = 0.01 \to \mathrm{p}\,[z < z_\alpha] = 0.99 \xrightarrow{tabla} z_\alpha = 2.33$$

Intervalo de aceptación:

$$I = \left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 \cdot q_0}{n}} \, , +\infty\right) = (0.6522 \, , +\infty)$$

Conclusión:

$$\widetilde{p} \in I \to \text{ Aceptamos } H_0$$

Con estos datos no podemos rechazar, a dicho nivel, que al menos 70% de las familias cena viendo la televisión.

3.Errores.

Podemos cometer dos tipos de errores en un contraste de hipótesis:

Rechazar una hipótesis verdadera (error de tipo I): la probabilidad de cometerlo es α y no depende del tamaño muestral.

Aceptar una hipótesis falsa (error de tipo II): la probabilidad de cometerlo depende del verdadero valor del parámetro y disminuye al aumentar n.

Ejercicios

1. [S/11] Una máquina está preparada para fabricar piezas de, a lo sumo, 10 cm de longitud. Se toma una muestra de 1000 piezas, comprobándose que la media sus longitudes es de 10.0037 cm. La longitud de las piezas fabricadas por esa máquina sigue una ley Normal con desviación típica 0.2 cm.

a) Plantee un contraste de hipótesis unilateral para comprobar si con los datos de esa muestra es posible afirmar que la media de la longitud de las piezas fabricadas por la máquina es de más de 10 cm.

b) Determine la región de aceptación de la hipótesis nula de ese contraste para un nivel de significación $\alpha = 0.025$.

c) Con los datos de la muestra y usando el contraste de hipótesis de (a), ¿qué conclusión se obtendría sobre la longitud media de las piezas fabricadas?

2. [S/11] El director de una televisión afirma que un nuevo programa que va a emitirse será visto, al menos, por un 30% de personas. Una vez emitido se realizó una encuesta a 500 personas, elegidas al azar, y ésta reveló que 130 de ellas habían visto ese programa.

a) Formule la hipótesis nula y la alternativa del contraste de hipótesis que permite determinar si los datos de la encuesta realizada son compatibles con la afirmación del director.

b) Halle la región crítica de ese contraste para un nivel de significación del 5.5%.

c) Según el dato obtenido en el apartado anterior ¿qué conclusión se obtiene sobre la afirmación realizada por el director de esa televisión?

3. [S/11] El director de un banco afirma que la cantidad media de dinero extraído, por cliente, de un cajero automático de su sucursal no supera los 120 euros. Para contrastar esta hipótesis elige al azar 100 extracciones de este cajero y obtiene una media muestral de 130 euros. Se sabe que la cantidad de dinero extraído por un cliente en un cajero automático se distribuye según una ley Normal de media desconocida y desviación típica 67 euros.

a) Plantee el contraste de hipótesis asociado.

b) Determine la región de aceptación, para un nivel de significación $\alpha = 0.05$.

c) Con los datos muestrales tomados, ¿existe evidencia estadística para rechazar la hipótesis de este director, con el mismo nivel de significación anterior?

4. [S/11] Suponiendo que la variable "años de vida de los individuos de un país" sigue una distribución Normal con desviación típica 8.9 años, se desea contrastar la hipótesis de que la vida media de los mismos no supera los 70 años.

A partir de una muestra aleatoria de 100 individuos se ha obtenido que su vida media ha sido 71,8 años.

a) Formule el contraste de hipótesis que indica el enunciado.

b) Determine la región crítica a un nivel de significación del 5%.

c) Con los datos muestrales, ¿existe evidencia estadística para rechazar la hipótesis a ese nivel de significación?

5. [S/11] Un estudio sociológico afirma que el 70% de las familias cena viendo la televisión. Se desea contrastar la veracidad de esta afirmación y, para ello, se toma una muestra de 500 familias, en la que se observa que 340 ven la televisión mientras cenan.

Decida, mediante un contraste de hipótesis, si la afirmación es cierta con un nivel de significación de 0.01.

6. Un laboratorio farmacéutico fabrica un producto para la caída del cabello que envasa en botes, y en el etiquetado indica que su contenido aproximado es de 100 cc. Se sabe que el contenido es una variable aleatoria normal de desviación típica 2 c.c.

Se eligen, al azar, 7 de estos botes y se miden sus contenidos dando el siguiente resultado (en cc):

97 ; 101 ; 102 ; 99 ; 98 ; 100 ; 103

¿Podemos asegurar que la capacidad media de los botes que se fabrican es la indicada en el bote? (Úsese $\alpha = 0.01$)

7. Un constructor afirma que por lo menos el 75 % de las casas que construye tienen calefacción. ¿Se estaría de acuerdo con tal afirmación si una inspección aleatoria muestra que 72 de 135 casas cuentan con calefacción? (Usar $\alpha = 0.1$)

8. En una ciudad, donde la proporción de fumadores con edad comprendida entre 18 y 20 años es del 30%, el ayuntamiento ha realizado una campaña contra el consumo de tabaco.

Dos meses después de terminar dicha campaña, se ha realizado una encuesta a 400 personas de estas edades, elegidas al azar, y se ha encontrado entre ellos a 92 fumadores.

¿Podemos afirmar, a un nivel de significación $\alpha = 0.05$, que esta campaña ha modificado la proporción de fumadores entre 18 y 25 años?

9. La talla de los individuos de una población sigue una distribución normal de desviación típica 8 cm. Se han determinado las tallas de 25 individuos, encontrándose una media de 168 cm. ¿Se podrá afirmar que la talla media de la población es menor de 170 cm? (Usar $\alpha = 0.03$)

Cuestiones

Autoevaluación

1. La altura en cm. de las cañas producidas por una variedad de carrizo en cada cosecha es una variable aleatoria que sigue una ley normal con desviación típica $\sigma = 16$ cm.

Para contrastar si la altura media de las cañas de la ultima cosecha es de 170 cm, se ha tomado una muestra aleatoria de 64 de estas cañas y se han medido sus longitudes, resultando como media muestral $\bar{x}=166$ cm.

¿Son suficientes estos datos para rechazar que la altura media de las cañas de la ultima cosecha es de 170 cm, a un nivel de significación $\alpha = 0.05$?

2. El peso en vacío de los envases fabricados por una empresa, según su método usual, es una variable aleatoria que sigue una ley normal con media 20 gramos y una desviación típica de 1 gramo.

Se desea contrastar si un nuevo proceso de fabricación no aumenta dicho peso medio. Para ello, se eligen al azar 25 envases fabricados por la nueva técnica y se encuentra que la media de su peso en vacío es de 20.5 gramos.

¿Se puede afirmar, a un nivel de significación α = 0,02, que el nuevo proceso ha aumentado el peso medio de los envases?

3. Una compañía textil afirma que a lo sumo el 20 % del público compra ropa de lana.

Verifica esta afirmación para $\alpha = 0.01$, si una encuesta aleatoria indica que 46 de 200 clientes compran ropa de lana.

4. Las autoridades educativas publican en un estudio que el 25 % de los estudiantes de Bachillerato de una cierta comunidad autónoma tienen ordenador portátil.

A partir de una muestra aleatoria de tamaño 300 se ha obtenido que sólo 70 de ellos tienen ordenador portátil.

¿Se podrá asegurar que las autoridades dicen la verdad? (Usar $\alpha = 0.06$).

Autoevaluación

1. En el contexto del problema, queremos saber si "la altura media de las cañas es 170 cm."

Hipótesis:

$H_0 : \mu = 170 = \mu_0 \quad [\sigma = 16] \quad \rightarrow \quad H_1 : \mu \neq 170$

Es bilateral sobre la media.

Muestra y estadístico:

Tamaño muestral: $n = 64$

Media muestral: $\bar{x} = 166$

A) Significación y nivel crítico:

Significación: $\alpha = 0{,}05$ → Valor crítico: $z_{\alpha/2} = 1{,}96$

$$p(z > z_{\alpha/2}) = 0{,}975 \rightarrow z_{\alpha/2} = 1{,}96$$

B) Intervalo de aceptación:

$$I = \left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \ , \ \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = (166{,}08 \, , 173{,}92)$$

C) Conclusión:

$$\bar{x} \notin I \rightarrow \text{Rechazamos } H_0$$

A la vista de los datos podemos rechazar, a dicho nivel, que la altura media de las cañas es 170 cm.

2. En el contexto del problema, queremos saber si "no ha aumentado el peso medio de los envases."

Hipótesis:

$H_0 : \mu \leq 20 = \mu_0 \ [\sigma = 1] \quad \rightarrow \quad H_1 : \mu > 20 = \mu_0$

Es unilateral sobre la media.

Muestra y estadístico:

Tamaño muestral: $n = 25$

Media muestral: $\bar{x} = 20.5$

Significación y nivel crítico:

Significación: $\alpha = 0.02$ → Valor crítico: $z_\alpha = 2.05$

$$\mathrm{p}\,[z > z_\alpha] = 0.02 \rightarrow \mathrm{p}\,[z < z_\alpha] = 0.98 \xrightarrow{tabla} z_\alpha = 2.05$$

Intervalo de aceptación:

$$I = \left(-\infty \, , \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = (-\infty \, , 20.41)$$

Conclusión:

$$\bar{x} \notin I \rightarrow \text{ Rechazamos } H_0$$

3. En el contexto del problema, queremos saber si "a lo sumo el 20% compra ropa de lana"

Hipótesis:

$H_0 : p \leq 0{,}20 = p_0 \quad \rightarrow \quad H_1 : p > 0{,}20$

Es unilateral sobre la proporción.

Muestra y estadístico:

Tamaño muestral: $n = 200$

Proporción muestral: $\tilde{p} = \dfrac{46}{200} = 0{,}23$

Significación y nivel crítico:

Significación $\alpha = 0{,}01$ → Valor crítico: $z_\alpha \approx 2{,}33$

$$p(z > z_\alpha) = 0{,}01 \rightarrow p(z < z_\alpha) = 0{,}99 \rightarrow z_\alpha \approx 2{,}33$$

Intervalo de aceptación:

$$I = \left(-\infty \ , \ p_0 + z_\alpha \cdot \sqrt{\frac{p_0 q_0}{n}}\right) = (-\infty \, , 0{,}2659)$$

Conclusión:

$$\tilde{x} \in I \rightarrow \text{ Aceptamos } H_0$$

Con estos datos no podemos rechazar, a dicho nivel, que a lo sumo el 20% compra ropa de lana.

4. En el contexto del problema, queremos saber si "el 25% de los estudiantes tiene portátil"

Hipótesis:

$H_0 : p = 0{,}25 = p_0 \quad \rightarrow \quad H_1 : p \neq 0{,}25$

Es bilateral sobre la proporción.

Muestra y estadístico:

Tamaño muestral: $n = 300$

Proporción muestral: $\tilde{p} = \dfrac{70}{300} = 0{,}2333\ldots$

D) Nivel de significación y nivel crítico:

Significación $\alpha = 0{,}06$ →Valor crítico: $z_{\alpha/2} = 1{,}88$

$$p(z > z_{\alpha/2}) = 0{,}06/2 = 0{,}03 \rightarrow p(z < z_{\alpha/2}) = 0{,}97 \rightarrow z_{\alpha/2} = 1{,}88$$

Intervalo de aceptación:

$$I = \left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 q_0}{n}} \ , \ p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 q_0}{n}}\right) = (0{,}203 \, , 0{,}297)$$

<u>Conclusión</u>:

$$\tilde{p} \in I \;\rightarrow\; \text{Aceptamos } \; H_0$$

Con estos datos no podemos rechazar, a dicho nivel, que el 25% de los estudiantes tenga un portátil.