Contrastes de hipótesis

Clase del 19/11/2024

Contenidos

Ejemplo

En una población normal con varianza conocida $\sigma^2 = 4$, se extrajo una muestra de tamaño $n = 16$, obteniéndose una media muestral de $\bar{x} = 6$. Queremos contrastar la hipótesis de que la media poblacional es $\mu = 5$ frente a la hipótesis alternativa de que $\mu > 5$, utilizando un nivel de significación de $\alpha = 0.05$.

Formulación de las Hipótesis

$$H_0 : \mu = 5 \quad \text{(la media poblacional es 5).}$$

$$H_a : \mu > 5 \quad \text{(la media poblacional es mayor que 5).}$$

Estadístico de Prueba

Dado que la población sigue una distribución normal y se conoce la varianza, utilizamos la z-prueba:

$$d = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Sustituyendo los valores:

$$\bar{x} = 6, \quad \mu_0 = 5, \quad \sigma = \sqrt{4} = 2, \quad n = 16,$$

obtenemos:

$$d = \frac{6-5}{2/\sqrt{16}} = \frac{6-5}{2/4} = \frac{1}{0.5} = 2.$$

Nivel de Significación y Región Crítica

Para un nivel de significación de $\alpha = 0.05$
(Probabilidad de rechazar H_0 cuando es verdadera)
el valor crítico $z_{\text{crítico}}$ se obtiene de la tabla de la distribución normal estándar: $z_{\text{crítico}} = 1.645$
La regla de decisión es la siguiente:

- Si $d \geq z_{\text{crítico}}$, rechazamos H_0.
- Si $d < z_{\text{crítico}}$, no rechazamos H_0.

Comparación del Estadístico Calculado con el Valor Crítico

El estadístico calculado es: $d = 2$. Por tanto como $d = 2 > z_{\text{crítico}} = 1.645$ entonces **rechazamos** H_0.
Equivalentemente, rechazamos H_0 si:

$$\bar{x} \geq \frac{\sigma}{\sqrt{n}} \cdot z_{crtico} + \mu_0$$

Alternativamente podemos razonar calculando el **nivel crítico**:

$$P(d > 2|N(0,1) = 0.023$$

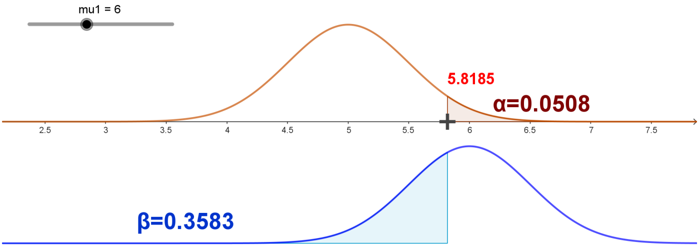
Como el nivel de significación (0.05) es mayor que el nivel crítico (0.023) , entonces rechazamos la hipótesis.

error de tipo II

El error de tipo II β viene dado por la probabilidad de no rechazar H_0 cuando es falsa. Supongamos, por ejemplo, que $\mu = 6$.
Tendríamos que $\bar{X} \sim N(6, 0.5)$.
$P(d \leq 1,65|\mu = 6) = P(z \leq -0.35) = 0.363$

Resultado del contraste

Con un nivel de significación de $\alpha = 0.05$, los datos proporcionan suficiente evidencia para concluir que la media poblacional es mayor que 5.


mu1 = 6
5.8185
α=0.0508
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
β=0.3583

Introducción

El contraste de hipótesis es una herramienta esencial en la inferencia estadística para tomar decisiones sobre una población basándose en información muestral. Podríamos establecer el paralelismo con la celebración de un juicio. Hay que establecer la culpabilidad de alguien y hay que aportar suficientes evidencias para determinar si es así o si nos quedamos con la hipótesis inicial de la inocencia.

Hipótesis Nula (H_0) y Alternativa (H_a)

- **Hipótesis nula (H_0):** Representa la afirmación inicial o el estado "por defecto". Usualmente indica ausencia de efecto o diferencia (por ejemplo, $\mu = 0$, $p = 0.5$).
- **Hipótesis alternativa (H_a):** Es la afirmación que se desea probar. Puede ser:
  - *Unilateral*: Por ejemplo, $\mu > 0$ o $\mu < 0$.
  - *Bilateral*: Por ejemplo, $\mu \neq 0$.

Estadístico de Prueba

Es una función de los datos muestrales que resume la información necesaria para el contraste. Su forma depende del tipo de prueba empleada, como t-prueba, z-prueba o χ^2-prueba. Este estadístico se compara con una distribución teórica bajo la suposición de que H_0 es verdadera.

Nivel de Significación (α)

Es la probabilidad máxima tolerada de cometer un **error tipo I** (rechazar H_0 cuando es verdadera). Los valores más comunes son $\alpha = 0.05$ o $\alpha = 0.01$.

Regla de Decisión

- **Valor p:** Es la probabilidad de obtener un resultado tan extremo como el observado, bajo la suposición de H_0. Si $p < \alpha$, se rechaza H_0.
- **Valor Crítico:** Se define un rango crítico en función de α y la distribución del estadístico. Si el estadístico cae dentro del rango crítico, se rechaza H_0.

Errores en la Decisión

- **Error tipo I (α):** Rechazar H_0 cuando es verdadera.
- **Error tipo II (β):** No rechazar H_0 cuando es falsa. verdadera.

Distribución bajo la Hipótesis Nula

Es la distribución teórica del estadístico de prueba cuando H_0 es verdadera. Por ejemplo, para pruebas sobre medias:

- z-distribución (si la varianza poblacional es conocida).
- t-distribución (si la varianza poblacional es desconocida y la muestra es pequeña).

Datos Muestrales

Los datos son la base para calcular el estadístico de prueba. Deben cumplir ciertos supuestos, como normalidad, independencia y homogeneidad de varianzas (dependiendo del contraste).

Interpretación del Resultado

- Si se **rechaza** H_0, se concluye que los datos ofrecen evidencia suficiente para respaldar H_a.
- Si **no se rechaza** H_0, no hay evidencia suficiente para descartar H_0, pero esto no implica que sea verdadera.

Contraste sobre proporciones: Formulación de las hipótesis

Caso bilateral

$$H_0 : p = p_0$$
$$H_a : p \neq p_0$$

Estadístico del contraste a un nivel de significación α

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Donde:

- $\hat{p} = \frac{x}{n}$: Proporción muestral, con x el número de éxitos en la muestra y n el tamaño muestral.
- p_0: Proporción teórica según H_0.
- n: Tamaño de la muestra.

Región crítica o regla de decisión

Se rechaza H_0 si $|Z| > Z_{\alpha/2}$

Contraste unilateral sobre proporciones

$$H_0 : p \leq p_0$$

$$H_a : p > p_0$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Se rechaza H_0 si $Z > Z_{\alpha/2}$

Contraste unilateral 2 sobre proporciones

$$H_0 : p \geq p_0$$

$$H_a : p < p_0$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Se rechaza H_0 si $Z < Z_{\alpha/2}$

Contraste sobre diferencia de proporciones

En esta ocasión se dispone de dos muestras de sendas poblaciones. A partir de ahí se pretende estudiar si la proporción de una determinada característica de una población es igual, mayor o menor que la de la otra población.

Contraste sobre diferencia de proporciones

$$\left.\begin{array}{lcl} H_0 & : & p_x = p_y \\ H_a & : & p_x \neq p_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\left| \frac{\hat{p}_x - \hat{p}_y}{\sqrt{p(1-p)(\frac{1}{n_x} + \frac{1}{n_y})}} \right| > Z_\alpha$$

$$p = \frac{n_x p_x + n_y p_y}{n_x + n_y}$$

Contraste sobre diferencia de proporciones (unilateral)

$$\left.\begin{array}{lcl} H_0 & : & p_x \leq p_y \\ H_a & : & p_x > p_y \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\hat{p}_x - \hat{p}_y}{\sqrt{p(1-p)(\frac{1}{n_x} + \frac{1}{n_y})}} > Z_\alpha$$

CH para la media de poblaciones normales. Varianza conocida

$$\left.\begin{array}{lcl} H_0 & : & \mu = \mu_0 \\ H_a & : & \mu \neq \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\left|\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}\right| > Z_{\frac{\alpha}{2}}$$

CH para la media de poblaciones normales.Varianza conocida. (Unilateral)

$$\left.\begin{array}{lcl} H_0 & : & \mu \leq \mu_0 \\ H_a & : & \mu > \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > Z_{\frac{\alpha}{2}}$$

CH para la media de poblaciones normales.Varianza conocida. (Unilateral 2)

$$\left.\begin{array}{lcl} H_0 & : & \mu \geq \mu_0 \\ H_a & : & \mu < \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} < -Z_{\frac{\alpha}{2}}$$

CH para la media de poblaciones normales. Varianza desconocida

$$\left.\begin{array}{lcl} H_0 & : & \mu = \mu_0 \\ H_a & : & \mu \neq \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\S} \right| > t_{n-1,\frac{\alpha}{2}}$$

CH para la media de poblaciones normales.Varianza desconocida. (Unilateral)

$$\left.\begin{array}{lcl} H_0 & : & \mu \leq \mu_0 \\ H_a & : & \mu > \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\S} > t_{n-1,\frac{\alpha}{2}}$$

CH para la media de poblaciones normales.Varianza desconocida. (Unilateral 2)

$$\left.\begin{array}{lcl} H_0 & : & \mu \geq \mu_0 \\ H_a & : & \mu < \mu_0 \end{array}\right\}$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{\S} < t_{n-1,\frac{\alpha}{2}}$$

Contraste para la varianza

$$\left.\begin{array}{lcl} H_0 & : & \sigma^2 = \sigma_0^2 \\ H_a & : & \sigma^2 \neq \sigma_0^2 \end{array}\right\}$$

No se rechaza H_0 si:

$$\chi^2_{n-1,1-\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1,\frac{\alpha}{2}}$$

Contraste para la varianza (unilateral1)

$$\left.\begin{array}{lcl} H_0 & : & \sigma^2 \leq \sigma_0^2 \\ H_a & : & \sigma^2 > \sigma_0^2 \end{array}\right\}$$

No se rechaza H_0 si:

$$\frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1,\alpha}$$

Contraste para la varianza (unilateral 2)

$$\left.\begin{array}{lcl} H_0 & : & \sigma^2 \geq \sigma_0^2 \\ H_a & : & \sigma^2 < \sigma_0^2 \end{array}\right\}$$

No se rechaza H_0 si:

$$\chi^2_{n-1,1-\alpha} < \frac{(n-1)S^2}{\sigma_0^2}$$

Contraste para el cociente de varianzas de dos poblaciones normales

$$\left.\begin{array}{lcl} H_0 & : & \sigma_x^2 = \sigma_y^2 \\ H_a & : & \sigma_x^2 \neq \sigma_y^2 \end{array}\right\}$$

No se rechaza H_0 si:

$$F_{n_x-1,n_y-1,1-\frac{\alpha}{2}} < \frac{S_x^2}{S_y^2} < F_{n_x-1,n_y-1,\frac{\alpha}{2}}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas conocidas

$$\left.\begin{array}{lcl} H_0 & : & \mu_x = \mu_y \\ H_a & : & \mu_x \neq \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \right| > Z_{\frac{\alpha}{2}}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas conocidas.

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \leq \mu_y \\ H_a & : & \mu_x > \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} > Z_\alpha$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas conocidas.

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \geq \mu_y \\ H_a & : & \mu_x < \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} < -Z_\alpha$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas e iguales

$$\left.\begin{array}{lcl} H_0 & : & \mu_x = \mu_y \\ H_a & : & \mu_x \neq \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\left| \frac{\bar{X} - \bar{Y}}{S_d \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \right| > t_{n_x+n_y-2, \frac{\alpha}{2}}$$

$$S_d^2 = \frac{(n_x-1)\cdot S_x^2 + (n_y-1)\cdot S_y^2}{n_x+n_y-2}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas e iguales

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \leq \mu_y \\ H_a & : & \mu_x > \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{S_d \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} > t_{n_x + n_y - 2, \alpha}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas e iguales

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \geq \mu_y \\ H_a & : & \mu_x < \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{S_d \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < -t_{n_x+n_y-2,\alpha}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas y distintas

$$\left.\begin{array}{lcl} H_0 & : & \mu_x = \mu_y \\ H_a & : & \mu_x \neq \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{X_x^2}{n_x} + \frac{S_y^2}{n_y}}} \right| > t_{g,\frac{\alpha}{2}}$$

$g = n_1 + n_2 - 2 - \Delta$, gr. de libertad
$\Delta =$ es el entero más próximo a $\frac{((n_y-1)S_x-(n_x-1)S_y)^2}{(n_y-1)S_x^2+(n_x-1)S_y^2}$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas y distintas

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \leq \mu_y \\ H_a & : & \mu_x > \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} > t_{g,\alpha}$$

Contraste sobre diferencia de medias en dos poblaciones normales. Varianzas desconocidas y distintas

$$\left.\begin{array}{lcl} H_0 & : & \mu_x \geq \mu_y \\ H_a & : & \mu_x < \mu_y \end{array}\right\} \equiv$$

Se rechaza H_0 al nivel de significación α si

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} < -t_{g,\alpha}$$

En los contrastes de hipótesis sobre dos poblaciones normales, se asume que son independientes. Si las variables aleatorias son dependientes o apareadas, se considera la variable aleatoria $D = X - Y \sim N(\mu_x - \mu_y, \sigma_D)$