

Pruebas de hipótesis

Mg. Ciro Ivan Machacuay Meza

Facultad de Economía

9 de marzo de 2025

Índice de contenidos

- 1 Introducción a las pruebas de hipótesis
- 2 Prueba de hipótesis de media con σ^2 conocida
- 3 Prueba de hipótesis de media con σ^2 desconocida
- 4 Prueba de hipótesis de la varianza poblacional
- 5 Prueba de hipótesis para la proporción

Introducción a las pruebas de hipótesis

Se denomina hipótesis estadística a cualquier afirmación o conjetura que se hace acerca de la distribución de una o más poblaciones. Esta puede referirse bien a la forma o tipo de distribución de probabilidad de la población en estudio o bien referirse al valor o valores de uno o más parámetros de la distribución conocida su forma.

Algunos ejemplos de hipótesis estadísticas son :

- La longitud media de un tipo de objetos es 10 centímetros.
- La proporción de objetos defectuosos producidos por ciertos procesos es superior al 8 %.
- La varianza de los contenidos de un producto que se comercializa en bolsas de 250 g es $0.25 g^2$.
- Son iguales las medias de dos tipos de mediciones independientes X e Y que se distribuyen normalmente con varianzas desconocidas supuestas iguales a σ^2 .

Se denomina **hipótesis simple** a cualquier hipótesis estadística que especifique un valor del parámetro. Por ejemplo, afirmar que $\mu = 500$.

Si la hipótesis no indica un valor específico del parámetro, se dice que es una **hipótesis compuesta**. Por ejemplo, afirmar que $\mu > 70$.

Se denomina **hipótesis nula** (H_0) a la hipótesis que es aceptada provisionalmente como verdadera y cuya validez será sometida a comprobación experimental. Por tanto, es la hipótesis principal a probar. Los resultados experimentales nos permitirán seguir aceptándola como verdadera o si, por el contrario debemos rechazarla como tal.

Se denomina **hipótesis alternativa** (H_1) a la hipótesis contraria a la hipótesis nula. Se acepta en caso de que la hipótesis nula sea rechazada. Por ejemplo, son H_0 y H_1 acerca del parámetro μ de la población normal $N(\mu, 9)$:

$$H_0 : \mu = 70 \quad H_1 : \mu \neq 70$$

$$H_0 : \mu \leq 70 \quad H_1 : \mu > 70$$

$$H_0 : \mu \geq 70 \quad H_1 : \mu < 70$$

Se prueba la hipótesis nula H_0 en contraposición de la hipótesis alterna H_1 . Por lo tanto, el primer paso es plantear las dos suposiciones contrarias, H_0 contra H_1 .

Se supone que H_0 es verdadera. Esta suposición nos lleva a obtener un intervalo de los valores del parámetro en donde se toma la decisión de aceptar o rechazar la hipótesis nula H_0 aplicando los resultados de una muestra al azar escogida de la población en estudio.

Si se acepta H_0 , entonces se rechaza H_1 y si se rechaza H_0 , entonces se acepta H_1 .

Puede ser que H_0 no sea realmente verdadera, en este caso, si la decisión es aceptarla, estaríamos cometiendo un error. Por otro lado, puede ser que H_0 sea realmente verdadera, en este caso, si la decisión es rechazarla, estaríamos cometiendo otro error.

Hay cuatro posibles acciones o situaciones que determinan si la decisión tomada es correcta o incorrecta. Dos son decisiones acertadas y otras dos son decisiones erradas.

Decisión / Hipótesis	H_0 verdadera	H_0 falsa
Rechazar H_0	Error tipo I <i>Probabilidad = α</i>	Decisión correcta <i>Probabilidad = $1 - \beta$</i>
Aceptar H_0	Decisión correcta <i>Probabilidad = $1 - \alpha$</i>	Error tipo II <i>Probabilidad = β</i>

Error tipo I

Es el error que se comete al tomar la decisión de rechazar la hipótesis nula H_0 cuando realmente es verdadera. La probabilidad de cometer el error tipo I se denota por α . Si se acepta H_0 cuando realmente es verdadera entonces ocurre con una probabilidad $1 - \alpha$. Esto se denomina **nivel de significancia de la prueba**.

Si para un valor dado de α se rechaza la hipótesis nula H_0 , entonces se dice que los resultados muestrales obtenidos, no solo son diferentes por efectos del azar, sino que son realmente significativamente diferentes al nivel $\alpha \times 100\%$.

Error tipo II

Es el error que se comete al tomar la decisión de aceptar la hipótesis nula H_0 cuando realmente es falsa. Rechazar H_0 cuando ésta en realidad es falsa es una decisión correcta y ocurre con probabilidad $1 - \beta$. La probabilidad β se denomina **potencia de la prueba**.

Con los resultados de una muestra aleatoria de tamaño n seleccionada de la población en estudio se comprueba que si α aumenta, entonces β disminuye, y si β aumenta, entonces α disminuye. En todo proceso de toma de decisiones sobre hipótesis estadísticas, es deseable disminuir las probabilidades de cometer esos dos tipos de errores.

Después de plantear la prueba de hipótesis nula H_0 contra su correspondiente H_1 y especificado la probabilidad de error tipo I o el nivel de significación α , se deberá determinar en la distribución de la estadística correspondiente al parámetro el intervalo donde se rechaza H_0 . Este intervalo se denomina **región de rechazo o región crítica (RC)** de H_0 y satisface la condición $Prob(RC) = \alpha$.

Una vez determinada la región RC se aplica la siguiente **regla de decisión**: Si la estadística especificada por H_0 y calculada de la muestra se ubica en el intervalo RC se rechazaría H_0 . En caso contrario, no se rechazaría H_0 . El nivel de significación va a generar la división de la distribución muestral de la estadística de prueba en dos partes mutuamente excluyentes. La **región crítica** (RC) de rechazo de H_0 y la **región de aceptación** (RA) o *no rechazo* de H_0 . La hipótesis alternativa H_1 indicará cómo realizar esta división.

El procedimiento se resume en tres pasos concretos:

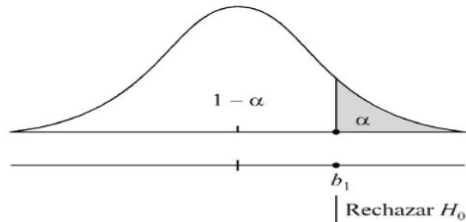
- **Hipótesis:** Plantear adecuadamente la hipótesis nula H_0 contra la hipótesis alternativa H_1 .
- **Estadística y región crítica:** Especificar la estadística apropiada de la prueba. Luego, con el valor dado del nivel de significación α y el tipo de contraste que indica H_1 , hallar la región crítica RC en la que se aplica la regla de decisión.
- **Decisión:** Calcular el valor de la estadística de la prueba aplicando los datos de la muestra y tomar la decisión de rechazar la hipótesis H_0 si la estadística calculada está en la región crítica.

Si la población X es normal (PRN) o no normal (TLC) se define:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

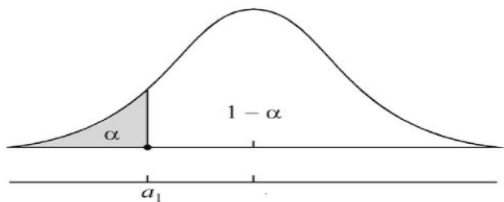
1) Prueba unilateral de cola a la derecha

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$, dado el nivel de significación α y dado que H_1 indica cola derecha en la distribución de Z especificada por H_0 supuesta verdadera, se hallan los valores críticos $z_{1-\alpha}$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P[Z > z_{1-\alpha}] = \alpha$.



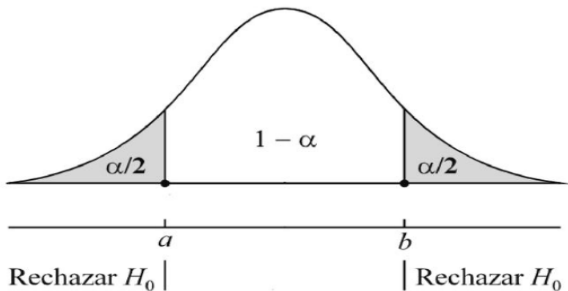
2) Prueba unilateral de cola a la izquierda

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$, dado el nivel de significación α y dado que H_1 indica cola izquierda en la distribución de Z especificada por H_0 supuesta verdadera, se hallan los valores críticos $z_{1-\alpha}$, tal que la probabilidad de rechazar H_0 si es verdadera sea $P[Z < -z_{1-\alpha}] = \alpha$.



3) Prueba bilateral o de dos colas

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$, dado el nivel de significación α y dado que H_1 indica dos colas, en la distribución del estadístico Z especificada por H_0 supuesta verdadera, se hallan los valores críticos $\pm z_{1-\frac{\alpha}{2}}$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P\left[Z < -z_{1-\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$ o $P\left[Z > z_{1-\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$.



Otro método de la toma de decisión en prueba de hipótesis es aplicando la probabilidad p que se obtiene de la estadística calculada de la muestra Z :

$$P = 2P[Z > Z_{calc}]$$

$$P = P[Z > Z_{calc}]$$

La primera prueba corresponde para una prueba bilateral y la segunda a una prueba unilateral. La regla de decisión es: Si p -valor es mayor a α , se acepta H_0 , caso contrario, se rechaza H_0 .

Prueba de hipótesis de media con σ^2 desconocida

Sean X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n escogida de una población definida por la variable X cuya distribución de probabilidades es normal $N(\mu, \sigma^2)$ con ambos parámetros desconocidos. Sean la media y la varianza muestral respectivas:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \qquad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Si la población es normal y con varianza desconocida, se tiene que:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

En esta distribución T se determina la **región crítica** (RC) de la prueba cuya probabilidad sea igual al nivel de significación α . De la muestra se calcula:

$$t_{calc} = \frac{\bar{x} - \mu_0}{ET} \rightarrow ET = \frac{s}{\sqrt{n}}$$

Donde ET denota el error típico de la media muestral.

1) Prueba bilateral o de dos colas

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$, dado el nivel de significación α y dado que H_1 indica dos colas, en la distribución del estadístico t_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $\pm t_{1-\frac{\alpha}{2}, n-1}$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P\left[T < -t_{1-\frac{\alpha}{2}, n-1}\right] = \frac{\alpha}{2}$ o $P\left[T > t_{1-\frac{\alpha}{2}, n-1}\right] = \frac{\alpha}{2}$. Luego, la región de rechazo de H_0 en el rango de variación de T es el intervalo:

$$RC = \left\{ T < -t_{1-\frac{\alpha}{2}, n-1} \quad \text{o} \quad T > t_{1-\frac{\alpha}{2}, n-1} \right\}$$

2) Prueba unilateral de cola a la derecha

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$, dado el nivel de significación α y dado que H_1 indica cola derecha en la distribución de t_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $t_{1-\alpha, n-1}$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P\left[T > t_{1-\alpha, n-1}\right] = \alpha$. Luego, la región de rechazo de H_0 en el rango de variación de T es el intervalo:

$$RC = \{T > t_{1-\alpha, n-1}\}$$

3) Prueba unilateral de cola a la izquierda

Si se prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$, dado el nivel de significación α y dado que H_1 indica cola izquierda en la distribución de t_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $t_{1-\alpha, n-1}$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P[T < -t_{1-\alpha, n-1}] = \alpha$. Luego, la región de rechazo de H_0 en el rango de variación de T es el intervalo:

$$RC = \{T < -t_{1-\alpha, n-1}\}$$

Prueba de hipótesis de la varianza poblacional

Sean X_1, X_2, \dots, X_n una muestra aleatoria de tamaño n escogida de una población definida por la variable X cuya distribución de probabilidades es normal $N(\mu, \sigma^2)$ con ambos parámetros desconocidos.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Esta estadística se utiliza para obtener la región crítica de la prueba de hipótesis acerca de una varianza. Si la hipótesis nula $H_0 : \sigma^2 = \sigma_0^2$ es supuesta verdadera, entonces la estadística de prueba de H_0 contra cualquier alternativa unilateral o bilateral es:

$$\chi^2 = \frac{(n - 1) \times S^2}{\sigma^2} \sim \chi_{n-1}^2$$

En esta distribución especificada por H_0 se determina la **región crítica** (RC) de la prueba cuya probabilidad sea igual al nivel de significación α . De la muestra se calcula:

$$\chi_{calc}^2 = \frac{(n - 1) \times S^2}{\sigma^2}$$

1) Prueba bilateral o de dos colas

Si se prueba $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$, dado el nivel de significación α y dado que H_1 indica dos colas, en la distribución del estadístico χ_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $\chi_{\frac{\alpha}{2}, n-1}^2$ y $\chi_{1-\frac{\alpha}{2}, n-1}^2$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P\left[\chi^2 < \chi_{\frac{\alpha}{2}, n-1}^2\right] = \frac{\alpha}{2}$ o $P\left[\chi^2 > \chi_{1-\frac{\alpha}{2}, n-1}^2\right] = \frac{\alpha}{2}$. Luego, la región de rechazo de H_0 es el intervalo:

$$RC = \left\{ \chi^2 < \chi_{\frac{\alpha}{2}, n-1}^2 \quad \text{o} \quad \chi^2 > \chi_{1-\frac{\alpha}{2}, n-1}^2 \right\}$$

2) Prueba unilateral de cola a la derecha

Si se prueba $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 > \sigma_0^2$, dado el nivel de significación α y dado que H_1 indica cola derecha en la distribución de χ_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $\chi_{1-\alpha, n-1}^2$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P\left[\chi^2 > \chi_{1-\alpha, n-1}^2\right] = \alpha$.

Luego, la región de rechazo de H_0 es el intervalo:

$$RC = \left\{ \chi^2 > \chi_{1-\alpha, n-1}^2 \right\}$$

3) Prueba unilateral de cola a la izquierda

Si se prueba $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 < \sigma_0^2$, dado el nivel de significación α y dado que H_1 indica cola izquierda en la distribución de χ_{calc} especificada por H_0 supuesta verdadera, se hallan los valores críticos $\chi_{\alpha, n-1}^2$, tales que la probabilidad de rechazar H_0 si de hecho es verdadera sea $P \left[\chi^2 < \chi_{\alpha, n-1}^2 \right] = \alpha$.

Luego, la región de rechazo de H_0 es el intervalo:

$$RC = \left\{ \chi^2 < \chi_{\alpha, n-1}^2 \right\}$$

Prueba de hipótesis para la proporción

Prueba de hipótesis para una proporción

Sea $\bar{P} = \frac{X}{n}$ la variable aleatoria que denota la proporción de éxito de muestras de tamaño n , X_1, X_2, \dots, X_n extraídas de la población Bernoulli $B(1, p)$. Entonces, si n es suficientemente grande:

$$Z = \frac{\bar{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \underset{TLC}{\sim} N(0, 1)$$

Si se prueba $H_0 : p = p_0$ contra $H_1 : p \neq p_0$:

$$RC = \left\{ Z < -z_{1-\frac{\alpha}{2}} \quad \text{o} \quad Z > z_{1-\frac{\alpha}{2}} \right\}$$

Si se prueba $H_0 : p = p_0$ contra $H_1 : p > p_0$:

$$RC = \{Z > z_{1-\alpha}\}$$

Si se prueba $H_0 : p = p_0$ contra $H_1 : p < p_0$:

$$RC = \{Z < -z_{1-\alpha}\}$$