

Introducción a la Estadística y Probabilidad

Tema 9. Inferencia estadística. Contraste de hipótesis.

MANUEL MONGE, Ph.D.

Departamento de Economía Aplicada y Métodos Cuantitativos

Facultad de Derecho, Economía y Gobierno

Universidad Francisco de Vitoria

1. Introducción

2. Procedimiento para contrastar una hipótesis

Paso 1. Se establecen las hipótesis nula (H_0) y alternativa (H_1)

Paso 2. Se selecciona un nivel de significancia

Paso 3. Se identifica el estadístico de prueba

Paso 4. Se formula la regla de decisión

Paso 5. Se toma una muestra y se decide

Paso 6. Se interpreta el resultado

3. Pruebas de significancia de una o dos colas

Valor p en la prueba de hipótesis

1. Introducción

1. Introducción

¿Qué es una hipótesis?

- Una **hipótesis** es una afirmación relativa a un parámetro de la población sujeta a verificación.

En términos estadísticos,

- Una **hipótesis estadística** es una suposición o conjetura (H), con respecto a una característica desconocida a la que llamamos *parámetro* (θ), de una población de interés, como por ejemplo, la media poblacional (μ), la varianza (σ^2), o la proporción de una característica poblacional (p).

2. Procedimiento para contrastar una hipótesis

2. Procedimiento para contrastar una hipótesis



Paso 1. Se establecen las hipótesis nula (H_0) y alternativa (H_1)

Paso 1. Se establecen las hipótesis nula (H_0) y alternativa (H_1)

- Formular un contraste de hipótesis consiste en plantear dos **hipótesis incompatibles**, es decir, no pueden ser ciertas al mismo tiempo.
- La **hipótesis nula** que se supone correcta antes de conocer los datos experimentales y se representa por H_0 .
- La **hipótesis alternativa** se denota por H_1 o H_A y es una suposición incompatible con la hipótesis nula, que los resultados experimentales, muestra, pueden identificar como correcta

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

La especificación de la hipótesis nula y de la hipótesis alternativa depende del problema.

Paso 1. Se establecen las hipótesis nula (H_0) y alternativa (H_1)

Ejemplo

Una empresa de cereales quiere averiguar si el peso medio de los paquetes es mayor de lo que éstos indica. Sea μ el peso medio poblacional, en gramos, por paquete. El contraste de hipótesis se formularía como sigue:

$$\begin{cases} H_0 : \mu \leq 1600 \\ H_1 : \mu > 1600 \end{cases}$$

Resolver el contraste de hipótesis consiste en decidir cuál de las suposiciones se encuentra avalada por la evidencia experimental que se obtiene a través de una muestra aleatoria.

Paso 2. Se selecciona un nivel de significancia

Paso 2. Se selecciona un nivel de significancia

- Después de establecer las hipótesis nula y alternativa, el siguiente paso consiste en determinar el nivel de significancia.
- El nivel de significancia o significatividad, que ya introdujimos en el tema anterior expresa el nivel de riesgo que se corre al rechazar la hipótesis nula cuando es verdadera.
- Los más utilizados en la literatura son 1 % (0.01), 5 % (0.05) y 10 % (0.10).
- Todo investigador debe elegir el nivel de significatividad antes de formular una regla de decisión y recopilar los datos de la muestra.

Paso 2. Se selecciona un nivel de significancia

Continuación del ejemplo anterior...

Parece obvio que para contrastar la hipótesis nula de que el peso medio de los paquetes de cereales es menor que 1600 gramos, obtenemos una muestra aleatoria de paquetes y se calcula su media muestral. Si la media muestral es considerablemente superior a 1600 gramos, podemos rechazar la hipótesis nula y aceptar la hipótesis alternativa. En general, cuanto más distante de 1600 sea la media muestral, mayor será la probabilidad de rechazar la hipótesis nula.

Pero...

Construir la regla de decisión en función de un valor muestral puede extraer conclusiones erróneas sobre el parámetro poblacional.

La siguiente tabla resume los posibles tipos de error que podemos cometer.

Decisión/Realidad	H_0 cierta	H_1 falsa
Aceptamos H_0	Correcto	Error de Tipo II
Rechazamos H_0	Error de Tipo I	Correcto

Por lo tanto, sabemos que la regla de decisión adoptada tiene alguna probabilidad de extraer una conclusión errónea.

- **Error tipo I:** Rechazar la hipótesis nula, H_0 , cuando es verdadera.
 $P(\text{Cometer Error de Tipo I}) = P(\text{Rechazar } H_0 | H_0 \text{ es cierta}) = \alpha$
 $P(\text{Aceptar } H_0 | H_0 \text{ es cierta}) = 1 - \alpha$
- **Error tipo II:** Aceptar la hipótesis nula cuando es falsa.
 $P(\text{Cometer Error de Tipo II}) = P(\text{Aceptar } H_0 | H_0 \text{ es falsa}) = \beta$
 $P(\text{Rechazar } H_0 | H_0 \text{ es falsa}) = 1 - \beta$

Paso 3. Se identifica el estadístico de prueba

Paso 3. Se identifica el estadístico de prueba

Como vimos en el tema anterior, también en este se puede dar dos supuestos en el contraste de hipótesis:

- Contraste de la media de la distribución normal cuando σ se conoce (z).
- Contraste de la media de la distribución normal cuando σ no se conoce (t *Student*).

En este tema, como en el anterior solo nos vamos a centrar en el supuesto de que conocemos σ .

Paso 3. Se identifica el estadístico de prueba

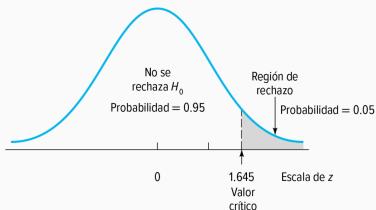
La prueba de hipótesis de la media (μ), cuando se conoce σ o el tamaño de la muestra es grande, es el estadístico de prueba z que se calcula de la siguiente manera:

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

**Paso 4. Se formula la regla de
decisión**

Paso 4. Se formula la regla de decisión

- Una regla de decisión es un enunciado sobre las condiciones específicas en que se rechaza la hipótesis nula y aquellas en las que no es así.
- La región o área de rechazo define la ubicación de todos los valores que son tan grandes o tan pequeños, que la probabilidad de que ocurran en una hipótesis nula verdadera es remota.
- La región de rechazo de una prueba de significancia se presenta en la siguiente figura:



En la figura se observa lo siguiente:

1. El área en que se acepta la hipótesis nula se localiza a la izquierda de 1.645
2. El área de rechazo se encuentra a la derecha de 1.645
3. Se aplica una prueba de una sola cola (este hecho se explicará más adelante).
4. Se eligió el nivel de significancia 0.05
5. La distribución muestral del estadístico z es normal
6. El valor de 1.645 es el **valor crítico**

El **valor crítico** es el punto de división entre la región en que se rechaza la hipótesis nula y aquella en la que se

Paso 5. Se toma una muestra y se decide

Paso 5. Se toma una muestra y se decide

El quinto paso en la prueba de hipótesis consiste en

1. Seleccionar la muestra y calcular el estadístico de prueba.
2. Compararlo con el valor crítico.
3. Tomar la decisión de rechazar o no la hipótesis nula.

Sin embargo, como la decisión se basa en la muestra, siempre es posible cometer cualquiera de los dos errores de decisión.

- Es posible cometer un error de tipo I si la hipótesis nula se rechaza cuando no debía ser así.
- Es posible cometer un error de tipo II si la hipótesis nula no se rechaza y debía haberlo sido.

Paso 6. Se interpreta el resultado

Paso 6. Se interpreta el resultado

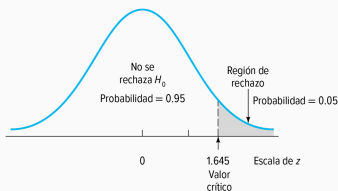
El paso final en el procedimiento de prueba de hipótesis es interpretar los resultados.

¿Qué se puede decir o reportar con base en los resultados de la prueba estadística?

3. Pruebas de significancia de una o dos colas

3. Pruebas de significancia de una o dos colas

Región de rechazo en la cola superior de la curva



- Este figura que he representado en una diapositiva anterior, describe **una prueba de una cola**.
- Se llama así porque la región de rechazo se localiza en una cola de la curva.
- En este caso, dicha región está situada en la parte derecha (o superior) de la curva.

Un ejemplo,

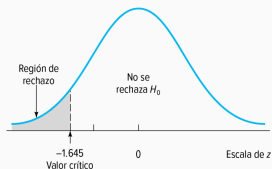
Suponga que el departamento de empaquetado de General Foods Corporation se preocupa porque algunas cajas de cereales exceden considerablemente el peso. Se sabe que el cereal se empaqueta en cajas de 453 gramos ($H_0 : \mu \leq 453$). Por tanto, la hipótesis alternativa es $H_1 : \mu > 453$.

$$\begin{cases} H_0 : \mu \leq 453 \\ H_1 : \mu > 453 \end{cases}$$

Como se puede observar, el signo de desigualdad en la hipótesis alternativa ($>$) señala la región de rechazo ubicada en la cola superior. (La condición de igualdad siempre aparece en H_0 y jamás en H_1)

3. Pruebas de significancia de una o dos colas

Región de rechazo en la cola inferior de la curva



- En esta figura se presenta un caso donde la región de rechazo se encuentra en la cola izquierda (inferior) de la distribución normal.

Un ejemplo,

Considere el problema de los fabricantes de automóviles; en este caso, las grandes compañías de alquiler de coches y otras empresas que compran grandes cantidades de neumáticos desean que duren un promedio de 60.000km en condiciones normales ($H_0 : \mu \geq 60,000$). Por lo tanto, rechazarán un envío de neumáticos si las pruebas revelan que la vida de estas es mucho menor en promedio ($H_1 : \mu < 60,000$). Por tanto, la prueba se plantea de manera que satisfaga la preocupación de los fabricantes de automóviles respecto de que la *vida útil de los neumáticos no sea menor a 60.000km*

$$\begin{cases} H_0 : \mu \geq 60,000 \\ H_1 : \mu < 60,000 \end{cases}$$

Como se puede observar, el signo de desigualdad en la hipótesis alternativa ($<$) señala la región de rechazo ubicada en la cola inferior.

3. Pruebas de significancia de una o dos colas

Región de rechazo de una prueba de dos colas



- Si no se especifica dirección alguna en la hipótesis alternativa, utilice una prueba de **dos colas**.

Un ejemplo,

- H_0 : el ingreso medio anual de los corredores de bolsa es de 65.000 dólares anuales.
- H_1 : el ingreso medio anual de los corredores de bolsa no es igual a 65.000 dólares anuales.

$$\begin{cases} H_0 : \mu = 65,000 \\ H_1 : \mu \neq 65,000 \end{cases}$$

Si se rechaza la hipótesis nula H_0 y se acepta H_1 , en el caso de las dos colas, el ingreso medio puede ser significativamente mayor o inferior a 65.000 dólares anuales.

Para dar cabida a ambas posibilidades, el área de 5% de rechazo se divide con equidad en las dos colas de la distribución muestral (2,5% cada una).

Observe que el área total en la distribución normal es de 1, que se calcula mediante $0,95 + 0,025 + 0,025$

3. Pruebas de significancia de una o dos colas

Ejemplo

Valor p en la prueba de hipótesis

Valor p en la prueba de hipótesis

- El valor p (p – *value*) es la probabilidad de observar un valor muestral tan extremo o más que el valor observado, si la hipótesis nula es verdadera.
- Trata de responder a la pregunta ¿cuánta confianza hay en el rechazo de la hipótesis nula?
- Este enfoque indica la probabilidad (en el supuesto de que la hipótesis nula H_0 sea verdadera) de obtener un valor estadístico de prueba por lo menos tan extremo como el valor real que se obtuvo.
- Este proceso compara la probabilidad, denominada **valor p** , con el nivel de significancia.
- Si el valor p es menor que este, H_0 se rechaza.
- Si es mayor, H_0 no se rechaza.
- La determinación del valor p no solo da como resultado una decisión respecto de H_0 , sino que brinda la oportunidad de observar la fuerza de la decisión.
- Un valor p muy pequeño, como 0.0001, indica que existe poca probabilidad de que H_0 sea verdadera; por otra parte, un valor p de 0.2033 significa que H_0 no se rechaza y que existe poca probabilidad de que sea falsa.

Ejemplo

¿Cómo se interpreta un valor p ?

Si el valor p es menor que

- 0.10, hay cierta evidencia de que H_0 no es verdadera.
- 0.05, hay evidencia fuerte de que H_0 no es verdadera.
- 0.01, hay evidencia muy fuerte de que H_0 no es verdadera.
- 0.001, hay evidencia extremadamente fuerte de que H_0 no es verdadera.