Maestría en Marketing

Métodos y técnicas de análisis cuantitativo y cualitativo

Test de Hipótesis

• Referencia: PR Caps 22, 23. KT Cap 17

Otoño 2003

Problema a tratar

• Nuestro objeto de estudio es una población en particular, pero nuestra manera de aproximarnos al problema es a través de una muestra de esa población.

• El problema entonces es cómo, mediante las cosas que observamos de la muestra, hacer inferencia sobre las características de la población.

Definición de Hipótesis

• Una hipótesis es una afirmación sobre una característica de la población.

• Es importante notar que las hipótesis no son aseveraciones sobre una persona en particular, sino sobre una característica de la población.

• De esta manera, para testear una hipótesis es necesario:

  • Especificar la población/es de interés.

  • Definir la característica o variable y cómo medirla.

  • Relacionar la hipótesis con las parámetros de la población

Hipótesis Nula y Alternativa

• Sea θ la característica o parámetro sobre el cual queremos conjeturar. Entonces la hipótesis nula (H_0) es una afirmación que representa la conjetura que queremos testar.

  • Hipótesis simples: H_0: θ = a

  • Hipótesis compuestas: H_0: a < θ < b

• La hipótesis alternativa (H_A) es la alternativa contra la cual queremos comparar la hipótesis nula. Si H_0 es un hipótesis simple, entonces puede haber 2 tipos de H_A:

  • Unilaterales: H_A: a < θ o H_A: θ < a

  • Bilaterales: H_A: $\theta \neq$ a

Regla de decisión

	H_0 es verdadera	H_0 es falsa
Aceptar	OK $1 - \alpha$	Error Tipo 2 β
Rechazar	Error Tipo 1 α	OK $1 - \beta$

• Probabilidad(Error Tipo 1) = α (nivel de significatividad)

Test sobre la media de una población normal (una cola)

• En una muestra de *n* observaciones independientemente obtenidas de una población que tiene distribución normal queremos testear

$$H_0:\ m = a$$

$$vs.$$

$$H_A:\ m > a$$

• Utilizaremos al promedio como un estimador de la media de la población.

• Dado un nivel de significatividad α, podemos escribir la probabilidad de competer error tipo 1 como

$$P(\bar{x} - a > K_\alpha) = \alpha$$

• O lo que es lo mismo

$$P\left(\frac{\bar{x} - a}{\sigma / \sqrt{n}} > \frac{K_\alpha}{\sigma / \sqrt{n}}\right) = \alpha$$

$$P(Z > Z_\alpha) = \alpha$$

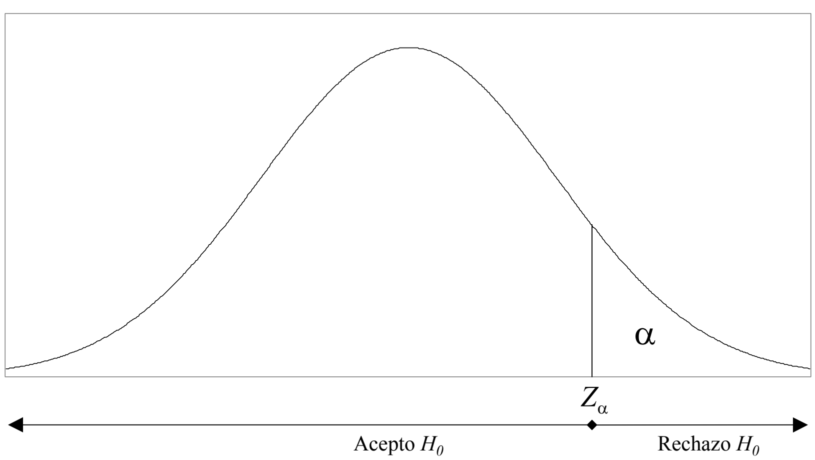
• Dado que las observaciones provienen de una muestra normal, se puede mostrar que si H_0 es cierta entonces $\bar{x} \sim N(m, \sigma^2 / n)$

• Entonces, $Z \sim N(0,1)$

• De esta manera, puedo encontrar Z_α mirando la tabla normal

• Luego, la regla de decisión consiste en computar Z y compararlo con Z_α. Si $Z > Z_\alpha$ entonces rechazo H_0

• Gráficamente



• En la práctica, la verdadera varianza de la población nunca es conocida, solo tenemos una estimación de este a traves del S^2.

• Lo que haremos es proceder reemplazando la verdadera varianza por el valor estimado.

• En realidad, cuando hacemos este reemplazo, el estadístico Z ya no tiene más distribución normal sino que tiene una distribución que se llama *t de Student*.

• Aún así, para tamaños de muestra grandes ($n>30$) la distribución t de Student es prácticamente indistinguible de la normal estandar. Luego, lo que haremos es proceder como si la distribución de la población fuese normal.

Test sobre la media de una población normal (dos colas)

• En una muestra de n observaciones independientemente obtenidas de una población que tiene distribución normal queremos testear

$$H_0: \ m = a$$

$$vs.$$

$$H_A: \ m \neq a$$

• Utilizaremos al promedio como un estimador de la media de la población.

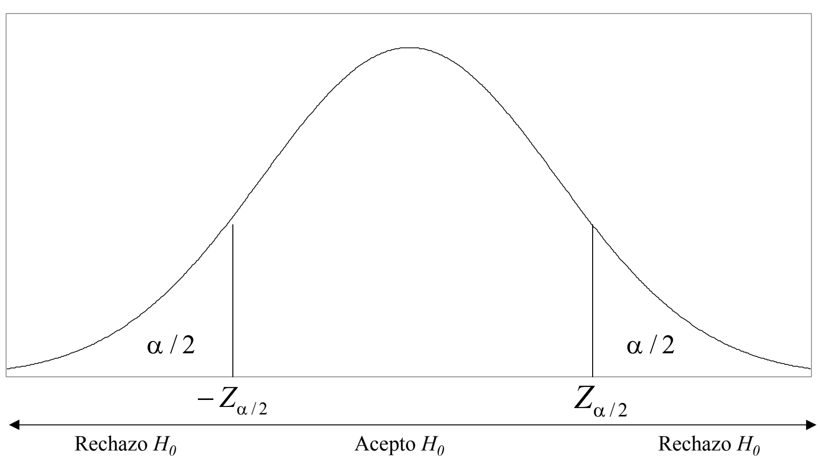
• Dado un nivel de significatividad α, en este caso podemos escribir la probabilidad de competer error tipo 1 como

$$P(Z > Z_{\alpha/2}) = \alpha/2 \;\text{ y }\; P(Z < -Z_{\alpha/2}) = \alpha/2$$

•De esta manera, puedo encontrar $Z_{\alpha/2}$ mirando la tabla normal.

• Luego, la regla de decisión consiste en computar Z y compararlo con $Z_{\alpha/2}$. Si $Z > Z_{\alpha/2}$ o $Z < -Z_{\alpha/2}$ entonces rechazo H_0

• Gráficamente



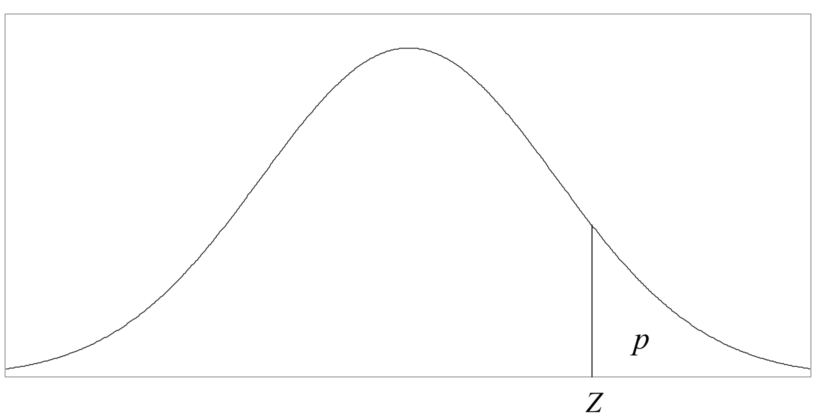
Método de p-valor

• En un test de un cola, el p-valor se define como la probabilidad acumulada del valor computado del estadístico Z para el test en particular.

• De esta manera, el p-valor será el mínimo nivel de significatividad para el cual se rechaza H_0.

• Luego, lo que debemos hacer es comparar este p-valor con el nivel de significatividad deseado para el test. Si el p-valor es menor al nivel de significatividad, entonces rechazamos H_0, sino no.

•Gráficamente,



• Para el test a dos colas, la forma es análoga a la que vimos para encontrar el valor crítico de la regla de decisión.

Intervalos de confianza

• Un intervalo de confianza de nivel *K*% para el verdadero valor de parámetro es un intervalo en el cual la probabilidad de que el verdadero valor del parámetro se encuentre en ese intervalo es del K%.

• Es decir, lo que queremos encontrar son *A* y *B* tales que,

$$P(A < m < B) = 1 - \alpha$$

• Para esto, recordemos que sabemos que

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

• De esta manera, sabemos que

$$P\left(-Z_{\alpha/2} < \frac{\bar{x} - a}{\sigma / \sqrt{n}} < Z_{\alpha/2} \right) = 1 - \alpha$$

$$P\left(-\frac{Z_{\alpha/2}\sigma}{\sqrt{n}} < \bar{x} - a < \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$P\left(\bar{x} - \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} < a < \bar{x} + \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

• Luego,

$$IC_{1-\alpha} = \bar{x} \pm \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}$$

• Notemos que realizar un intervalo de confianza de nivel (1-α) para el valor de un parámetros es equivalente a realizar un test de hipótesis simple a dos colas, con nivel de significatividad α, de que el verdadero valor del parámetro está en el centro del intervalo de confianza.

Generalización

• Aún cuando las observaciones provengan de una población que no es normal, se puede mostrar que un estimador $\hat{\theta}$ para θ que pueda escribirse como un promedio, cuando la muestra es "grande", cumple que

$$Z = \frac{\hat{\theta} - \theta}{desvest(\hat{\theta}) / \sqrt{n}} \sim N(0,1)$$

• De esta manera, la metodología utilizada con el supuesto de distribución normal es aplicable para los casos que veremos.

Ejemplo: Test para una proporción

• Queremos testear

$$H_0:\ p = a \quad vs. \quad H_A:\ p \neq a$$

donde p es la proporción de individuos con una característica determinada en la población.

• El tener o no tener un característica tiene distribución Bernoulli. Luego una proporción es la suma de Bernoullis (una Binomial) dividido n. Luego, la varianza estimada de una proporción es $p(1-p)$

• Luego el estadístico que debemos mirar (que tiene distribución normal estandar) es

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