Tema 8. Contrastes de hipótesis

Eva Romero Ramos evarom03@ucm.es

Universidad Complutense de Madrid

Outline

Motivación

Un fabricante afirma que la duración media de las baterías para móvil que fabrica es de 36 horas.

Una empresa que fabrica móviles utiliza estás baterías en sus dispositivos y en los últimos meses esta recibiendo reviews negativas en relación con la duración de las baterías. Esta empresa ha firmado un contrato para la adquisición de dichas baterías a un precio prefijado durante 5 años y aún quedan 4 años para que venza dicho contrato. Además, los precios de las baterías en el mercado están subiendo, por lo que si decide cambiar de distribuidor tendrá que comprarlas a un precio superior.

Si asumimos que la duración de las baterías se comporta como una distribución normal con desvición típica 10 horas, **¿Podemos ayudar a la empresa a tomar la mejor decisión?**

Outline

1. Definición e interpretación

2. Contrastes de hipótesis en poblaciones normales

Contraste de hipótesis

Definición

El contraste de hipótesis es una herramienta que, como su propio nombre indica, nos permite contrastar una cierta hipótesis relacionada con algún parámetro poblacional o con la propia estructura de datos de la población.

Las hipótesis

Definición (Hipótesis nula)

La hipótesis nula es el enunciado para el que se plantea para corroborar o descartar su validez por medio del contraste de hipótesis.

Esta hipótesis debe estar referida a información poblacional que comprobaremos mediante la información muestral.

La hipótesis nula se denota por H_0.

Definición (Hipótesis alternativa)

La hipótesis alternativa es el resultado esperado si finalmente concluimos que no se cumple la hipótesis nula.

La denotaremos por H_1.

Elementos del contraste de hipótesis

Definición (Estadístico de contraste)

El estadístico de contraste será la función de elementos muestrales o estadístico que utilicemos para la realización del contraste.

Definición (Región crítica)

Llamaremos Región Crítica al conjunto de valores que puede tomar nuestro estadístico de contraste y que constituyen una evidencia suficiente para afirmar que la hipótesis nula es falsa.

Definición (Región de aceptación)

Llamaremos Región de Aceptación al conjunto de valores que puede tomar nuestro estadístico de contraste que no constituyen una evidencia suficiente para rechazar la hipótesis nula.

Tipos de errores

Al realizar el contraste de hipótesis podemos tener las siguientes situaciones:

Decisión	**Hipótesis cierta (H_0)**	**Hipótesis cierta (H_1)**
Aceptar H_0	Correcto	Error tipo II
Rechazar H_0	Error tipo I	Correcto

Tipos de erróres

- La probabilidad de cometer error de tipo I se denomina nivel de significación del contraste y se designa por α.
- El nivel de significación es fijado por el investigador y habitualmente se trata de minimizar lo máximo posible.
- Son valores de uso habitual para el nivel de significación $0,1$, $0,05$ o $0,01$.
- La probabilidad de cometer error de tipo II se denota por β y nos ayuda a calcular lo que denominamos potencia del contraste.
- La potencia del contraste es la probabilidad de rechazar la hipótesis nula cuando es realmente falsa y se calcula como $1-\beta$.
- También la podemos expresar como la probabilidad de aceptar la hipótesis alternativa cuando es cierta.

El p-valor

Definición

El p-valor es el nivel de significación más pequeño con el que se debe rechazar la hipótesis nula, teniendo en cuenta la información de la muestra observada.

También se puede entender como:

- Probabilidad de que el valor observado del estadístico de contraste esté en la región crítica, cuando la hipótesis nula es cierta.
- Probabilidad de que el estadístico de contraste observado nos indique que debemos rechazar la hipótesis nula, cuando es cierta.

Para calcularlo debemos obtener:

$$\boxed{P\left(\text{ El valor observado del estadístico este dentro de la región crítica} \middle| H_0\right)}$$

El p-valor

- El p-valor proporciona una regla de decisión tan válida como la región crítica y más sencilla de manejar, ya que la mayoría de los software estadísticos incluyen su cálculo en los contrastes de hipótesis que incorporan.
- La regla de decisión es que, una vez fijado el nivel de significación, si este es mayor que el p-valor obtenido rechazaremos la hipótesis nula y de lo contrario diremos que no tenemos evidencias suficientes para rechazarla, es decir:
  - Si $p - valor > \alpha \to$ no tendremos evidencias suficientes para rechazar la hipótesis nula.
  - Si $p - valor < \alpha \to$ tendremos evidencias suficientes para rechazar la hipótesis nula.

El nivel de significación no debe determinarlo el p-valor obtenido , debe decidirse antes de realizar el muestreo y en ningún caso dependerá el p-valor o del valor del estadístico de contraste.

Outline

Contraste de hipótesis para la media en poblaciones normales con varianza conocida

Nos planteamos contrastar:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Si $X \sim N(\mu, \sigma^2) \rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Teniendo en cuenta esto, usaremos como estadístico de contraste:

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

.

Por lo que la región crítica será:

$$\boxed{RC = \{|T| > z_{\alpha/2}\}}$$

Contraste de hipótesis para la media en poblaciones normales con varianza conocida

Si nos planteasemos con contrastar:

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Usaremos el mismo estadístico de contraste, pero la región crítica será:

$$\boxed{RC = \{T > z_\alpha\}}$$

Y para:

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

$$\boxed{RC = \{T < -z_\alpha\}}$$

Contraste de hipótesis para la media en poblaciones normales con varianza desconocida

Nos planteamos de nuevo contrastar:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Como la varianza es desconocida no podemos usar el estadístico anterior, pero sabemos que:

$$T = \frac{\bar{X} - \mu_0}{S'/\sqrt{n}} \sim t(n-1)$$

Con lo que podremos usarlo como estadístico de contraste y obtendremos la región crítica:

$$\boxed{RC = \{|T| > t_{n-1,\alpha/2}\}}$$

Contraste de hipótesis para la media en poblaciones normales con varianza desconocida

Si nos planteasemos con contrastar:

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Usaremos el mismo estadístico de contraste, pero la región crítica será:

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

Y para:

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

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

Contraste de hipótesis para la varianza en poblaciones normales

Se plantea contrastar:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

Sabemos que:

$$T = \frac{(n-1)S'^2}{\sigma_0^2} \sim \chi^2(n-1)$$

Con lo que podremos usarlo como estadístico de contraste y obtendremos la región crítica:

$$\boxed{RC = \{T < \chi^2_{n-1,1-\alpha/2} \text{ ó } T > \chi^2_{n-1,\alpha/2}\}}$$

Contraste de hipótesis para la varianza en poblaciones normales

Y el mismo estadístico de contraste nos permitirá contrastar:

$$\begin{cases} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases} \rightarrow \boxed{RC = \{T > \chi^2_{n-1,\alpha}\}}$$

y

$$\begin{cases} H_0 : \sigma^2 \geq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases} \rightarrow \boxed{RC = \{T < \chi^2_{n-1,1-\alpha}\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas conocidas

Tenemos ahora dos variables aleatorias: $X_1 \sim N(\mu_1, \sigma_1^2)$ & $X_2 \sim N(\mu_2, \sigma_2^2)$, y se plantea contrastar:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

En este caso sabemos que:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Usando T como estadístico de contraste obtenemos la región crítica:

$$\boxed{RC = \{|T| > Z_{\alpha/2}\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas conocida

Usando el mismo estadístico de contraste podemos contrastar:

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \rightarrow \boxed{RC = \{T > Z_\alpha\}}$$

y

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \rightarrow \boxed{RC = \{T < -Z_\alpha\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas desconocidas e iguales

Si se tienen dos variables aleatorias: $X_1 \sim N(\mu_1, \sigma^2)$ & $X_2 \sim N(\mu_2, \sigma^2)$, con varianzas desconocidas pero iguales y se plantea contrastar:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

En este caso sabemos que:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_c \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

donde: $S_c = \sqrt{\frac{(n-1) \cdot S_1'^2 + (m-1) \cdot S_2'^2}{n+m-2}}$

Usando T como estadístico de contraste obtenemos la región crítica:

$$\boxed{RC = \{|T| > t_{n+m-2, \alpha/2}\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas desconocidas e iguales

Usando el mismo estadístico de contraste podemos contrastar:

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \rightarrow \boxed{RC = \{T > t_{m+n-2,\alpha}\}}$$

y

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \rightarrow \boxed{RC = \{T < -t_{m+n-2,\alpha}\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas desconocidas y distintas

Para dos variables aleatorias: $X_1 \sim N(\mu_1, \sigma_1^2)$ & $X_2 \sim N(\mu_2, \sigma_2^2)$, nos plantea contrastar:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

En este caso sabemos que:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1'}{n} + \frac{S_2'}{m}}} \sim t(k) \text{ donde: } K = \frac{\left(S_1'^2/n + S_2'^2/m\right)^2}{\left[\frac{\left(S_1'^2/n\right)^2}{n-1}\right] + \left[\frac{\left(S_2'^2/m\right)^2}{m-1}\right]}$$

Usando T como estadístico de contraste obtenemos la región crítica:

$$\boxed{RC = \{|T| > t_{k,\alpha/2}\}}$$

Contraste de igualdad de medias en poblaciones normales con varianzas desconocidas y distintas

Usando el mismo estadístico de contraste podemos contrastar:

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \quad \rightarrow \boxed{RC = \{T > t_{k,\alpha}\}}$$

y

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases} \quad \rightarrow \boxed{RC = \{T < -t_{k,\alpha}\}}$$

Contraste de igualdad de medias en poblaciones normales con datos pareados

Para dos variables aleatorias: $X_1 \sim N(\mu_1, \sigma_1^2)$ & $X_2 \sim N(\mu_2, \sigma_2^2)$, nos plantea contrastar:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

En este caso, al tratarse de datos pareados tiene sentido obtener directamente la diferencia entre ellos D y contrastar si la diferencia es 0. Usaremos entonces el estadístico de contraste:

$$T = \frac{\bar{D}}{S_D'^2 / \sqrt{n}} \sim t(n-1)$$

Por lo que la región crítica será:

$$\boxed{RC = \{|T| > t_{n-1,\alpha/2}\}}$$

Contraste de igualdad de medias en poblaciones normales con datos pareados

Y con el mismo estadístico podremos contrastar:

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases} \rightarrow \boxed{RC = \{T > t_{n-1,\alpha}\}}$$

Y

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_1 : \mu_< \mu_2 \end{cases} \rightarrow \boxed{RC = \{T < -t_{n-1,\alpha}\}}$$

Contraste de igualdad de varianzas en poblaciones normales

Para dos variables aleatorias de poblaciones normales nos plantea contrastar:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Para el coeciente de varianzas sabemos que:

$$T = \frac{S_1'^2}{S_2'^2} \sim F(n-1, m-1)$$

De modo que la región crítica será:

$$\boxed{RC = \{T < F_{n-1,m-1,1-\alpha/2} \text{ ó } T > F_{n-1,m-1,\alpha/2}\}}$$

Contraste de igualdad de varianzas en poblaciones normales

Y con el mismo estadístico podremos contrastar:

$$\begin{cases} H_0 : \sigma_1^2 \leq \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases} \quad \rightarrow \quad \boxed{RC = \{ T > F_{n-1,m-1,1-\alpha}}$$

Y

$$\begin{cases} H_0 : \sigma_1^2 \geq \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases} \quad \rightarrow \quad \boxed{RC = \{ T < F_{n-1,m-1,1-\alpha} \}}$$

Contraste para la proporción

Si p es la proporción de individuos de una población grande que presenta cierta característica, podemos contrastar:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

Teniendo en cuenta que:

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

De modo que la región crítica será:

$$\boxed{RC = \{|T| > z_{\alpha/2}\}}$$

Contraste para la proporción

Y con el mismo estadístico podremos contrastar:

$$\begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{cases} \rightarrow \boxed{RC = \{T > z_\alpha\}}$$

Y

$$\begin{cases} H_0 : p \geq p_0 \\ H_1 : p < p_0 \end{cases} \rightarrow \boxed{RC = \{T < -z_\alpha\}}$$

Contraste de igualdad de proporciones

Consideraremos dos poblaciones grandes y denotaremos por p_1 y p_2 a las proporciones de individuos de cada población que presentan cierta característica de interés, nos plantearemos contrastar:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

Teniendo en cuenta que:

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}} \sim N(0,1)$$

De modo que la región crítica será:

$$\boxed{RC = \{|T| > z_{\alpha/2}\}}$$

Contraste para la proporción

Y con el mismo estadístico podremos contrastar:

$$\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases} \rightarrow \boxed{RC = \{T > z_\alpha\}}$$

Y

$$\begin{cases} H_0 : p_1 \geq p_2 \\ H_1 : p_1 < p_2 \end{cases} \rightarrow \boxed{RC = \{T < -z_\alpha\}}$$