Técnicas de Inferencia Estadística II

Tema 2. Contrastes de hipótesis paramétricos

Parte I. Contrastes para una muestra

M. Concepción Ausín
Universidad Carlos III de Madrid

Grado en Estadística y Empresa
Curso 2015/16

Contenidos

Contrastes para una muestra de una población normal

Suponemos una muestra aleatoria simple $(X_1, X_2, \ldots, X_n)$ de una población normal, $N(\mu, \sigma^2)$. Queremos resolver contrastes del tipo:

Bilateral	Unilateral por la derecha	Unilateral por la izquierda
$H_0 : \theta = \theta_0$	$H_0 : \theta = \theta_0$	$H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$	$H_1 : \theta > \theta_0$	$H_1 : \theta < \theta_0$

donde θ representa el parámetro de interés, que puede ser la media, μ, o la varianza, σ^2.

- Las región de rechazo se obtiene usando un estadístico de contraste, que es una medida de discrepancia entre la muestra de datos y la hipótesis nula.
- Un estadístico de contraste será cualquier función de la muestra y del parámetro especificado en H_0 (con distribución conocida cuando H_0 es cierta) que permita decidir hasta qué punto la muestra de datos está de acuerdo o no con la hipótesis nula.

Contrastes para la media con varianza conocida

Suponemos primero una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ de una población normal, $N(\mu, \sigma^2)$, con la varianza, σ^2 conocida.

Queremos resolver contrastes para la media del tipo:

$$\begin{array}{lll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 & H_1 : \mu > \mu_0 & H_1 : \mu < \mu_0 \end{array}$$

El estadístico de contraste en este caso es:

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim_{H_0} N(0, 1)$$

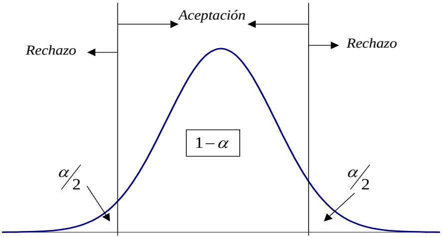
- Este estadístico proporciona una medida de discrepancia entre los datos y la hipótesis nula.
- Para elegir la región de rechazo, de la hipótesis nula debemos de fijar el nivel de significación, α, donde recordamos que:

$$\alpha = \Pr(\text{rechazar } H_0 \mid H_0 \text{ cierta})$$

Contrastes para la media con varianza conocida

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$
  - Dados los datos, $(x_1, \ldots, x_n)$, si $\left| \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \right|$ es grande, se rechazará H_0.
  - La región de rechazo es:

$$R = \left\{ \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2} \right\}$$



  - p-valor $= 2\,\mathrm{Pr}\left(Z > \left| \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \right| \right)$

Ejemplo 2.1.

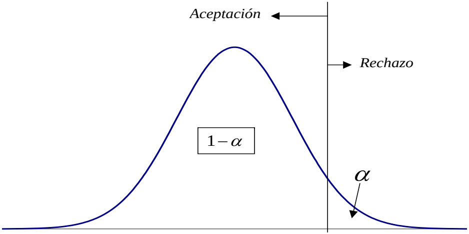
Uno de los productos de una empresa es café molido en paquetes de 200 gramos. Se diseña un experimento en el que se pesan con precisión el peso de 15 paquetes, seleccionados aleatoriamente. Los pesos son 208, 206, 210, 199, 202, 196, 198, 209, 211, 204, 206, 197, 196, 203 y 207. Se supone que el peso de estos paquetes sigue una distribución normal y que su desviación típica es conocida (no realista) e igual a 4.5 gramos.

- *La empresa desea saber si el peso medio de los paquetes es distinto de los 200 gramos que figuran en la etiqueta. Contrastar dicha hipótesis usando el p-valor para* $\alpha = 0.05$ *y 0.01.*
- *Construir dos intervalos de confianza al 95 % y al 99 % para el valor real del peso medio de un paquete de café.*

Contrastes para la media con varianza conocida

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$
  - Dados los datos, $(x_1, \ldots, x_n)$, si $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ es grande, se rechazará H_0.
  - La región de rechazo es:

$$R = \left\{ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\}$$

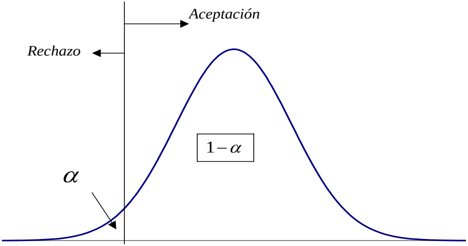


  - p-valor $= \Pr\left(Z > \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)$

Contrastes para la media con varianza conocida

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$
  - Dados los datos, $(x_1, \ldots, x_n)$, si $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ es pequeño, se rechazará H_0.
  - La región de rechazo es:

$$R = \left\{ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha \right\}$$



  - p-valor $= \Pr\left(Z < \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \right)$

Ejemplo 2.2.

Uno de los productos de una empresa es café molido en paquetes de 200 gramos. Se diseña un experimento en el que se pesan con precisión el peso de 15 paquetes, seleccionados aleatoriamente. Los pesos son 208, 206, 210, 199, 202, 196, 198, 209, 211, 204, 206, 197, 196, 203 y 207. Se supone que el peso de estos paquetes sigue una distribución normal y que su desviación típica es conocida (no realista) e igual a 4.5 gramos.

- *A la vista del resultado anterior, la empresa desea saber si el peso medio de los paquetes es de hecho superior a los 200 gramos que figuran en la etiqueta. Contrastar dicha hipótesis usando el p-valor para* $\alpha = 0.05$ *y 0.01.*

Contrastes para la media con varianza desconocida

En la práctica, la varianza poblacional σ es casi siempre desconocida.

Consideramos ahora el caso para una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ de una población normal, $N(\mu, \sigma^2)$, con la varianza, σ^2, desconocida.

Queremos resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : \mu = \mu_0 & \quad H_0 : \mu = \mu_0 & \quad H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 & \quad H_1 : \mu > \mu_0 & \quad H_1 : \mu < \mu_0 \end{array}$$

El estadístico de contraste en los tres casos es:

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim_{H_0} t_{n-1}$$

Gracias a la simetría de la distribución t, las regiones de rechazo se obtienen de manera equivalente a las de los contrastes para la media con varianza conocida.

Contrastes para la media con varianza desconocida

Cálculo del p-valor

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$

$$\text{p-valor} = \Pr\left(t_{n-1} > \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$

$$\text{p-valor} = \Pr\left(t_{n-1} < \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

- $H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$

$$\text{p-valor} = 2\Pr\left(t_{n-1} > \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

Ejemplo 2.3.

Uno de los productos de una empresa es café molido en paquetes de 200 gramos. Se diseña un experimento en el que se pesan con precisión el peso de 15 paquetes, seleccionados aleatoriamente. Los pesos son 208, 206, 210, 199, 202, 196, 198, 209, 211, 204, 206, 197, 196, 203 y 207. Se supone que el peso de estos paquetes sigue una distribución normal y que su desviación típica es desconocida.

- *La empresa desea saber si el peso medio de los paquetes es distinto de los 200 gramos que figuran en la etiqueta y, en ese caso, averiguar si es superior a 200 gramos. Contrastar dicha hipótesis usando el p-valor para* $\alpha = 0.05$ *y 0.01.*
- *Construir dos intervalos de confianza al 95 % y al 99 % para el valor real del peso medio de un paquete de café.*

Contrastes para la varianza

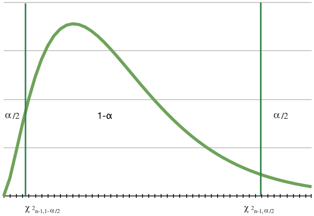
Consideramos una muestra aleatoria $(X_1, X_2, \dots, X_n)$ de una población normal, $N(\mu, \sigma^2)$, con μ, desconocida. Queremos contrastar:

$$\begin{array}{lll} H_0 : \sigma^2 = \sigma_0^2 & \qquad H_0 : \sigma^2 = \sigma_0^2 & \qquad H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 & \qquad H_1 : \sigma^2 > \sigma_0^2 & \qquad H_1 : \sigma^2 < \sigma_0^2 \end{array}$$

El estadístico de contraste en los tres casos es:

$$\frac{(n-1)S^2}{\sigma_0^2} \sim_{H_0} \chi^2_{n-1}$$

Las regiones de rechazo se obtienen de manera equivalente a los casos anteriores, pero teniendo en cuenta que la distribución χ^2_{n-1} es asimétrica.



Contrastes para la varianza

Cálculo del p-valor

- $H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 > \sigma_0^2$

$$\text{p-valor} = \Pr\left(\chi^2_{n-1} > \frac{(n-1)s^2}{\sigma_0^2}\right)$$

- $H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 < \sigma_0^2$

$$\text{p-valor} = \Pr\left(\chi^2_{n-1} < \frac{(n-1)s^2}{\sigma_0^2}\right)$$

- $H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$

$$\text{p-valor} = \text{mín}\left\{2\Pr\left(\chi^2_{n-1} > \frac{(n-1)s^2}{\sigma_0^2}\right), 2\Pr\left(\chi^2_{n-1} < \frac{(n-1)s^2}{\sigma_0^2}\right)\right\}$$

Ejemplo 2.4.

Un inversor quiere saber si la variación del precio de las acciones de una compañía este mes será superior a la variación del mes pasado, que fue de 114.09. Ha observado que la varianza muestral de los precios de los primeros 10 días de este mes ha sido igual a 110.2. Asumiendo que los 10 datos pueden considerarse una muestra aleatoria de una población normal, contrastar al 5 % la hipótesis anterior.

Contrastes para la media a partir de una muestra grande

Supongamos que se tiene una muestra $(X_1, \ldots, X_n)$ de una población cualquiera con n grande ($n > 30$).

Aunque la población no sea normal, se pueden resolver contrastes para la media:

$$\begin{array}{lll} H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 & H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 & H_1 : \mu > \mu_0 & H_1 : \mu < \mu_0 \end{array}$$

utilizando el Teorema Central del Límite, que garantiza que:

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \rightarrow_{H_0} N(0,1)$$

Ejemplo 2.5.

La vida media de una muestra de 55 tubos fluorescentes producidos por una empresa es de 1750 horas con una cuasi-desviación típica de 120 horas. Contrastar la hipótesis de que la vida media sea distinta de 1600 horas, utilizando un nivel de significación de 0.05.