

CONTRASTE DE HIPÓTESIS

Introducción

Hasta ahora hemos visto dos formas de efectuar una estimación de un parámetro de la población a partir de una muestra de tamaño n : la estimación puntual y la estimación por intervalos de confianza.

En este tema veremos una tercera forma de inferencia estadística denominada Contraste de hipótesis y que, como apunta la expresión, trata de dar crédito o no (contraste) a una afirmación que haremos previamente (hipótesis) a raíz de los resultados obtenidos en las muestras.

Por ejemplo deseo contrastar que la proporción de recién nacidos varones en un determinado hospital es del 56%. El estudio mediante un contraste de hipótesis me permitirá aceptar el dato como válido o no, con un determinado nivel de confianza.

Deseo contrastar si el precio medio de las latas de conserva de venta en un gran almacén mayor que 7 euros...etc.

Conceptos

Se llama **hipótesis nula**, y se representa por H_0 a la afirmación que quiero contrastar, en contraposición con la **hipótesis alternativa**, representada por H_1 .

El proceso de contrastar me dirá si debo aceptar o rechazar la hipótesis nula. Si los valores muestrales difieren mucho de los teóricos que se obtendrían de ser H_0 cierta, rechazaríamos la hipótesis, diciendo que las diferencias son estadísticamente significativas.

Las hipótesis estadísticas pueden ser de dos tipos: paramétricas (hipótesis sobre los valores de un parámetro de la población); no paramétricas (hipótesis sobre otros criterios, como por ejemplo criterio de independencia entre variables, etc.)

En este tema nosotros vamos a ceñirnos a las paramétricas. Contrastaremos medias, varianzas y proporciones.

La hipótesis alternativa no tiene que ser necesariamente complementaria de la hipótesis nula. Aclaremos esta afirmación con un ejemplo.

Supongamos que estoy estudiando que la proporción de recién nacidos varones en un determinado hospital es del 56%. La hipótesis nula sería $H_0 =$ "La proporción $p = 0,56$ ", mientras que la alternativa sería $H_1 =$ "La proporción $p \neq 0,56$ ". En este caso se llama **hipótesis alternativa bilateral** porque considera los valores menores y mayores que 0,56.

Pero si lo que contrasto es por ejemplo que el peso medio de las latas de una fabrica de conserva es de 1000 gr. como indica la información al consumidor, solo me interesaría como hipótesis alternativa que el peso fuese menor que 1000 gr, ya que si fuese mayor no haría más que reforzar la hipótesis nula que interesa al consumidor. En este caso la hipótesis nula sería $H_0 =$ "El peso medio de las latas es de 1000 gr." y la alternativa sería $H_1 =$ "El peso medio sería inferior a 1000 gr.", en cuyo caso estaríamos ante una **hipótesis alternativa unilateral** (en este caso de cola a la izquierda). La hipótesis nula deberá ser siempre una igualdad.

La hipótesis nula la mantendremos siempre hasta que los resultados nos digan lo contrario, lo cual conlleva una toma de decisiones que no siempre son acertadas. Esto provoca una clasificación en los errores que se establece del siguiente modo:

Error tipo I. El que se produce al rechazar la hipótesis nula cuando en realidad es cierta.

Error tipo II. El que se produce al no rechazar la hipótesis nula cuando esta no es cierta.

Llamaremos α a la probabilidad de cometer un error de tipo I, llamado también nivel de confianza de contraste estadístico.

Pasos para la construcción de un contraste de hipótesis

Se toma una muestra, obteniendo un valor numérico a partir de ella, para lo que se usará un estadístico pivote con distribución conocida.

Una vez establecido el nivel de significación del contraste, se divide la recta real en dos regiones (de **aceptación** y de **rechazo** de la hipótesis nula). Observaremos el valor numérico obtenido en el pivote aceptando o rechazando la hipótesis en función de si el valor cae o no en el intervalo de aceptación. Por este motivo el pivote usado se llama **medida de discrepancia**.

En todo caso, los pasos a seguir son los siguientes:

- 1º) Establecer claramente la hipótesis nula y la alternativa.
- 2º) Fijar el nivel de significación α , llamada también probabilidad de Error tipo I. Nos interesa que el error a cometer sea pequeño, por lo que α será de un valor próximo a 0 (0,05; 0,025;...etc)
- 3º) Elegir el estadístico pivote o medida de discrepancia, que dependerá del parámetro que estamos contrastando (aquí recurriremos al tema anterior de los intervalos de confianza)
- 4º) Determinar la región de aceptación y de rechazo.
- 5º) Tomar una muestra de la población calculando con ella el valor del estadístico pivote.
- 6º) Decidir si la hipótesis es aceptable o no en función de si el valor anterior cae en la región de aceptación o en la de rechazo.

Contraste de hipótesis para la media de una distribución normal

Hipótesis nula $H_0: \mu = \mu_0$

Estadísticos pivote:

$$d = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \in N(0,1) \text{ si conocemos la desviación típica o}$$

$$d = \frac{\bar{X} - \mu_0}{\frac{S_{n-1}}{\sqrt{n}}} \in t\text{-Student con } n-1 \text{ grados de libertad si desconocemos la } d, \text{ típica.}$$

Dependiendo de la hipótesis alternativa, podemos establecer el siguiente cuadro para las regiones de aceptación

Contraste	Hip. nula	Hip. alt.	Intervalo de aceptación	
			σ conocida	σ no conocida
Bilateral	$\mu = \mu_0$	$\mu \neq \mu_0$	$(-Z_{\alpha/2}, Z_{\alpha/2})$	$(-t_{\alpha/2}, t_{\alpha/2})$
Unilat. izquierda	$\mu = \mu_0$	$\mu < \mu_0$	$(-Z_{\alpha}, \infty)$	$(-t_{\alpha}, \infty)$
Unilat. derecha	$\mu = \mu_0$	$\mu > \mu_0$	$(-\infty, Z_{\alpha})$	$(-\infty, t_{\alpha})$

Ejemplo:

Una organización de consumidores está interesada en comprobar que el peso medio de una determinada marca de quesos es de 1000 gr. Para verificar esto, eligen al azar cinco quesos, obteniendo los siguientes pesos 992, 998, 990, 999, 1001. ¿Puede mantenerse la hipótesis de que la media es de 1000 gramos con un nivel de significación de 0,05?

Paso 1) Establecer sin ambigüedad la **hipótesis nula** y la **alternativa**.

$$H_0: \mu = 1000 \text{ gr.} \quad H_1: \mu < 1000$$

Paso 2) **Nivel de significación** o probabilidad de error tipo I $\alpha = 0,05$

Paso 3) **Estadístico pivote:** $d = \frac{\bar{X} - \mu_0}{S_{n-1} / \sqrt{n}} \in t - Student$ con 4 grados de libertad

Paso 4) **Región de aceptación:**

Como la hipótesis alternativa es bilateral izquierda, la región de rechazo está a la izquierda de la gráfica de la t-Student, con un valor de probabilidad 0,05. Por tanto hay que buscar en la tabla de la t-Student con 4 grados de libertad el valor de t que deja a la derecha un valor de probabilidad de 0,95, que se corresponde con -2,1318

Así pues el intervalo de aceptación es $(-2,1318, +\infty)$

Paso 5) **Medida de discrepancia:**

$$d = \frac{\bar{X} - \mu_0}{S_{n-1} / \sqrt{n}} = \frac{996 - 1000}{4,7434 / \sqrt{5}} = -1,89$$

Paso 6º) **Decisión final.** Como el valor del estadístico -1,89 está dentro de la región de aceptación, podemos aceptar la hipótesis nula, es decir que la media sea de 1000 gr.

Contraste para la varianza en una población normal

Hipótesis nula $H_0: \sigma^2 = \sigma_0^2$

Estadístico pivote:

$$d = \frac{(n-1)S_{n-1}^2}{\sigma_0^2} \in X_{n-1}^2 \text{ chi-cuadrado con } n-1 \text{ grados de libertad}$$

Dependiendo de la hipótesis alternativa, podemos establecer el siguiente cuadro para las regiones de aceptación

Contraste	Hip. nula	Hip. alt.	Intervalo de aceptación
Bilateral	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$(X^2_{1-\alpha/2}, X^2_{\alpha/2})$
Unilat. izquierda	$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(X^2_{1-\alpha}, +\infty)$
Unilat. derecha	$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$(0, X^2_{\alpha})$

Ejemplo:

Un cliente está dispuesto a demostrarle al dueño de una cafetería que la desviación típica de los churros que le sirven en el desayuno es superior a 1,5 cm con un nivel de significación de 0,01. Toma 20 churros y obtiene que la cuasi desviación típica es 2,3 cm. ¿Cómo realizará el contraste?

Paso 1) Establecer sin ambigüedad la **hipótesis nula** y la **alternativa**.

$$H_0: \sigma^2 = 2,25 \quad H_1: \sigma^2 > 2,25$$

Paso 2) **Nivel de significación** o probabilidad de error tipo I $\alpha = 0,01$

Paso 3) **Estadístico pivote:** $d = \frac{(n-1)S^2_{n-1}}{\sigma_0^2} \in X^2_{n-1}$ con 19 grados de libertad

Paso 4) **Región de aceptación:**

Como la hipótesis alternativa es bilateral derecha, la región de rechazo está a la derecha de la gráfica de la chi-cuadrado, con un valor de probabilidad 0,01. Por tanto hay que buscar en la tabla de la chi-cuadrado con 19 grados de libertad el valor de X^2 que deja a la derecha un valor de probabilidad de 0,01, que se corresponde con 36,191

Así pues el intervalo de aceptación es $(0, 36,191)$

Paso 5) **Medida de discrepancia:**

$$d = \frac{(n-1)S^2_{n-1}}{\sigma_0^2} = \frac{19 \cdot 5,29}{2,25} = 44,67$$

Paso 6º) **Decisión final.** Como el valor del estadístico 44,67 no está dentro de la región de aceptación, podemos rechazar la hipótesis nula, es decir que la varianza es mayor que 2,25 y en consecuencia la desviación típica mayor que 1,5 que es lo que pretendía demostrar el cliente.

Contraste de hipótesis para una proporción

Hipótesis nula $H_0: p = p_0$

Estadísticos pivote:

$$d = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \in N(0,1)$$

Dependiendo de la hipótesis alternativa, podemos establecer el siguiente cuadro para las regiones de aceptación

Contraste	Hip. nula	Hip. alt.	Intervalo de aceptación
Bilateral	$p=p_0$	$p \neq p_0$	$(-Z_{\alpha/2}, Z_{\alpha/2})$
Unilat. izquierda	$p=p_0$	$P < p_0$	$(-Z_{\alpha}, \infty)$
Unilat. derecha	$p=p_0$	$p > p_0$	$(-\infty, Z_{\alpha})$

Ejemplo:

Un partido político está interesado en contrastar si revalidará la mayoría absoluta en las próximas elecciones. Des resultado de una encuesta obtiene el apoyo de 550 de los 1000 votantes de que constaba la muestra. Realícese el contraste de hipótesis oportuno con un nivel de significación de 0,01.

Paso 1) Establecer sin ambigüedad la **hipótesis nula** y la **alternativa**.

$H_0: p=0,5$. $H_1: p>0,5$

Paso 2) **Nivel de significación** o probabilidad de error tipo I $\alpha=0,01$

Paso 3) **Estadístico pivote:**
$$d = \frac{p' - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \in N(0,1)$$

Paso 4) **Región de aceptación:**

Como la hipótesis alternativa es bilateral derecha, la región de rechazo está a la derecha de la gráfica de la $N(0,1)$, con un valor de probabilidad 0,01. Por tanto hay que buscar en la tabla de la normal el valor de z que deja a la derecha un valor de probabilidad de 0,01, o lo que es lo mismo, a la izquierda un valor de probabilidad de 0,99 que se corresponde con 2,33

Así pues el intervalo de aceptación es $(-\infty, 2,33)$

Paso 5) **Medida de discrepancia:**

$$d = \frac{p' - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0,55 - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{1000}}} = 3,16$$

Paso 6º) **Decisión final.** Como el valor del estadístico 3,16 no está dentro de la región de aceptación, podemos rechazar la hipótesis nula, es decir que la proporción es mayor que 0,5 y en consecuencia el partido político va a revalidar la mayoría.

Se pueden encontrar problemas propuestos y resueltos de Contrastes de hipótesis en <http://www.iesxunqueira1.com> en la sección “Taboleiro/departamento”.