Pruebas de hipótesis

NATALIA ACEVEDO PRINS

Natalia.acevedop@udea.edu.co

Contrastes de hipótesis de una población

Métodos para contrastar hipótesis que nos permiten contrastar la validez de una conjetura o de una afirmación utilizando datos muestrales

Procedimiento para **probar la validez** de un enunciado relativo a un parámetro poblacional.

Hipótesis: Afirmación relativa a un parámetro de la población sujeta a verificación.

Prueba de hipótesis: Procedimiento basado en evidencia de la muestra y la teoría de la probabilidad para determinar si la hipótesis es una afirmación razonable

Proceso para probar una hipótesis

1. **Establecer las hipótesis**
   - H_0
   - H_1
2. **Selección del nivel de significancia**
   - Error tipo I
   - Error tipo II
3. **Seleccionar el Estadístico de prueba**
   - A partir de la muestra
4. **Formular regla de decisión**
   - Valor crítico
5. **Tomar la decisión**
   - Comparación con el valor crítico

Definición de las hipótesis

Hipótesis nula H_0:
Hipótesis que se mantiene que es verdadera, a menos que se obtenga suficiente evidencia en contra.

Hipótesis alternativa H_1 :
Hipótesis frente a la que se contrasta la hipótesis nula y que se mantiene que es verdadera si se declara que la hipótesis nula es falsa.

Hipótesis simple:
Hipótesis que especifica un único valor para un parámetro poblacional de interés.

Hipótesis compuesta:
Hipótesis que especifica un rango de valores para un parámetro poblacional.

Definición de las hipótesis

Hipótesis alternativa unilateral:

Hipótesis alternativa que implica todos los valores posibles de un parámetro poblacional a un lado o al otro (es decir, mayores o menores que) del valor especificado por una hipótesis nula simple.

Hipótesis alternativa bilateral:

Hipótesis alternativa que implica todos los valores posibles de un parámetro poblacional distintos del valor especificado por una hipótesis nula simple (es decir, tanto mayores como menores que este valor).

Definición de errores

Decisiones sobre H0	Estados de la Naturaleza	
	H0 Verdadera	H0 Falsa
No rechazar H0	Decisión correcta $Probabilidad = 1 - \alpha$	Error tipo II $Probabilidad = \beta$
Rechazar H0	Error Tipo I $Probabilidad = \alpha$	Decisión correcta $Probabilidad = 1 - \beta$

Error de Tipo I:
Es la probabilidad de rechazar la hipótesis nula cuando esta es verdadera.

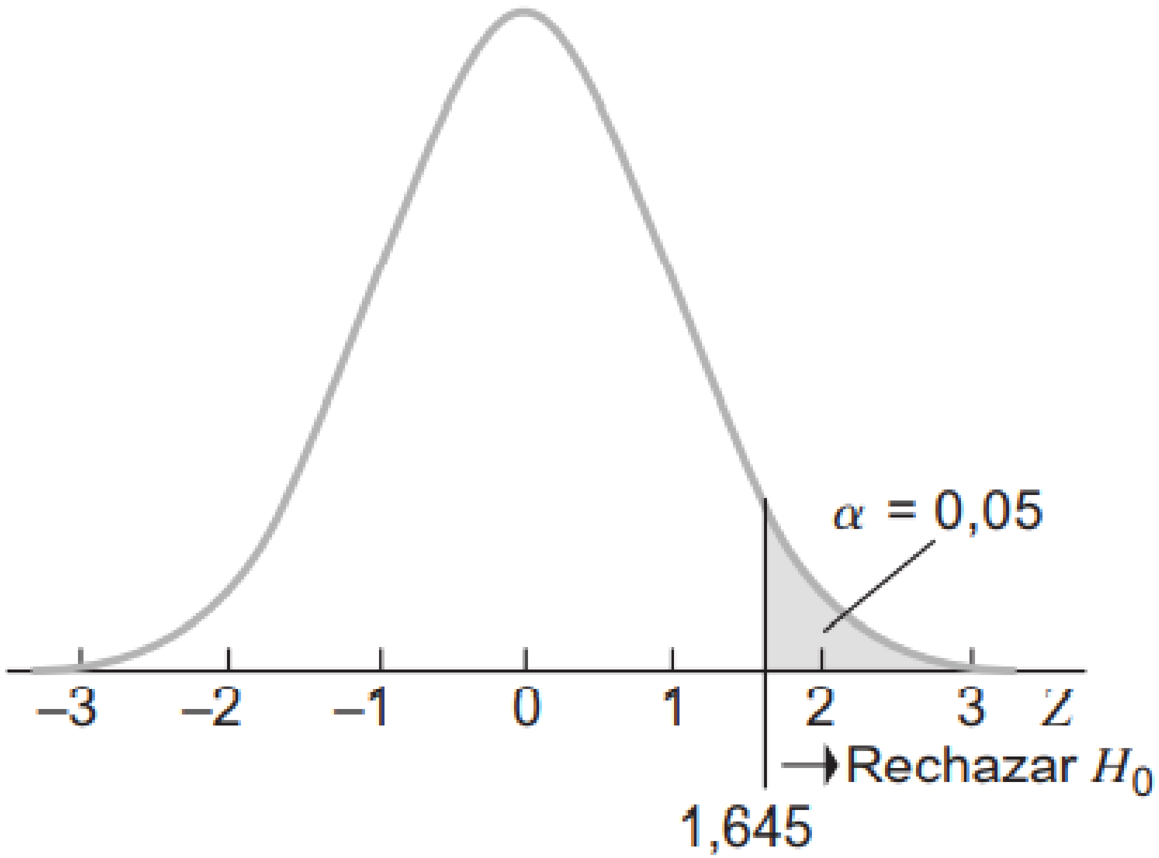
Error de Tipo II:
cuando no se rechaza una hipótesis nula falsa.

Tipos de contrastes

- Contraste de la media cuando se conoce σ en muestras grandes
- Contrastes de la media de una distribución normal: varianza poblacional desconocida
- Contrastes de la proporción poblacional (grandes muestras)
- Contrastes de la varianza de una distribución normal
- Contrastes de hipótesis de dos poblaciones

Contraste de la media cuando se conoce σ en muestras grandes

Para la prueba de hipótesis de la media μ, cuando se conoce σ o el tamaño de la muestra es grande entonces el estadístico es:



$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Contraste de la media cuando se conoce σ

Tenemos una muestra aleatoria de n observaciones procedentes de una población que sigue una distribución normal de media μ y varianza conocida σ^2. Calcule la media muestral $\bar{x}$. Se obtiene un contraste con un nivel de significación a de la hipótesis nula

Hipótesis Nula:

$$H_0\colon \mu = \mu_0$$

Hipótesis alternativa:

$$H_1\colon \mu > \mu_0$$

Regla de decisión:

$$Rechazar\ H_0\, si\ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$$

Ejemplo

Supongamos que el director de la fábrica quiere averiguar si el verdadero peso medio de las cajas de cereales es de más de 16 onzas.

Las normas del sector dicen que si la media poblacional del peso de las cajas es de 16,1 onzas o menos en una población de cajas que indican que su peso es de 16 onzas, el fabricante pagará una cuantiosa multa. Por tanto, nuestro objetivo es conseguir pruebas contundentes de que el peso medio de las cajas, μ, es superior a 16,1 onzas. En este caso, nuestra hipótesis nula sería:

$$H_0\colon\ \mu = \mu_0 = 16{,}1$$

Hipótesis Alternativa:

$$H_1\colon\ \mu > \mu_0 = 16{,}1$$

Nuestro contraste de la media poblacional utiliza la media muestral $\bar{x}$. Si la media muestral es considerablemente superior a $\mu_0 = 16{,}1$ entonces rechazamos la hipótesis nula.

Regla de decisión:

$$Z_{\alpha=0{,}05} = 1{,}645 \approx 1{,}65$$

$$Rechazar H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > 1{,}65$$

Valor p
otra forma de contraste

El *p-valor* es la probabilidad de obtener un valor del estadístico del contraste igual de extremo o más que el valor efectivo obtenido cuando la hipótesis nula es verdadera.

Es el menor nivel de significación al que se puede rechazar una hipótesis nula

Consideremos una muestra aleatoria de n observaciones procedente de una población que sigue una distribución normal de media μ y desviación típica σ, y la media muestral calculada resultante, $\bar{x}$. Se contrasta la hipótesis nula

El p-valor del contraste es:

$$H_0\colon \mu = \mu_0$$

$$H_1\colon \mu > \mu_0$$

$$p - value = P\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq Zp \mid H_0\colon \mu = \mu_0\right)$$

Ejemplo Valor -p

El director de producción de Ventanas Norte, S.A., le ha pedido que evalúe un nuevo método propuesto para producir su línea de ventanas de doble hoja. El proceso actual tiene una producción media de 80 unidades por hora con una desviación típica poblacional de $\boldsymbol{\sigma = 8}$. El director no quiere sustituirlo por el nuevo método, a menos que existan pruebas contundentes de que el **nivel medio de producción es mayor con ese nuevo método**. Para el proceso Obtenemos una muestra aleatoria de $n = 25$ horas de producción utilizando el nuevo método propuesto y una media muestral de $\bar{x} = 83$.

El director solo adoptará el nuevo método si existen pruebas contundentes a su favor. Por tanto, las hipótesis son:

$$H_0:\ \mu \leq 80$$

$$H_1:\ \mu > 80$$

Ejemplo Valor -p

La regla de decisión es:

$$Rechazar\ H_0\ si\ \frac{\bar{x}-80}{8/\sqrt{25}} > 1.65$$

$$Z_{\alpha=0,05} = 1.65$$

$$z = \frac{83-80}{8/\sqrt{25}} = 1875 > 1.65$$

Dada esta media muestral, también podríamos calcular el p-valor de la forma siguiente:

$$p - value = P(Z > 1.875) = 0.03$$

Rechazaríamos la hipótesis nula y concluiríamos que tenemos pruebas contundentes para apoyar la conclusión de que el nuevo método aumenta la productividad.

Hipótesis alternativa bilateral

Hay algunos problemas en los que las desviaciones demasiado altas o demasiado bajas tienen la misma importancia.

En esas situaciones, consideramos el contraste de la hipótesis nula:

$$H_0:\ \mu = \mu_0$$

Frente a la hipótesis alternativa:

$$H_0:\ \mu \neq \mu_0$$

La regla de decisión:

$$Rechazar H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < Z_{\alpha/2} \quad \text{o} \quad Rechazar H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > Z_{\alpha/2}$$

Igualmente:

$$Rechazar\ H0\ si\ \bar{x} < \mu 0 - Z_{\frac{\alpha}{2}} \sigma/\sqrt{n} \qquad Rechazar\ H0\ si\ \bar{x} > \mu 0 + Z_{\frac{\alpha}{2}} \sigma/\sqrt{n}$$

Valor p para Hipótesis alternativa bilateral

Se pueden calcular los p-valores observando que la probabilidad de la cola correspondiente se duplicaría para reflejar un p-valor que se refiere a la suma de las probabilidades de la cola superior y la cola inferior para los valores positivos y negativos de Z. El p-valor correspondiente al contraste de dos colas es:

$$p - value = 2P\left(\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > Z_{p/2} | H_0: \mu = \mu_0\right)$$

Donde $Z_{p/2}$ es el valor normal estándar correspondiente a la menor probabilidad de rechazar la hipótesis nula en cualquiera de las dos colas de la distribución de probabilidad.

Ejemplo Valor p para Hipótesis alternativa bilateral

El director de producción de Circuitos Ilimitados le ha pedido ayuda para analizar un proceso de producción. Este proceso consiste en hacer taladros cuyos diámetros siguen una distribución normal que tiene una media poblacional de dos centímetros y una desviación típica poblacional de 0,06 centímetros. Una muestra aleatoria de 9 mediciones tenía una media muestral de 1,95 centímetros. Utilice un nivel de significación de $\alpha = 0.05$ para averiguar si la media muestral observada es excepción y, por tanto, se debe ajustar la taladradora.

Solución Ejemplo

En este caso, el diámetro podría ser demasiado grande o demasiado pequeño. Por tanto, realizamos un contraste de hipótesis de dos colas planteando las siguientes la hipótesis:

$$H_0\text{: } \mu = 2.0$$

$$H_0\text{: } \mu \neq 2.0$$

La regla de decisión es rechazar H_0 en favor de $H1$ si

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < Z_{\alpha/2} \quad \text{o} \quad \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < Z_{\alpha/2}$$

$$\frac{1.95 - 2.0}{0.06/\sqrt{9}} = -2.5$$

Como -2,50 es menor que -1,96, rechazamos la hipótesis nula y concluimos que es necesario ajustar la taladradora.

hipótesis nula y alternativa compuestas con σ conocida

El método adecuado para contrastar, a un nivel de significación α, la hipótesis nula

$$H_0\colon \mu \leq \mu_0$$

La hipótesis alternativa:

$$H_1\colon \mu > \mu_0$$

Es precisamente igual que el que se emplea cuando la hipótesis nula es $H_0\colon\ \mu = \mu_0$.

Además, los p-valores también se calculan exactamente de la misma forma.

hipótesis nula y alternativa compuestas o simples con σ conocida

El método adecuado para contrastar al nivel de significación a la hipótesis nula

$$H_0\colon \mu = \mu_0 \qquad \text{O} \qquad H_0\colon \mu \geq \mu_0$$

La hipótesis alternativa:

$$H_1\colon \mu < \mu_0$$

La regla de decisión:

$$Rechazar H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < -Z_\alpha$$

Se pueden calcular los p-valores utilizando las probabilidades de la cola inferior

Ejemplo

El director de producción de Rodamientos Niquelados, S.A. le ha pedido ayuda para evaluar un proceso modificado de producción de rodamientos. Cuando el proceso funciona correctamente, produce rodamientos cuyos pesos siguen una distribución normal de media poblacional cinco onzas y desviación típica poblacional 0,1 onzas. Se ha recurrido a un nuevo proveedor de materia prima para un lote reciente de producción y el director quiere saber si, como consecuencia del cambio, el peso medio de los rodamientos es menor. No hay razón alguna para sospechar que el nuevo proveedor plantea problemas, por lo que el director continuará recurriendo a él a menos que existan pruebas contundentes de que están produciéndose rodamientos de menor peso que antes. Obtenemos una muestra aleatoria de $n = 16$ observaciones y la media muestral es 4,962

Solución

Nos interesa saber si existen pruebas contundentes para concluir que están produciéndose rodamientos de menor peso. Por tanto, contrastamos la hipótesis nula:

$$H_0\colon \mu = \mu_0 = 5$$

Solo emprendemos acciones si se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

La hipótesis alternativa:

$$H_1\colon \mu < 5$$

La regla de decisión para el problema:

$$Rechazar H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -1.65 \qquad \frac{4.962 - 5}{0.1/\sqrt{16}} = -1.52$$

No rechazamos la hipótesis nula.

También podemos calcular el p-valor correspondiente

$$p - value = P(Z_p < -1.52) = 0.0643$$

Contrastes de la media de una distribución normal: varianza poblacional desconocida

Tenemos una muestra aleatoria de n observaciones procedentes de una población normal que tiene una media μ. Utilizando la media muestral y la desviación típica muestral, $\bar{x}$ y s, respectivamente, podemos utilizar 3 tipos de contrastes con el nivel de significación α.

1. Para contrastar cualquiera de las hipótesis nulas :

$$H_0\colon \mu = \mu_0 \quad \text{o} \quad H_0\colon \mu \leq \mu_0$$

Hipótesis alternativa

$$H_1\colon \mu > \mu_0$$

Regla de decisión

$$Rechazar\ H_0\ si\ t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$$

Contrastes de la media de una distribución normal: varianza poblacional desconocida

2. Para contrastar cualquiera de las dos hipótesis nulas

$$H_0\colon \mu = \mu_0 \qquad \text{o} \qquad H_0\colon \mu \geq \mu_0$$

Hipótesis alternativa

$$H_1\colon \ \mu < \mu_0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ t = \frac{\bar{x}\ - \mu_0}{s/_{\sqrt{n}}} < tn_{\ -\ 1,\ \alpha}$$

Contrastes de la media de una distribución normal: varianza poblacional desconocida

3. Para contrastar la hipótesis nula

$$H_0\text{: } \mu = \mu_0$$

Hipótesis alternativa

$$H_0\text{: } \mu \neq \mu_0$$

La regla de decisión

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < tn_{-1,\alpha} \quad \text{o} \quad Rechazar\ H_0\ si\ \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > tn_{-1,\alpha}$$

$t_{n-1,\alpha}$ es el valor de la t-student con n-1 grados de libertad y probabilidad $\alpha/2$

Los p-valores de estos contrastes se calculan de la misma forma que con varianza conocida, pero el valor de la Z normal se sustituye por el valor de la t-Student.

Ejemplo Contraste con varianza poblacional desconocida

Un productor de una amplia variedad de verduras congeladas, le ha pedido que averigüe si las ventas semanales de bolsas de brócoli congelado de 16 onzas han aumentado. En los seis últimos meses, se ha vendido una media semanal de 2.400 bolsas. Ha obtenido una muestra aleatoria de datos de ventas de 134 tiendas para realizar el estudio. Este estudio arroja que la media muestral es 3.593 y que la desviación típica muestral es 4.919.

Solución

Se contrasta la hipótesis nula de que la μ= 2.400 frente a la alternativa de que las ventas han aumentado utilizando un nivel de significación $\alpha = 0{,}05$. Las hipótesis son

$$H_0\text{: } \mu = 2400$$

$$H_1\text{: } \mu > 2400$$

$$t = \frac{3593 - 2400}{4919/_{134}} = \frac{3593 - 2400}{425} = 2.81$$

Con $t_{133,\, 0.05} = 1.645 < 2.81$ se rechaza H_0

Contrastes de la proporción poblacional (grandes muestras)

Supongamos que tenemos una muestra aleatoria de n observaciones procedentes de una población que tiene una proporción P cuyos miembros poseen un atributo especial.

Si $nP(1-P) > 5$ y la proporción muestral es $\hat{p}$, los siguientes 3 contrastes tienen el nivel de significación α.

1. Para contrastar cualquiera de las dos hipótesis:

$$H_0: P = P_0 \qquad H_0: P \leq P_0$$

Hipótesis alternativa:

$$H_1: P > P_0$$

La regla de decisión es:

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1-P_0)/_n}} > Z_\alpha$$

Contrastes de la proporción poblacional (grandes muestras)

2. Para contrastar una de las dos hipótesis nula siguientes:

$$H_0: P = P_0 \qquad \text{O} \qquad H_0: P \geq P_0$$

La hipótesis alternativa:

$$H_1: P < P_0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} < -Z_\alpha$$

Contrastes de la proporción poblacional (grandes muestras)

3. Para contrastar la hipótesis nula siguiente:

$$H_0: P = P_0$$

La hipótesis alternativa:

$$H_1: P \neq P_0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} < -Z_{\alpha/2}$$ O $$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} > Z_{\alpha/2}$$

En todos estos contrastes, el Valor-p es el nivel de significación más bajo al que se puede rechazar la hipótesis nula.

Ejemplo Contraste proporción poblacional

Una empresa de estudios de mercado quiere saber si los compradores son sensibles a los precios de los artículos que se venden en un supermercado. Obtiene una muestra aleatoria de 802 compradores y observa que 378 son capaces de decir cuál es el precio correcto de un artículo inmediatamente después de colocarlo en el carro. Contraste al nivel del 7 % la hipótesis nula de que al menos la mitad de todos los compradores son capaces de decir cuál es el precio correcto.

Solución

P es la proporción poblacional de compradores de los supermercados que es capaz de decir cuál es el precio correcto. El Contraste de las hipótesis son:

$$H_0 : P \geq P_0 = 0.5$$

$$H_1 : P < P_0 = 0.5$$

Solución

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} < -Z_\alpha$$

Para el ejemplo:

$$\hat{p} = \frac{378}{802} = 0.471$$

$$n = 802$$

$$\frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} = \frac{0.471 - 0.5}{\sqrt{0.5(1 - 0.5)/_{802}}} = -1.64$$

$$Z_{\alpha=0.07} = 1.474$$

Como -1.64 < -1.474 entonces rechazamos la hipótesis nula.

Concluimos que menos de la mitad de los compradores puede decir correctamente el precio

Contrastes de la varianza de una distribución normal

Tenemos una muestra aleatoria de n observaciones procedentes de una población que sigue una distribución normal que tiene una varianza σ^2. Si observamos la varianza muestral s^2, los siguientes 3 contrastes tienen el nivel de significación α.

1. Para contrastar cualquiera de la dos hipótesis nulas:

$$H_o\colon \sigma^2 = \sigma_0^2 \qquad \text{O} \qquad H_o\colon \sigma^2 \leq \sigma_0^2$$

La hipótesis alternativa:

$$H_1\colon \sigma^2 > \sigma_0^2$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{(n-1)s^2}{\sigma_0^2} > X^2_{n-1,\alpha}$$

Contrastes de la varianza de una distribución normal

2. Para contrastar cualquiera de las dos hipótesis nulas

$$H_o\colon \sigma^2 = \sigma_0^2 \qquad \text{O} \qquad H_o\colon \sigma^2 \geq \sigma_0^2$$

Hipótesis Alternativa

$$H_1\colon \sigma^2 < \sigma_0^2$$

La regla de decisión

$$Rechazar\ H_0\, si\ \frac{(n-1)s^2}{\sigma_0^2} < X^2_{n-1,\, 1-\alpha}$$

Contrastes de la varianza de una distribución normal

3. Para contrastar la Hipótesis Nula

$$H_o\colon \sigma^2 = \sigma_0^2$$

Hipótesis alternativa bilateral:

$$H_1\colon \sigma^2 \neq \sigma_0^2$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{(n-1)s^2}{\sigma_0^2} > X^2_{n-1,\,\alpha/2}$$

O

$$Rechazar\ H_0\ si\ \frac{(n-1)s^2}{\sigma_0^2} < X^2_{n-1,\,\alpha/2}$$

Donde X^2_{n-1} es una variable aleatoria ji-cuadrado y $P(X^2_{n-1} > X^2_{n-1,\alpha}) = \alpha$

El valor-p de estos contrastes es la probabilidad de obtener un valor al menos tan extremo como el obtenido, dada la hipótesis nula.

Ejemplo

El director de control de calidad de Industrias Químicas Asociadas le ha pedido que averigüe si la varianza de las impurezas de sus envíos de fertilizante está dentro de la norma establecida. Esta norma establece que la varianza de los kilos de impurezas de los sacos de 100 kilos no puede ser superior a 4. Se obtiene una muestra aleatoria de 20 sacos y se miden los kilos de impurezas de cada saco. Se calcula que la varianza muestral es 6,62. Utilice un nivel de confianza del 95%.

Solución

La Hipótesis Nula

$$H_o: \sigma^2 \leq \sigma_0^2 = 4$$

La Hipótesis Alternativa:

$$H_1: \sigma^2 >= 4$$

Ejemplo

La regla de decisión para un contraste de nivel de significación a es rechazar H_0 en favor de H_1 si:

$$\frac{(n-1)s^2}{\sigma_0^2} > X^2_{n-1,\alpha}$$

Para el ejemplo:

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{(20-1)6.62}{4} = 31.445 > X^2_{n-1,\alpha} = 30.144$$

Por tanto, rechazamos la hipótesis nula y concluimos que la variabilidad de las impurezas es superior a lo que establece la norma

El valor p:

$$p - value = P\left(\frac{20-1}{\sigma_0^2} > X^2_{19} = 31.444\right) = 0.036$$

$Vp < 0.05$ entonces rechazamos H_0

Valoración de la potencia de un contraste: Probabilidad de cometer error tipo II

Solo se puede cometer un error de Tipo II si la hipótesis alternativa es verdadera. Por tanto, consideraremos el error de Tipo II y la potencia que se dan cuando el parámetro poblacional adopta valores específicos que están incluidos en la hipótesis alternativa.

$$H_0:\ \mu = \mu_0$$

$$Rechazar\ H_0\ si = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > Z_\alpha$$

$$H_1:\ \mu > \mu_0$$

Esta es la probabilidad de cometer un error de Tipo II. Por tanto, consideramos una μ=μ* tal que μ*>μ0. Entonces, para μ*, la probabilidad de cometer un error de Tipo II es

$$\beta = P(\bar{x} < \bar{x}_c | \mu = \mu^*) = P\left(z < \frac{\bar{x} - \mu^*}{\sigma/\sqrt{n}}\right)$$

$$Potencia = 1 - \beta$$

Ejemplo

Consideremos un ejemplo en el que contrastamos H_0 de que la media poblacional (μ) del peso de los rodamientos de un proceso de producción es de 5 onzas frente H_1 de que es de más de 5 onzas. Tenemos una muestra aleatoria de 16 observaciones y un $\alpha = 0.05$. Se supone que la distribución poblacional es una distribución normal que tiene una desviación típica de 0,1 onzas.

$$H_0: \mu = 5$$

$$H_0: \mu > 5$$

La regla de decisión es

$$Rechazar H_0 \text{ si } \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > Z_\alpha$$

Para el ejemplo:

$$\frac{\bar{x} - 5}{0.1/\sqrt{16}} > 1.645$$

Despejando $\bar{x}$:

$$\bar{x} > 5 + 1.645\left(0.1/\sqrt{16}\right) > 5.041$$

Si la media muestral es inferior o igual a 5,041, entonces, acetaríamo H_0. Pero para medias mayores rechazaríamos H_0 y Aceptaríamos H_1

Ejemplo

Supongamos que queremos hallar la probabilidad de que no se rechace la hipótesis nula si el verdadero peso medio es de 5,05 onzas.

Con esta media la hipótesis alternativa sería correcta, y queremos hallar la probabilidad de que no rechacemos la hipótesis nula, por tanto, que cometamos un error de Tipo II. Es decir, queremos hallar la probabilidad de que la media muestral sea de menos de 5,041 si la media poblacional es realmente 5,05.

Tenemos que:

$$\beta = P(\bar{x} \leq \bar{x}_c | \mu = \mu^*) = P\left(z \leq \frac{\bar{x} - \mu^*}{\sigma/\sqrt{n}}\right)$$

Para el ejemplo:

$$\beta = P(\bar{x} \leq 5.041 | \mu = 5.05) = P\left(z \leq \frac{5.041 - 5.05}{0.1/\sqrt{16}}\right) = P(z \leq -0.36) = 0.3594$$

Por tanto la potencia es:

$$Potencia = 1 - \beta = 1 - 0.3594 = 0,6406$$

Potencia de contrastes de proporciones poblacionales

También podemos hallar la probabilidad de cometer un error de Tipo II en los contrastes de proporciones. La probabilidad, β, de cometer un error de Tipo II dada una proporción poblacional P_1 incluida en H_1 se halla así:

Partiendo de la regla de decisión del contraste, se halla el rango de valores de la proporción muestral que llevan a no rechazar la hipótesis nula.

Con el valor P_1 para la proporción poblacional (donde P_1 está incluida en la hipótesis alternativa) se halla la probabilidad de que la $\hat{p}$ esté en el intervalo de no rechazo del paso 1 para muestras de n observaciones.

Ejemplo

El presidente de Inversores Electrónicos le ha pedido que analice las predicciones de los beneficios empresariales por acción realizadas por un grupo de analistas financieros. Estos analistas tenían interés en saber:

¿Cuál era la proporción de predicciones que eran superiores al nivel efectivo de beneficios ?
¿Cuál era la proporción de predicciones que eran inferiores al nivel efectivo de beneficios?

Sabiendo que Se obtiene una muestra aleatoria de $n = 600$ predicciones y se averigua que 382 son superiores a los beneficios efectivos. El nivel de significación es : $\alpha = 0.05$

Solución

Comencemos asumiendo que las proporciones son diferentes del 50 %. Si P es la proporción de predicciones superiores al nivel efectivo entonces:

$$H_0: P = P_0 = 0.5$$

$$H_1: P \neq 0.5$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} < -Z_{\alpha/2}$$

O

$$Rechazar\ H_0\ si\ \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} > Z_{\alpha/2}$$

Para nuestro ejemplo H_0 se rechaza si:

$$\frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} < -1.96 \quad o. \quad \frac{\hat{p} - P_0}{\sqrt{P_0(1 - P_0)/_n}} > 1.96$$

H_0 también se rechaza si:

$$\hat{p} < 0.5 - 1.96\sqrt{\frac{0.5(1 - 0.5)}{600}} = 0.5 - 0.04 = 0.46$$

O

$$\hat{p} > 0.5 + 0.04 = 0.54$$

Solución

Si la proporción muestral observada es $\hat{p} = \frac{382}{600} = 0.637$

se rechaza H_0 al nivel de 5 %.

Ahora queremos hallar la probabilidad de cometer un error de Tipo II. Supongamos que la verdadera proporción poblacional es $P1 = 0{,}55$. Queremos hallar la probabilidad de que la proporción muestral se encuentre entre 0,46 y 0,54 si la proporción poblacional es 0,55.

$$\beta = P(0.46 \leq \hat{p} \leq 0.54) = P\left[\frac{0.46 - P_1}{\sqrt{\frac{P_1(1-P_1)}{n}}} \leq Z \leq \frac{0.54 - P_1}{\sqrt{\frac{P_1(1-P_1)}{n}}}\right]$$

$$\beta = P\left[\frac{0.46 - 0.55}{\sqrt{\frac{0.55(1-0.55)}{600}}} \leq Z \leq \frac{0.54 - 0.55}{\sqrt{\frac{0.55\,(1-0.55)}{600}}}\right]$$

$$\beta = P(-4.43 \leq Z \leq 0.49) = 0.3121$$

Cometer un error de Tipo II si no se rechaza H_0 cuando la verdadera proporción es 0,55 es $\beta = 0.3121$

Contrastes de hipótesis de dos poblaciones: muestras dependientes

Supongamos que tenemos una muestra aleatoria de n pares de observaciones enlazadas de distribuciones que tienen las medias μ_x y μ_y. Sean d_1 y sd la media muestral y la desviación típica muestral observadas de las n diferencias $(x_i - y_i)$. Si la distribución poblacional de las diferencias es una distribución normal, los siguientes 3 contrastes tienen un nivel de significación α:

1. Para contrastar cualquiera de las dos hipótesis:

$$H_0: \mu_x - \mu_y = 0 \qquad \text{o} \qquad H_0: \mu_x - \mu_y < 0$$

Hipótesis alternativa:

$$H_1: \mu_x - \mu_y > 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{d}}{s_d/\sqrt{n}} > t_{n-1,\alpha}$$

Contrastes de hipótesis de dos poblaciones: muestras dependientes

2. Para contrastar cualquiera de las dos hipótesis:

$$H_0\!: \mu_x - \mu_y = 0 \qquad \text{o} \qquad H_0\!: \mu_x - \mu_y \geq 0$$

Hipótesis alternativa:

$$H_1\!: \mu_x - \mu_y < 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{d}}{s_d/\sqrt{n}} < t_{n-1,\alpha}$$

Contrastes de hipótesis de dos poblaciones: muestras dependientes

3. Para contrastar la hipótesis:

$$H_0: \mu_x - \mu_y = 0$$

Hipótesis alternativa bilateral:

$$H_1: \mu_x - \mu_y \neq 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{d}}{s_d/\sqrt{n}} < -t_{n-1,\alpha/2} \qquad \text{o} \qquad Rechazar\ H_0\ si\ \frac{\bar{d}}{s_d/\sqrt{n}} > t_{n-1,\alpha/2}$$

t_{n-1} sigue una distribución t-Student con $(n-1)$ grados de libertad.

Los Valores-p de todos estos contrastes son la probabilidad de obtener un valor al menos tan extremo como el obtenido, dada H_0.

Contrastes de la diferencia entre dos medias poblacionales normales: muestras independientes

Supongamos que tenemos muestras aleatorias independientes de n_x y n_y observaciones procedentes de distribuciones normales que tienen las medias μ_x y μ_y y las varianzas σ_x^2 y σ_y^2 respectivamente. Si las medias muestrales observadas son $\bar{x}$ e $\bar{y}$, entonces los siguientes 3 contrastes tienen un nivel de significación α.

1. Contrastar cualquiera de las siguiente 2 hipótesis:

$$H_0: \mu_x - \mu_y = 0 \qquad H_0: \mu_x - \mu_y \leq 0$$

Hipótesis alternativa:

$$H_1: \mu_x - \mu_y > 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} > Z_\alpha$$

Contrastes de la diferencia entre dos medias poblacionales normales: muestras independientes

2. Contrastar cualquiera de las siguiente 2 hipótesis:

$$H_0: \mu_x - \mu_y = 0 \qquad\qquad H_0: \mu_x - \mu_y \geq 0$$

Hipótesis alternativa:

$$H_1: \mu_x - \mu_y < 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} < -Z_\alpha$$

Contrastes de la diferencia entre dos medias poblacionales normales: muestras independientes

3. Contrastar la hipótesis:

$$H_0\colon \mu_x - \mu_y = 0$$

Hipótesis alternativa bilateral:

$$H_1\colon \mu_x - \mu_y \neq 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} < -Z_{\alpha/2} \quad \text{o} \quad Rechazar\ H_0\ si\ \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > Z_{\alpha/2}$$

Si los tamaños de las muestras son grandes (n>100), se puede n aproximar las varianzas poblacionales por las varianzas muestrales. Por teorema del límite central se obtienen buenas aproximaciones aunque las poblaciones no sean con distribución normal.

Dos medias, muestras independientes con σ desconocida que se supone que son iguales

En estos contrastes, se supone que tenemos una muestra aleatoria independiente de tamaño n_x y n_y observaciones extraídas de poblaciones que siguen una distribución normal que tiene las medias μ_x y μ_y y una varianza común. Se utilizan las varianzas muestrales s_x^2 y s_y^2 para calcular un estimador agrupado de la varianza:

Media ponderada de las dos varianzas muestrales

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{(n_x - n_y - 2)}$$

Utilizando las medias muestrales observadas $\bar{x}$ e $\bar{y}$, los siguientes 3 contrastes tienen un nivel de significación α:

Dos medias, muestras independientes con σ desconocida que se supone que son iguales

1. Para contrastar cualquiera de las dos hipótesis nulas:

$$H_0: \mu_x - \mu_y = 0 \qquad\qquad H_0: \mu_x - \mu_y \leq 0$$

La hipótesis alternativa:

$$H_1: \mu_x - \mu_y > 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} > t_{n_x+n_y-2,\alpha}$$

Dos medias, muestras independientes con σ desconocida que se supone que son iguales

2. Para contrastar cualquiera de las dos hipótesis nulas:

$$H_0: \mu_x - \mu_y = 0 \qquad\qquad H_0: \mu_x - \mu_y \geq 0$$

La hipótesis alternativa:

$$H_1: \mu_x - \mu_y < 0$$

La regla de

decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} < -t_{n_x+n_y-2,\alpha}$$

Dos medias, muestras independientes con σ desconocida que se supone que son iguales

3. Para contrastar la hipótesis nula:

$$H_0: \mu_x - \mu_y = 0$$

La hipótesis alternativa bilateral:

$$H_1: \mu_x - \mu_y \neq 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} < -t_{n_x+n_y-2,\alpha/2} \qquad Rechazar\ H_0\ si\ \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} > t_{n_x+n_y-2,\alpha/2}$$

Aquí $t_{n_x+n_y-2,\alpha}$ es el número para el que

$$P(\, t_{n_x+n_y-2} > t_{n_x+n_y-2,\alpha}) = \alpha$$

Obsérvese que los grados de libertad de la t-Student son $n_x + n_y - 2$ para todos estos contrastes.

Dos medias, muestras independientes con σ desconocida que no se suponen iguales

Estos contrastes suponen que tenemos muestras aleatorias independientes de n_x y n_y observaciones procedentes de poblaciones normales que tienen las medias μ_x y μ_y y varianzas desiguales. Se utilizan las varianzas muestrales s_x^2 y s_y^2. El número de grados de libertad v del estadístico t-Student viene dado por

$$v = \frac{\left[\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right]^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{(n_x - 1)} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{(n_y - 1)}}$$

Utilizando las medias muestrales observadas $\bar{x}$ e $\bar{y}$ los siguientes 3 contrastes tienen un nivel de significación α:

Dos medias, muestras independientes con σ desconocida que no se suponen iguales

1. Para contrastar cualquiera de las dos hipótesis nulas:

$$H_0: \mu_x - \mu_y = 0 \qquad\qquad H_0: \mu_x - \mu_y \leq 0$$

La hipótesis alternativa:

$$H_1: \mu_x - \mu_y > 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} > t_{v,\alpha}$$

Dos medias, muestras independientes con σ desconocida que no se suponen iguales

2. Para contrastar cualquiera de las dos hipótesis nulas:

$$H_0: \mu_x - \mu_y = 0 \qquad\qquad H_0: \mu_x - \mu_y \geq 0$$

La hipótesis alternativa:

$$H_1: \mu_x - \mu_y < 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} < -t_{v,\alpha}$$

Dos medias, muestras independientes con σ desconocida que no se suponen iguales

3. Para contrastar la hipótesis nula:

$$H_0: \mu_x - \mu_y = 0$$

La hipótesis alternativa:

$$H_1: \mu_x - \mu_y \neq 0$$

La regla de decisión:

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} < -t_{v,\alpha/2}$$

o

$$Rechazar\ H_0\ si\ \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} > t_{v,\alpha/2}$$

Aquí, $t_{v,\alpha}$ es el número para el que

$$P(t_v > t_{v,\alpha}) = \alpha$$

Contrastes de la igualdad de las varianzas entre dos poblaciones distribuidas normalmente

Tenemos dos muestras aleatorias independientes con n_x y n_x observaciones procedentes de dos poblaciones normales que tienen las varianzas σ_x^2y σ_y^2 . Si las varianzas muestrales son s_x^2 y s_y^2 , entonces la variable aleatoria

$$F = \frac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2}$$

Sigue una distribución F con $(n_x - 1)$ grados de libertad en el numerador y $(n_y - 1)$ grados de libertad en el denominador.

Una distribución F con v_1 grados de libertad en el numerador y v_2 grados de libertad en el denominador se representa de la forma siguiente: $F_{v_1.v_2}$. $F_{v_1.v_{2,\alpha}}$ es el número para el que

$$P(F_{v_1.v_2} > F_{v_1.v_{2,\alpha}}) = \alpha$$

Este contraste es muy sensible al supuesto de normalidad

Contrastes de la igualdad de las varianzas entre dos poblaciones distribuidas normalmente

Sean s_x^2 y s_y^2 las varianzas muestrales observadas de muestras aleatorias independientes de tamaño n_x y n_y procedentes de poblaciones distribuidas normalmente que tienen las varianzas σ_x^2 y σ_y^2 . Sea s_x^2 la varianza mayor. En ese caso, los siguientes 2 contrastes tienen un nivel de significación a.

1. Para contrastar cualquiera de las dos hipótesis nulas:

$$H_0 = \sigma_x^2 = \sigma_y^2 \qquad\qquad H_0 = \sigma_x^2 \leq \sigma_y^2$$

La hipótesis alternativa:

$$H_1 = \sigma_x^2 > \sigma_y^2$$

La regla de decisión:

$$Rechazar\ H_0\ si\ F = \frac{s_x^2}{s_y^2} > F_{n_x-1,\,n_y-1,\alpha}$$

Contrastes de la igualdad de las varianzas entre dos poblaciones distribuidas normalmente

2. Para contrastar la hipótesis nula:

$$H_0 = \sigma_x^2 = \sigma_y^2$$

La hipótesis alternativa bilateral:

$$H_1 = \sigma_x^2 \neq \sigma_y^2$$

La regla de decisión:

$$Rechazar\ H_0\ si\ F = \frac{s_x^2}{s_y^2} > F_{n_x-1,\,n_y-1,\alpha/2}$$

Donde s_x^2 es la mayor de las dos varianzas muestrales. Dado que cualquiera de las dos varianzas muestrales podría ser mayor, esta regla se basa en un contraste de dos colas por tanto, utilizamos $\alpha/2$ como la probabilidad de la cola superior.

Dada la complejidad de la distribución F, solo se calculan los valores críticos.

Normalmente los valores p se calculan utilizando un paquete estadístico.