Intro a la Probabilidad y estadística

Martes y Jueves Aula B17

Dra Ana Georgina Flesia

Nociones de test de hipótesis

- Una industria quesera familiar compra leche a varios proveedores como el principal ingrediente del queso.
- El gerente de producción sospecha que algunos proveedores están agregando agua a la leche para incrementar sus ganancias.
- Se sabe que el exceso de agua puede ser detectado determinando el punto de congelamiento de la leche.

Nociones de test de hipótesis

- La temperatura a la cual la leche se congela varía normalmente con media $\mu = -0.540C^0$ y desviación estándar de $0.008C^0$ y agregar agua eleva el punto de congelamiento de la leche.
- En cambio un punto de congelamiento menor puede ser resultado de exceso de crema en la leche.

Nociones de test de hipótesis

- El técnico del laboratorio de la compañía analiza 5 lotes consecutivos de leche de un mismo proveedor y encuentra que el punto de congelamiento medio de la muestra es de $-0.538C^0$.
- ¿Es ésta evidencia substancial de que el productor está agregando agua a la leche?

Nociones de test de hipótesis

- Supongamos que queremos decidir si una moneda es honesta. Si la lanzamos 100 veces y obtenemos 87 caras, es probable que la moneda sea sesgada?¿Qué hay de 61 caras? ¿O 53 caras?
- La mayoría de la gente diría que 87 caras es evidencia fuerte de que la moneda es sesgada, mientras que 53 caras no constituye ninguna prueba. Por otra parte 61 caras es menos claro.

Nociones de test de hipótesis

Para responder estas preguntas, usualmente se plantea la posición del escéptico, que dice que las variaciones observadas son solo producto del azar, y la del investigador que argumenta a favor de un cambio real. En nuestro ejemplos, un crítico escéptico se preguntaría

- ¿cuál es la probabilidad de obtener una muestra como la observada si la moneda no es honesta?
- ¿cuál es la probabilidad de obtener una muestra como la observada si la leche no fue aguada?

Nociones de test de hipótesis

Si estas probabilidades resultan muy chicas, podrían ocurrir dos cosas

- En ambos casos, se tuvo mucha mala suerte y se observo una muestra muy poco común, pero la leche no esta aguada y la moneda esta balanceada.
- La leche fue aguada y la moneda no es balanceada y se obtuvo un resultado probable de la distribución del número de caras (de una moneda desbalanceada) y se observó un resultado esperable de la distribución del punto de congelamiento de ese tipo de leche adulterada.

El escéptico debe decidir que probabilidad se considera suficientemente chica, pues la frecuencia con que se encuentran muestras inusuales depende de esa probabilidad.

Lógica del test de hipótesis

El razonamiento que se emplea es el siguiente, deseamos saber si la hipótesis de estudio, llamada hipótesis alternativa, es verdadera.

- La hipótesis nula dice que no hay efecto aparte del que podría ser producido por el azar.
- La hipótesis alternativa dice que hay un efecto aparte del que podría ser producido por el azar.

Lógica del test de hipótesis

- Para probar que hay alguna clase de efecto, hay que probar falsa la posibilidad de que los resultados puedan haber sido producidos por el azar.
- Hay que probar falsa la hipótesis nula para validar la alternativa. Hay que convencer al escéptico.

Lógica del test de hipótesis

- El planteo del test de hipótesis tiene un axioma que hay aceptar, se supone que **siempre ocurre lo más probable**, lo cual no siempre es cierto.
- Si uno acepta esto, entonces consideraríamos como verdadera la hipótesis más probable.

¿Como planteamos las hipótesis?

- La hipótesis nula,llamada H_0, representa la idea de "no hay efecto".
- La hipótesis alternativa H_A representa la idea "hay efecto".
- Para poder cuantificar las probabilidades de ocurrencia, debemos expresarlas a través de un parámetro poblacional estimable.
- La decisión se tomará a partir del valor del estadístico que estima el parámetro (al cual debemos conocerle su distribución de probabilidad exacta o aproximada si es válida la hipótesis nula).

¿Como tomamos la decisión?

- Cuando H_0 es verdadera, esperamos que el estimador tome valores cerca del parámetro representado por H_0.
- Si la probabilidad de observar un valor al menos tan extremo como el observado, bajo hipótesis nula, es muy chico, (esto es, estoy lejos del parámetro verdadero) entonces consideramos que hay evidencia en contra de la hipótesis nula.

Ejemplo de la leche

- Tratemos de definir las hipótesis de estudio en el ejemplo de la posible leche adulterada.
- La sospecha del gerente es la hipótesis alternativa, y la hipótesis nula es la del escéptico, que no vé cambio en la leche comprada.
- Para precisar estas hipótesis en términos de parámetros estimables, observamos que el punto de congelamiento de la leche es una variable que puede ayudar a discriminar si la leche fue adulterada.
- Agregar agua eleva el punto de congelamiento, mientras que una leche con mucha crema tiene un punto de congelamiento menor a lo usual.

Ejemplo de la leche

- El punto de congelamiento usual no es fijo, es una variable aleatoria con distribución normal, con media $\mu = -0.545C^0$ y desviación estándar de $0.008C^0$.
- Por lo cual un cambio a favor del adulteramiento con agua seria observar un valor correspondiente a una distribución normal con una media mayor a la usual:

$$H_0 : \mu = -0.545 \qquad vs \qquad H_A : \mu > -0.545$$

Ejemplo de la leche

- El técnico del laboratorio de la compañía analiza 5 lotes consecutivos de leche de un mismo proveedor y encuentra que el punto de congelamiento medio de la muestra es de $-0.538C^0$.
- ¿Es ésta evidencia substancial de que el productor está agregando agua a la leche?

Ejemplo de la leche

- Si definimos la región de rechazo como

$$RR = \{x/x \geq -0.538\},$$

el error cometido al rechazar la hipótesis nula en favor de la hipótesis de leche adulterada, cuando en realidad H_0 era verdadera, es el área bajo una curva normal a la derecha del valor -0.538,

$$P(\overline{X} \geq -0.538)$$

En ese caso, el estimador $\overline{X}$ es una variable normal con media $\mu = -0.545$ y desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{5} = 0.008/\sqrt{5}$

$$\begin{aligned} P(\overline{X} \geq -0.538) &= P\left[\frac{\overline{X} - (-0.545)}{0.008/\sqrt{5}} \geq \frac{-0.538 - (-0.545)}{0.008/\sqrt{5}}\right] \\ &= P(Z \geq 1.96) \\ &= 1 - 0.975 = 0.025 \end{aligned}$$

Ejemplo de la leche

- Esto quiere decir que una temperatura de congelamiento media como la observada solo ocurre 2 veces y media cada 100 mediciones de leche no adulterada,
- Por lo cual hay evidencia de que la leche ha sido adulterada por el productor.

Test de hipótesis: Ingredientes

- H_0: la hipótesis nula. Este es el supuesto por defecto para el modelo que genera los datos.
- H_A: la hipótesis alternativa. Si rechazamos la hipótesis nula, aceptamos esta alternativa como la mejor explicación para los datos.
- X: el estadístico de prueba. Calculamos esto a partir de los datos.
- Distribución nula: la distribución de probabilidad de X asumiendo H_0.
- Región de rechazo: si X está en la región de rechazo se rechaza H_0 a favor de H_A.
- Región de no rechazo: el complemento a la región de rechazo. Si X está en esta región no rechazamos H_0.

Planteo general

La hipótesis nula H_0 y la hipótesis alternativa H_A desempeñan diferentes roles.

- Por lo general, elegimos queH_0 sea una hipótesis simple o por defecto (e.g. las diferencias observadas se deben simplemente al azar), que solo rechazaremos si tenemos pruebas suficientes contra ella. Si θ es el parámetro de interés, la hipótesis nula es un valor θ_0.
- Posibles hipótesis alternativas:

$$H_a = \begin{cases} \theta \neq 0 & \text{Prueba bilateral} \\ \theta > \theta_0 & \text{Prueba unilateral a cola superior} \\ \theta < \theta_0 & \text{Prueba unilateral a cola inferior} \end{cases}$$

En el ejemplo de la leche realizamos un test a cola derecha.

Pvalor

Se llama P-valor a la probabilidad calculada asumiendo válida la hipótesis nula, de que el estadístico tome un valor al menos tan extremo como el observado.
Más chico es el P-valor, más grande es la evidencia en contra de H_0 dada por los datos.

P-valor-leche aguada

- Volvamos al ejemplo del gerente y sospecha acerca de la leche aguada.
- El P-valor en este caso es la probabilidad

$$
\begin{aligned}
P(\overline{X} \geq -0.538) &= P\left[\frac{\overline{X} - (-0.545)}{0.008/\sqrt{5}} \geq \frac{-0.538 - (-0.545)}{0.008/\sqrt{5}}\right] \\
&= P(Z \geq 1.96) \\
&= 1 - 0.975 = 0.025
\end{aligned}
$$

Significado estadístico

Si el P-valor es menor o igual que un valor fijo α, decimos que los datos son estadística mente significativos con nivel α.

Test de nivel α

1. Se proponen las hipótesis H_0 nula y alternativa H_a. El test se diseña para evaluar la fuerza de la evidencia en contra de H_0. H_a es la propuesta que vamos a aceptar si la evidencia nos permite rechazar H_0.
2. Se especifica el nivel de significación α, que permite decir por anticipado cuanta evidencia se necesita para rechazar H_0.
3. Se calcula el estadístico del test, cuya distribución debe ser conocida suponiendo válida la hipótesis nula, y que nos permite medir cuán bien los datos soportan H_0.
4. Se calcula el P-valor de los datos observados. Esta es la probabilidad calculada asumiendo valida H_0, de que el estadístico del test vaya a pesar tanto o más en contra de H_0 como lo hace con estos datos. Si el P-valor es menor o igual a α, el test es significativo con nivel α.

Errores

De acuerdo con el paradigma de Neyman-Pearson, podemos incurrir en dos clases de errores.

1. Uno de ellos es el llamado error de tipo I, la probabilidad de rechazar la hipótesis nula cuando era verdadera. Si observamos la definición de nivel de significación α, este es el error de tipo I.
2. El error de tipo II es la probabilidad de aceptar la hipótesis nula cuando es falsa, y se llama β. Para ser calculado hay que proponer una alternativa completamente identificable.

Potencia

El valor $(1-\beta)$ es la probabilidad de que H_0 sea rechazada cuando es falsa, y se llama potencia del test. Permite comparar dos tests diferentes para resolver el mismo problema, el que tiene potencia mayor es el mejor.

Test de nivel α

- Un test ideal debería tener los errores α y β iguales a cero.Pero eso no ocurre salvo en casos triviales.
- Es más, en la practica, ocurre casi siempre que para un tamaño de muestra fija, achicar el α implica agrandar el β, por lo cual si se quiere un nivel de significación mayor termina obteniéndose un test menos potente, un test con menos capacidad de rechazar cuando en realidad debería haber rechazado.

Errores

Al realizar una prueba de hipótesis se pueden cometer dos tipos de errores que se señalan en la siguiente tabla:

	Aceptar H_0	Rechazar H_0
H_0 es verdadera	Decisión correcta	Error de tipo I
H_0 es falsa	Error de tipo II	Decisión correcta

Pruebas de Hipótesis para la media poblacional

Hipótesis Nula:

$$H_0 : \mu = \mu_0$$

CASO A: Sea $X_1, \ldots, X_n$ una m.a. de distribución $N\left(\mu, \sigma^2\right)$ con σ^2 conocido. Bajo estos supuestos sabemos que

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

o equivalentemente

$$\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0, 1)$$

Tomamos como Estadístico de test:

$$\frac{\left(\bar{X} - \mu_0\right)\sqrt{n}}{\sigma} \sim N(0, 1), \quad \text{bajo } H_0$$

RR y potencia

H_a	RR_α	$\beta(\mu)$
$\mu < \mu_0$	$z \leq -z_\alpha$	$1 - \Phi\left(-z_\alpha + \frac{(\mu_0-\mu)\sqrt{n}}{\sigma}\right)$
$\mu > \mu_0$	$z \geq z_\alpha$	$\Phi\left(z_\alpha + \frac{(\mu_0-\mu)\sqrt{n}}{\sigma}\right)$
$\mu \neq \mu_0$	$\|z\| \geq z_{\alpha/2}$	$\Phi\left(z_{\alpha/2} + \frac{(\mu_0-\mu)\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} + \frac{(\mu_0-\mu)\sqrt{n}}{\sigma}\right)$

Tamaño de muestra

Para este caso, si la prueba de hipótesis es unilateral, el mínimo tamaño de muestra necesario para que una prueba de nivel α sea tal que $\beta(\mu') \leq \beta_0$ es

$$n \geq \left[\frac{(z_{\beta_0} + z_{\alpha})}{(\mu_0 - \mu')}\sigma\right]^2$$

Si la prueba de hipótesis es bilateral, el mínimo tamaño de muestra necesario para que una prueba de nivel α sea tal que $\beta(\mu') \leq \beta_0$ es

$$n \geq \left[\frac{\left(z_{\beta_0} + z_{\frac{\alpha}{2}}\right)}{(\mu_0 - \mu')}\sigma\right]^2$$

Pruebas de Hipótesis para la media poblacional

Hipótesis Nula:

$$H_0 : \mu = \mu_0$$

CASO B: Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria con valor medio μ y varianza σ^2 respectivamente.

- Si $n \geq 30$ y σ es conocida, entonces por TCL el Estadístico de Prueba:

$$Z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \dot{\sim} N(0, 1), \quad \text{bajo } H_0$$

Pruebas de Hipótesis para la media poblacional

H_a	RR_α	$\beta(\mu)$
$\mu < \mu_0$	$z \leq -z_\alpha$	$1 - \Phi\left(-z_\alpha + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right)$
$\mu > \mu_0$	$z \geq z_\alpha$	$\Phi\left(z_\alpha + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right)$
$\mu \neq \mu_0$	$\|z\| \geq z_{\alpha/2}$	$\Phi\left(z_{\alpha/2} + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right)$

Pruebas de Hipótesis para la media poblacional

- Si $\boldsymbol{n} \geq 40$ y σ es desconocida, se reemplaza σ por S_{n-1} y el Estadístico de Prueba:

$$Z = \frac{(\bar{X} - \mu)\sqrt{n}}{S_{n-1}} \dot{\sim} N(0,1), \quad \text{bajo } H_0$$

Pruebas de Hipótesis para la media poblacional

H_a	RR_α	$\beta(\mu)$
$\mu < \mu_0$	$z \leq -z_\alpha$	$1 - \Phi\left(-z_\alpha + \frac{(\mu_0-\mu)\sqrt{n}}{s_{n-1}}\right)$
$\mu > \mu_0$	$z \geq z_\alpha$	$\Phi\left(z_\alpha + \frac{(\mu_0-\mu)\sqrt{n}}{s_{n-1}}\right)$
$\mu \neq \mu_0$	$\lvert z \rvert \geq z_{\alpha/2}$	$\Phi\left(z_{\alpha/2} + \frac{(\mu_0-\mu)\sqrt{n}}{s_{n-1}}\right) - \Phi\left(-z_{\alpha/2} + \frac{(\mu_0-\mu)\sqrt{n}}{s_{n-1}}\right)$

Pruebas de Hipótesis para la media poblacional

- Para este caso, si la prueba de hipótesis es unilateral, el mínimo tamaño de muestra necesario para que una prueba de nivel α sea tal que $\beta\left(\mu^{t}\right) \leq \beta_0$ es

$$n \geq \left[\frac{(z_{\beta_0} + z_{\alpha})}{(\mu_0 - \mu')}\sigma\right]^2$$

- Si la prueba de hipótesis es bilateral, el mínimo tamaño de muestra necesario para que una prueba de nivel α sea tal que $\beta\left(\mu'\right) \leq \beta_0$ es

$$n \geq \left[\frac{\left(z_{\beta_0} + z_{\frac{\alpha}{2}}\right)}{(\mu_0 - \mu')}\sigma\right]^2$$

Ejemplo

Un fabricante de automóviles afirma que un 0km de determinado modelo tiene un rendimiento medio de, al menos, 12 km por litro de nafta. Un grupo de investigación de mercados, sospechando de tal afirmación, realiza un test. Se considera que la distancia recorrida por litro tiene distribución normal con desviación estándar $\sigma = 2$, y se toma una muestra aleatoria de 30 automóviles.

(a) Hallar la región de rechazo del test para el nivel $\alpha = 0.02$.

(b) Hallar la probabilidad de error de tipo II si el rendimiento real es de 11 km por litro.

(c) Tome una decisión si el valor observado promedio es 11.32.

Resolución

Tenemos una muestra de variables normales con media desconocida y desviación estándar $\sigma = 2$. Tenemos que hacer inferencia sobre la media de la distribución, y queremos desacreditar al fabricante, por lo cual la hipótesis nula va a ser la tesis del fabricante y la alternativa es que hace menos de 12km por litro,

$$H_0 : \mu = 12km \qquad H_A : \mu < 12km$$

.

Podemos aplicar el test Z con nivel $\alpha = 0.02$. La región de rechazo con ese nivel es el conjunto

$$\{\overline{x}_{obse} \mid \frac{(\overline{x}_{obse} - \mu_0)}{\sigma/\sqrt{n}} < -z_\alpha\}$$

, dado que

$$P(\frac{(\overline{X} - \mu_0)}{\sigma/\sqrt{n}} < -z_\alpha) = \alpha$$

por definición de valor crítico.

Resolución

Si $\alpha = 0.02$, entonces $-z_\alpha = -2.0537$ y la región de rechazo es

$$\{z \mid z < -2.0537\}$$

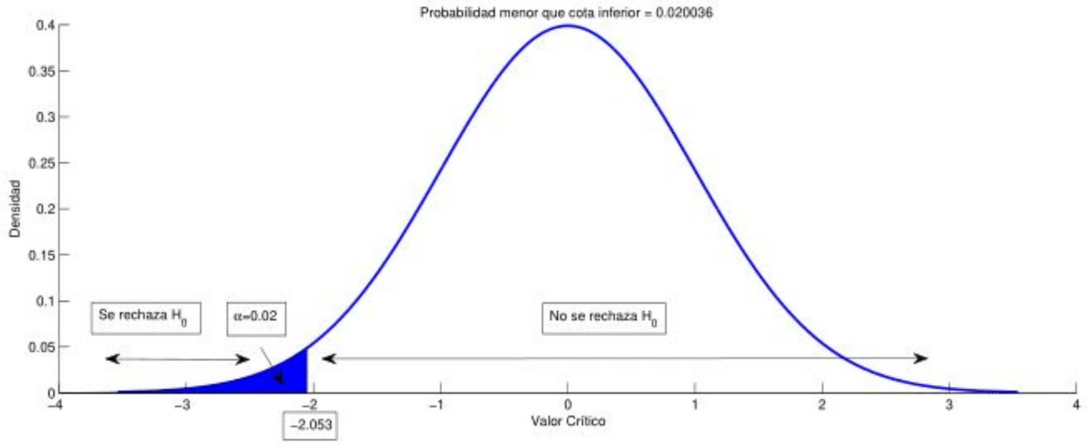


Figure: Región de rechazo en Z de un test a cola izquierda

Resolución

Podemos traducir la misma región a valores de $\overline{X}$,

$$\{\overline{x}_{obse} \mid \overline{x}_{obse} < -2.0537\frac{2}{\sqrt{30}} + 12 = 11.25\}$$

Como

$$\overline{x}_{obse} = 11.32$$

no puedo rechazar la hipótesis con ese nivel de error.

Resolución

El error de tipo II, β es la probabilidad de no rechazar cuando debía haber rechazado la hipótesis nula. Nos dan como dato que el verdadero valor de μ es 11 km.

$$
\begin{aligned}
\beta = P(\overline{X} > 11.25 \mid \mu = 11km) &= P(\frac{\overline{X} - 11}{2/\sqrt{30}} > \frac{11.25 - 11}{2/\sqrt{30}}) \\
&= P(Z > \frac{0.25\sqrt{30}}{2}) \\
&= P(Z > 0.684) = 0.2469
\end{aligned}
$$

La potencia de este test es $1 - \beta = 1 - 0.2469 = 0.75$ bastante alta, por lo cual tenemos confianza de que este test va a rechazar la hipótesis nula cuando realmente valga la hipótesis alternativa.

Ejemplo

Un contratista encarga un gran número de vigas de acero con longitud promedio de 5 metros. Se sabe que la longitud de una viga se halla normalmente distribuida con una desviación estándar de 0.02 metros. Usualmente, después de recibir un embarque, el contratista selecciona 16 vigas al azar y mide sus longitudes para decidir si acepta o rechaza el encargo.

(a) Si la probabilidad de rechazar un embarque bueno es 0.04, ¿cuál debe ser el rango de valores de la media muestral para que el embarque sea regresado al fabricante?

(b) Si la longitud promedio real es de 4.98 metros, ¿cuál es la potencia del test realizado?

Resolución

Primero debemos plantear el test de hipótesis. Si el contratista desea saber si el lote que recibió corresponde con el estándar de calidad que pidió, debe corroborar que la media de la población de barras es realmente $\mu = 5m$. La alternativa es a dos colas, barras con media menor o mayor no pueden ser aceptadas.

$$H_0 : \mu = 5m \qquad H_A : \mu \neq 5m$$

El error de quedar mal con el fabricante devolviendo algo que si alcanzaba el estándar es de $\alpha = 0.04$, bastante chico. El problema es que si la potencia es baja, puede no rechazarse el cargamento porque no hay evidencia suficiente en la muestra, no porque el cargamento alcance el estándar.

Resolución

La región de rechazo del test es el rango de valores de la media muestral para los cuales el embarque va a ser regresado al fabricante. Como la hipótesis es a dos colas, el valor crítico es $z_{0.04/2} = 2.053$

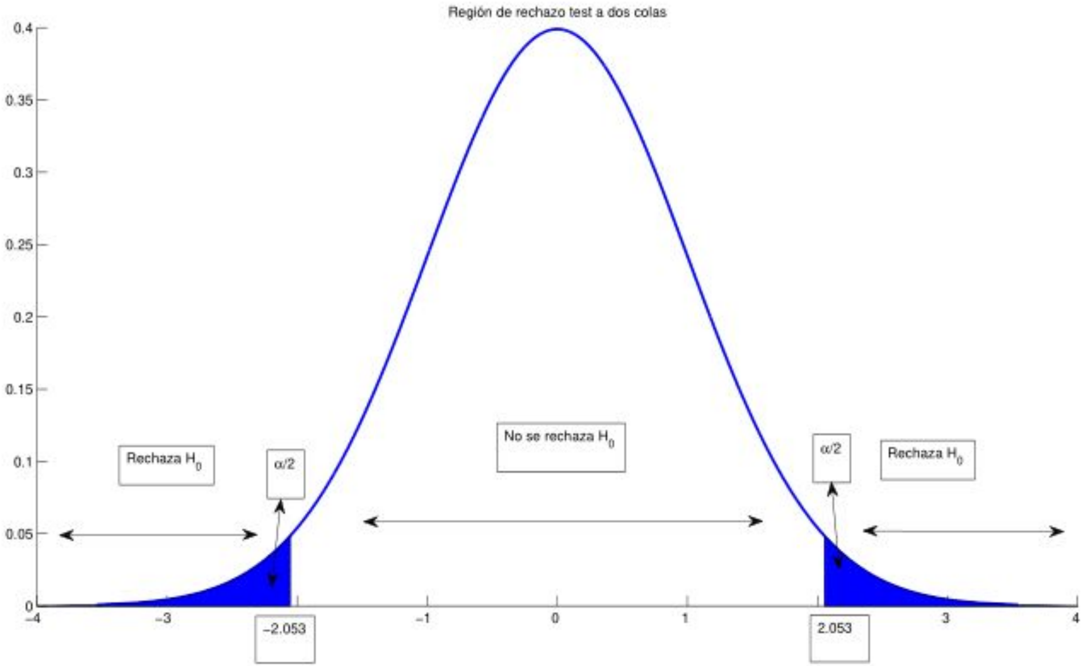
$$0.04 = P(|\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}| > z_{(1-0.04/2)})$$

y

$$\begin{aligned}
RR &= \{\overline{x}_{obse} \mid |\frac{\overline{x}_{obse} - \mu_0}{\sigma/\sqrt{n}}| > z_{0.04/2}\} \\
&= \{\overline{x}_{obse} \mid \frac{\overline{x}_{obse} - \mu_0}{\sigma/\sqrt{n}} > z_{0.04/2}\} \cup \{\overline{x}_{obse} \mid \frac{\overline{x}_{obse} - \mu_0}{\sigma/\sqrt{n}} < -z_{0.04/2}\} \\
&= \{\overline{x}_{obse} \mid \overline{x}_{obse} > z_{0.04/2}\sigma/\sqrt{n} + \mu_0\} \cup \{\overline{x}_{obse} \mid \overline{x}_{obse} < -z_{0.04/2}\sigma/\sqrt{n} + \mu_0\} \\
&= \{\overline{x}_{obse} \mid \overline{x}_{obse} > 5.01\} \cup \{\overline{x}_{obse} \mid \overline{x}_{obse} < 4.99\}
\end{aligned}$$

Resolución

Por lo cual el embarque va ser regresado al fabricante si el promedio de los largos de las vigas de la muestra extraída del lote es mayor a 5.01 o menor a 4.98. En la figura 2 vemos la región de rechazo sobre la normal patrón.



Resolución

Si la distribución de los largos de las vigas del lote es normal pero con media $\mu = 4.98m$ en vez de $5m$, el test realizado usando esta alternativa

$$H_0 : \mu = 5m \qquad H_a : \mu = 4.98$$

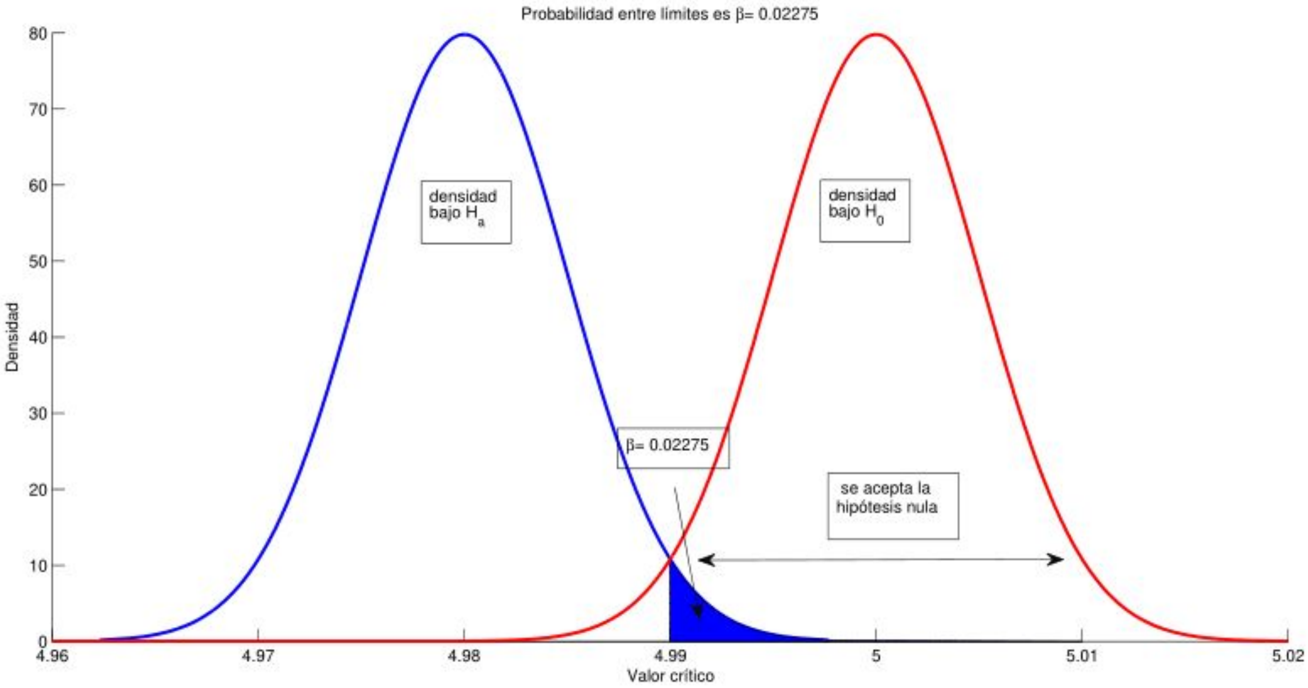
tiene potencia

$$\begin{aligned}
1-\beta &= 1 - P(4.99 \leq \overline{X} \leq 5.01 |\, \mu = 4.98) \\
&= 1 - P\left(\frac{4.99 - 4.98}{0.02/\sqrt{16}} \leq \frac{\overline{X} - 4.98}{0.02/\sqrt{16}} \leq 5.9\right) \\
\text{2?} \quad &= 1 - P(1.2 \leq Z \leq 5.9) \text{ ?} \qquad \text{6?} \\
&= 1 - 0.02275 = 0.97725
\end{aligned}$$

Resolución

La potencia del test es alta, por lo cual hay alta confianza en que el test rechace si la verdadera media del lote es 4.98. La figura muestra el error de tipo II, β para este test. En rojo está dibujada la densidad bajo hipótesis nula, y en azul la alternativa. El error es la intersección del área bajo la curva alternativa que entra dentro del área correspondiente a la región de aceptación de la hipótesis nula. Vemos que pocas muestras bajo la alternativa pueden ser confundidas como verdaderas bajo hipótesis nula, de allí el nombre de $(1 - \beta)$ como potencia del test.

Resolución



Probabilidad entre límites es β= 0.02275
densidad
bajo H_a
densidad
bajo H_0
β= 0.02275
se acepta la
hipótesis nula
Densidad
Valor crítico
80
70
60
50
40
30
20
10
0
4.96
4.97
4.98
4.99
5
5.01
5.02