UNIVERSIDAD DE MANAGUA

Al más alto nivel

Estadística Inferencial

Encuentro #11

Inferencia Estadística: Contraste de Hipótesis

Prof.: MSc. Julio Rito Vargas A. **IIIC-2017**

Inferencia Estadística: Contraste de Hipótesis:

Una hipótesis es una afirmación acerca de un hecho cualquiera, la cual se quiere contrastar o verificar con la realidad.

Una **hipótesis estadística** es una hipótesis, proposición o supuesto sobre la distribución de una variable aleatoria o sobre los parámetros de la distribución de una v.a.

Un contraste o test de hipótesis es una técnica de Inferencia Estadística que permite comprobar si la información que proporciona una muestra observada concuerda (o no) con la hipótesis estadística formulada sobre el modelo de probabilidad en estudio y, por tanto, se puede aceptar (o no) la hipótesis formulada.

Este es uno de los aspectos más útiles de la inferencia estadística, puesto que muchos tipos de problemas de toma de decisiones, pruebas o experimentos en el mundo de la ingeniería, pueden formularse como problemas de prueba de hipótesis.

Suponga que se tiene interés en la rapidez de combustión de un agente propulsor sólido utilizado en los sistemas de salida de emergencia para la tripulación de aeronaves. El interés se centra sobre la rapidez de combustión promedio. De manera específica, el interés recae en decir si la rapidez de combustión promedio es o no 50 cm/s. Esto puede expresarse de manera formal como

H_o; μ= 50 cm/s

H_1; $\mu\neq$50 cm/s

En una prueba de hipótesis surgen dos hipótesis excluyentes:

La proposición H_o; μ = 50 cm/s, se conoce como **hipótesis nula**, mientras que la proposición H_1; $\mu \neq$ 50 cm/s, recibe el nombre de **hipótesis alternativa.**

TIPOS DE HIPÓTESIS:

1) **Simples** cuando se refieren a un solo valor del parámetro poblacional**.**

$$H_o : \mu = \mu_o$$
$$H_1 : \mu \neq \mu_o$$

Prueba de 2 colas

2) **Compuestas** cuando se refieren a más de un valor del parámetro poblacional.

$$H_o : \mu \geq \mu_o$$
$$H_1 : \mu < \mu_o$$

Prueba de cola izquierda

$$H_o : \mu \leq \mu_o$$
$$H_1 : \mu > \mu_o$$

Prueba de cola derecha

Es importante recordar que las hipótesis siempre son proposiciones sobre la población o distribución bajo estudio, no proposiciones sobre la muestra. Por lo general, el valor del parámetro de la población especificado en la hipótesis nula se determina en una de tres maneras diferentes:

1. Puede ser resultado de la experiencia pasada o del conocimiento del proceso, entonces el objetivo de la prueba de hipótesis usualmente es determinar si ha cambiado el valor del parámetro.
2. Puede obtenerse a partir de alguna teoría o modelo que se relaciona con el proceso bajo estudio. En este caso, el objetivo de la prueba de hipótesis es verificar la teoría o modelo.
3. Cuando el valor del parámetro proviene de consideraciones externas, tales como las especificaciones de diseño o ingeniería, o de obligaciones contractuales. En esta situación, el objetivo usual de la prueba de hipótesis es probar el cumplimiento de las especificaciones.

Un procedimiento que conduce a una decisión sobre una hipótesis en particular recibe el nombre de **prueba de hipótesis**. Los procedimientos de prueba de hipótesis dependen del empleo de la información contenida en la muestra aleatoria de la población de interés. Si esta información es consistente con la hipótesis, se concluye que ésta es verdadera; sin embargo si esta información es inconsistente con la hipótesis, se concluye que esta es falsa. Debe hacerse hincapié en que la verdad o falsedad de una hipótesis en particular nunca puede conocerse con certidumbre, a menos que pueda examinarse a toda la población. Usualmente esto es imposible en muchas situaciones prácticas. Por tanto, es necesario desarrollar un procedimiento de prueba de hipótesis teniendo en cuenta la probabilidad de llegar a una conclusión equivocada.

La **hipótesis nula,** representada por H_o, es la afirmación sobre una o más características de poblaciones que al inicio se supone cierta (es decir, la "creencia a priori").

La **hipótesis alternativa,** representada por H_1, es la afirmación contradictoria a H_o, y ésta es la hipótesis del investigador.

La hipótesis nula se rechaza en favor de la hipótesis alternativa, sólo si la evidencia muestral sugiere que H_o es falsa. Si la muestra no contradice decididamente a H_o, se continúa creyendo en la validez de la hipótesis nula. Entonces, las dos conclusiones posibles de un análisis por prueba de hipótesis son ***rechazar H_o o no rechazar H_o.***

PASOS PARA ESTABLECER UNA PRUEBA DE HIPOTESIS INDEPENDIENTEMENTE DE LA DISTRIBUCION QUE SE ESTE TRATANDO

1. Precisar el problema y el nivel de significancia.
2. Definir las suposiciones que generan los datos (Normalidad de la distribución, si se conoce o no la varianza, el tamaño de la muestra, etc)
3. Formular las hipótesis concernientes al objetivo que se persigue

   $H_o : \mu = \mu_o$ $\quad$ $H_o : \mu \geq \mu_o$ $\quad$ $H_o : \mu \leq \mu_o$

   $H_1 : \mu \neq \mu_o$ $\quad$ $H_1 : \mu < \mu_o$ $\quad$ $H_1 : \mu > \mu_o$

4. Fijar los valores críticos. Estos son estadísticos de acuerdo a los supuestos, basados en el nivel de significación deseado y de la hipótesis planteada.

   Para el caso Normal que es el que nos concierne en este momento seria:

   a) Prueba de dos colas o bilateral

   $H_o : \mu = \mu_o$

   $H_1 : \mu \neq \mu_o$

   b) Prueba de cola derecha

   $H_o : \mu \geq \mu_o$

   $H_1 : \mu < \mu_o$

   c) Prueba de cola izquierda

   $H_o : \mu \leq \mu_o$

   $H_1 : \mu > \mu_o$

5. Establecer la regla de decisión: Sirve para determinar cual es la región de aceptación y de rechazo. A partir del punto determinado por los valores críticos, en dirección a la cola o las colas se inicia la región o regiones de rechazo.

   Y la regla de decisión expresa que se rechazará H_o si el valor del estadístico cae en la región de rechazo y que no se rechazará H_o si el valor del estadístico cae en la región de aceptación.

6. Determinar el estadístico de prueba, se realiza de acuerdo a los datos y las suposiciones que se establecieron.
7. Decisión y Conclusiones.

Error tipo I y II

- Las hipótesis nula y alternativa son aseveraciones sobre la población que compiten entre sí
- No siempre es posible que las conclusiones sean verdaderas o correctas

	H_0 verdadera	H_a verdadera
Aceptar H_0	Conclusión Correcta	Error tipo II
Rechazar H_0	Error tipo I	Conclusión Correcta

- No se puede eliminar la posibilidad de errores en la prueba de hipótesis, pero si es posible considerar su probabilidad
- Se define como:

  α=probabilidad de cometer un error tipo I

  β=probabilidad de cometer error tipo II

- La máxima probabilidad permisible se le llama nivel de significancia para la prueba. Los valores acostumbrados son de 0.05 y 0.01
- En la mayoría de las aplicaciones se controla la probabilidad de cometer error tipo I, luego existe la incertidumbre con respecto al error tipo II
- Si los datos muestrales son consistentes con H_0 se adopta en la práctica la conclusión de "no rechazar H_0", ya que de esta forma evitamos el riesgo de cometer error tipo II

  La conclusión de "aceptar H_0" se toma sólo cuando se haya determinado el error tipo II.

  Suponga que se va a implantar un nuevo método de producción si una prueba de hipótesis respalda la conclusión de que con ese método se reduce la media del costo de operación por hora.

  1) Enuncie las hipótesis nula y alterna si la media del costo para el método actual de producción es de $220 por hora
  2) ¿Cuál es el error de tipo I en este caso y sus consecuencias?
  3) ¿Cuál es el error tipo II en este caso y sus consecuencias?

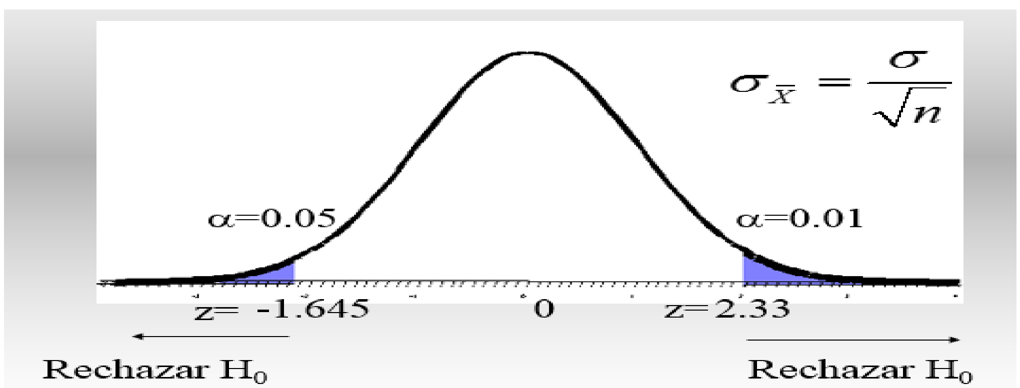
Pruebas unilaterales para la media

Muestra Grande

- En este caso (n>30) se asume distribución normal
- Para pruebas de hipótesis acerca de la media de una población se emplea el estadígrafo z

$$z = \frac{\left(\bar{X} - \mu\right)}{\sigma / \sqrt{n}}$$

- Se determina si la desviación del valor numérico en estudio es lo suficiente para justificar el rechazo de la hipótesis nula
- La probabilidades 0.05 y 0.01 de cometer error tipo I están relacionadas con un valor de z de –1.645 y –2.33 respectivamente
- Luego se debe rechazar H_0 si el valor de z es menor a –1.645 o –2.33 dependiendo del nivel de significancia
- El valor z establece el límite de la región de rechazo denominada valor crítico

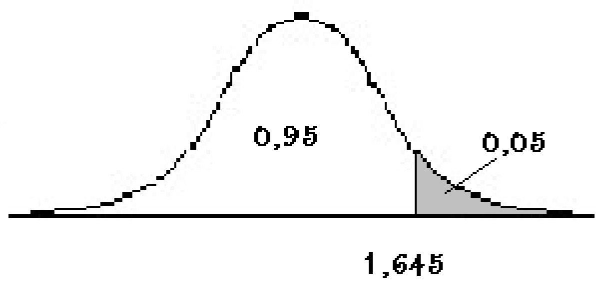


Resumen de pruebas unilaterales sobre media de una población. Si n≥30

$H_0 : \mu \geq \mu_0$	$H_0 : \mu \leq \mu_0$
$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$
$z = \frac{\left(\bar{X} - \mu_0\right)}{\sigma / \sqrt{n}}; \quad z = \frac{\left(\bar{X} - \mu_0\right)}{s / \sqrt{n}}$	$z = \frac{\left(\bar{X} - \mu_0\right)}{\sigma / \sqrt{n}}; \quad z = \frac{\left(\bar{X} - \mu_0\right)}{s / \sqrt{n}}$
Rechazar H_0 si $z < -z_\alpha$	Rechazar H_0 si $z > z_\alpha$

Ejemplo 1:

Una muestra aleatoria de 100 muertes registradas en Estados Unidos el año pasado muestra una vida promedio de 71.8 años. Suponga una desviación estándar poblacional de 8.9 años, ¿esto parece indicar que la vida media hoy en día es mayor que 70 años? Utilice un nivel de significancia de 0.05.

1) Significancia del 5%
2) Se trata de una distribución muestral de medias con desviación estándar conocida, y tamaño de muestra grande (N=100)
3) $H_o : \mu = \mu_o$ $\quad$ $H_1 : \mu > \mu_o$ $\quad$ Será que la vida media mayor a 70 años?
4) 

5) Si el estadístico de prueba es mayor a 1.645 Rechazar Hipótesis nula
6) Calculo del estadístico de prueba:

$$z = \frac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} = \frac{(71.8 - 70)}{8.9/\sqrt{100}} = \frac{1.8}{0.89} = 2.022$$

7) Como el Z es mayor a 1.645 se RECHAZA la hipótesis nula es decir que la vida media si es mayor a 70 años.

Ejemplo 2:

Un agente de bienes raíces afirma que 60% de todas las viviendas privadas que se construyen actualmente son casas con tres dormitorios. Para probar esta afirmación se inspecciona una muestra grande de viviendas nuevas. Se registra la proporción de las casas con 3 dormitorios y se utiliza como estadístico de prueba. Plantee las hipótesis nula y alternativa que se utilizaran en esta prueba y determine la ubicación de la región crítica.
Solución:

Si el estadístico de prueba fuera considerablemente mayor o menor que p = 0.6, rechazaríamos la afirmación del agente. En consecuencia, deberíamos plantear las siguientes hipótesis:
H_0: $p = 0.6$,
H_1: $p \neq 0.6$.

La hipótesis alternativa implica una prueba de dos colas con la región critica dividida por igual en ambas colas de la distribución de P^, nuestro estadístico de prueba.

Muestra Pequeña

- En este caso (n < 30) se asume que la población tiene una distribución normal
- Con distribución t se pueden hacer inferencias acerca de la media de la población

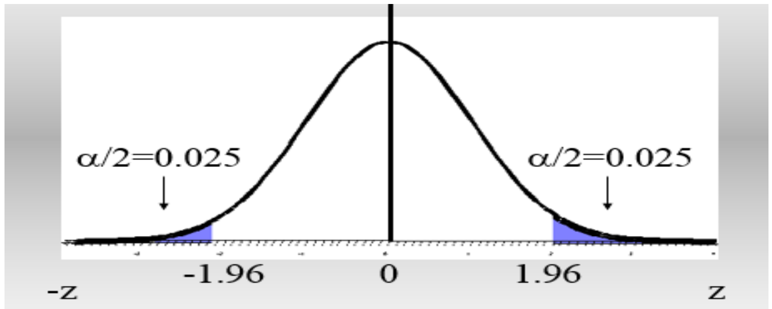
$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

- Para este estadígrafo se debe considerar los grados de libertad asociados al tamaño de la muestra (n-1) para definir el valor crítico que llevará al rechazo de H_0. Por las características de la tabla resulta complicado calcular el valor de p por lo que se expresa en intervalos

Pruebas bilaterales para la media

Muestra grande

- La diferencia de esta prueba con respecto a las unilaterales está en que la región de rechazo está ubicada simultáneamente en ambas colas
- En las pruebas bilaterales de hipótesis siempre se determina la región de rechazo colocando un área de probabilidad igual a $\alpha/2$ en cada cola de distribución
- Para este caso el valor de z para un nivel de significancia de 0.05 corresponderá a ± 1.96



Resumen de pruebas bilaterales sobre media de una población. Si n≥30

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

$$z = \frac{\left(\bar{X} - \mu_0\right)}{\sigma / \sqrt{n}}; \qquad z = \frac{\left(\bar{X} - \mu_0\right)}{s / \sqrt{n}}$$

$$\text{Rechazar H}_0 \text{ si } z < -z_{\alpha/2} \quad z > z_{\alpha/2}$$

Ejemplo 2:

Un diseñador quiere reducir el tiempo de secado de una pintura. Se prueban dos fórmulas de pintura. La fórmula 1 es la normal y la fórmula 2 posee un ingrediente secante que se espera reduzca el tiempo de secado. Se sabe que el tiempo de secado tiene una desviación estándar de 8 min y que ésta no se afecta con la adición del nuevo ingrediente. Se pintan 10 especímenes con la fórmula 1, y 10 con la fórmula 2, obteniéndose tiempos promedio de secado de x_1=121 minutos y x_2=112 minutos respectivamente. ¿A qué conclusión se llega sobre la eficacia del nuevo ingrediente?

Solución:

$\overline{X_1}$=121 ; $\overline{X_2}$=112; $n_1 = 10$; $n_2 = 10$; $\sigma_1 = 8\, min.$; $\sigma_2 = 8\, min.$

1) Cantidad de interés: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$
2) H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$
3) H_1: $\boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$ (se busca evidencia fuerte que indique que el tiempo de secado promedio de la muestra 2 es menor)
4) α =0.05
5) El estadístico de prueba es $Z = (\bar{X}_1 - \bar{X}_2) / \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
6) H_0 se rechazará si $z>z_{0.05}$ = 1.645
7) **Sustituyendo los datos, obtenemos $\mathbf{z=(121-112)/(12.8)^{1/2}=2.52}$**

Conclusión: Puesto que z = 2.52 > 1.645 se rechaza H_0 con un nivel de significancia a=0.05 concluyéndose el nuevo ingrediente sí disminuye el tiempo de secado

I. A las personas que sufren de tensión alta, se les recomienda seguir una dieta libre de sal. Queremos realizar un estudio para comprobar si esta dieta es efectivamente ventajosa. Para el estudio se estudió una muestra de 8 personas y se tomó la tensión antes de empezar la dieta y dos semanas después. Los resultados obtenidos fueron:

Antes	93	106	87	92	102	95	88	110
Después	92	102	89	92	101	96	88	105

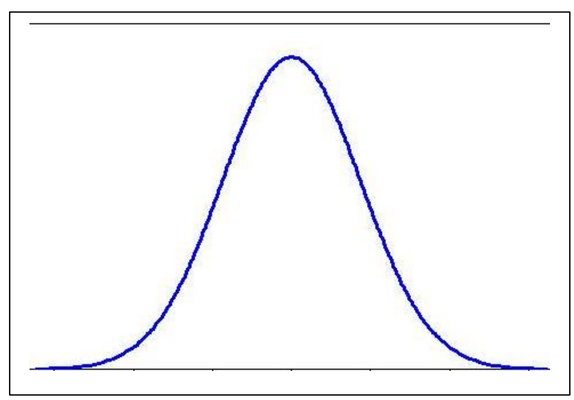
Denotamos μ_A y μ_B a las medias poblacionales de tensión antes y después de empezar la dieta, respectivamente. De este modo, el contraste de hipótesis que debemos plantear es:

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A < \mu_B$$

Haga el contraste para un nivel de confianza del 96%. Responda si hay cambio significativo después de empezar la dieta?

Recomendación:
Tiene que usar la siguiente fórmula como estadístico de prueba.

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



$$donde: \quad S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Debe calcular: $\bar{X}_1$ *y* $\bar{X}_2$
S_1^2 *y* S_2^2
Sustituir en S_p *y luego calcular* t_o

Hacer el contraste de hipótesis buscando en la tabla t-student
$t_{0.02,7}$ =________ y $t_{0.98,7}$ =__________

II. *Una compañía de transporte de carga desea escoger la mejor ruta para llevar la mercancía de un depósito a otro. La mayor preocupación es el tiempo de viaje. En el estudio se seleccionaron al azar 5 choferes de un grupo de 10 y se asignaron* *a la ruta A; los cinco restantes se asignaron a la ruta B. Los datos obtenidos fueron:*

Ruta	*Tiempo del viaje en horas*				
A	*18*	*24*	*30*	*21*	*32*
B	*22*	*29*	*34*	*25*	*35*

a) Existen diferencias significativas entre las rutas?
b) Plantee la prueba de hipótesis estadística correspondiente?

Recomendación: siga los pasos orientados en el problema anterior.

III. Un sociólogo ha pronosticado, que en una determinada ciudad, el nivel de abstención en las próximas elecciones será del 40% como mínimo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar. Determinar con un nivel de significación del 95%, si se puede admitir el pronóstico.

a) Formula la hipótesis
b) Haga el contraste y determine si se admite al pronóstico.

Recomendación:

Formula esta hipótesis:

Ho: p= 40%

H_A: p≠ 40%

Use el siguiente estadístico de prueba: $z_o = \frac{\hat{p} - np}{\sqrt{np(1-p)}}$

Busque en la tabla normal estándar el valor de z para 0.025 o 0975

Para las dos colas.

IV. El control de calidad una fábrica de pilas y baterías sospecha que hubo defectos en la producción de un modelo de batería para teléfonos móviles, bajando su tiempo de duración. Hasta ahora el tiempo de duración en conversación seguía una distribución normal con media de 300 minutos y desviación típica de 30 minutos. Sin embargo, en la inspección del último lote producido, antes de enviarlo al mercado, se obtuvo que de una muestra de 60 baterías el tiempo medio de duración en conversación fue de 290 minutos. Suponiendo que ese tiempo sigue siendo Normal con la misma desviación típica:

¿Se puede concluir que las sospechas del control de calidad son ciertas a un nivel de significación del 98%?

Recomendación: formule la siguiente hipótesis.

H_o: $\boldsymbol{\mu}$=300 minutos

H_A: $\boldsymbol{\mu}$≠300 minutos.

Use el siguiente estadístico de prueba

$$z_0 = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Busque en la tabla normal el valor z. y compara como en el ejemplo III.

V. Un experimento se comparó el ahorro de combustible para dos tipos de camiones: compactos que funcionan con diésel y están equipados de forma similar. Suponga que se utilizaron 12 camiones Volkswagen y 10 Toyota en pruebas con una velocidad constante de 90 kilómetros por hora. Si los 12 camiones Volkswagen promedian 16 kilómetros por litro con una desviación estándar de 1.0 kilómetros por litro, y los 10 Toyota promedian 11 kilómetros por litro con una desviación estándar de 0.8 kilómetros por litro, construya un intervalo de confianza del 90% para la diferencia de medias entre los kilómetros promedio por litro de estos dos camiones compactos. Suponga que las distancias por litro para cada modelo de camión están distribuidas de forma aproximadamente normal y que tienen varianzas iguales.

VI. Una empresa de taxis trata de decidir si comprará neumáticos de la marca *A* o de la marca *B* para su flotilla de taxis. Para estimar la diferencia entre las dos marcas realiza un experimento utilizando 12 neumáticos de cada marca, los cuales utiliza hasta que se desgastan. Los resultados son:

$$\begin{aligned} \text{Marca } A: \bar{x}_1 &= 36{,}300 \text{ kilómetros}, \\ s_1 &= 5000 \text{ kilómetros}. \\ \text{Marca } B: \bar{x}_2 &= 38{,}100 \text{ kilómetros}, \\ s_2 &= 6100 \text{ kilómetros}. \end{aligned}$$

Calcule un intervalo de confianza del 95% para $\mu_A - \mu_B$, suponiendo que las poblaciones se distribuyen de forma aproximadamente normal. Puede no suponer que las varianzas son iguales.

Ejemplo de contraste de hipótesis para la media, μ con σ conocida (dos colas)

Se desea contrastar con un nivel de significación del 5 % la hipótesis de que la talla media de los hombres de 18 o más años de un país es igual a 180. Suponiendo que la desviación típica de las tallas en la población vale 4, contraste dicha hipótesis frente a la alternativa de que es distinta.

$$H_0 : \mu = 180$$

frente a la alternativa:

$$H_1 : \mu \neq 180$$

Los datos constituyen una muestra de n=15 hombres seleccionados al azar, cuyas alturas son:

167 167 168 168 168 169 171 172 173 175 175 175 177 182 195

Es necesario determinar la media de la muestra, $\overline{X}$, y los valores de los cuantiles, $z_{\frac{\alpha}{2}}$, en la distribución normal. En el modelo normal, el cuantil de orden 0.975 es $z_{0,025} = 1{,}96$.

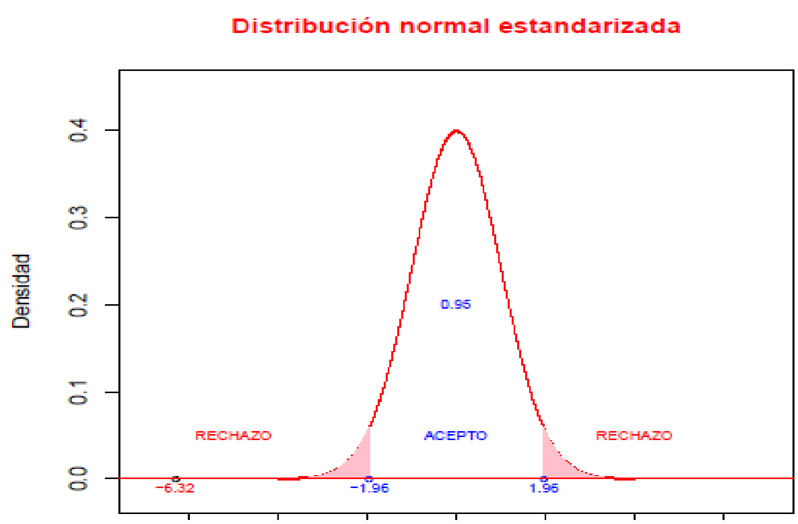
La media de la muestra es igual a 173.47.

Sustituyendo los datos en la expresión del estadístico de contraste, tenemos:

$$z_c = \frac{173{,}47 - 180}{\frac{4}{\sqrt{15}}} = -6{,}32$$

El valor del estadístico de contraste está en la zona de rechazo. Por lo que se rechaza la hipótesis nula que establece una talla media igual a 180 cm.

Gráficamente la situación es la siguiente:



Ejemplo de Contraste de hipótesis para la media con σ conocida(una cola)

Se desea contrastar con un nivel de significación del 5 % la hipótesis de que la talla media de los hombres de 18 o más años de un país es igual o mayor a 175. Suponiendo que la desviación típica de las tallas en la población vale 4, contraste dicha hipótesis frente a la alternativa de que es menor, con una muestra de n=15 hombres seleccionados al azar, cuyas alturas son las del apartado anterior:

$$H_0 : \mu \geq 175$$

frente a la alternativa:

$$H_1 : \mu < 175$$

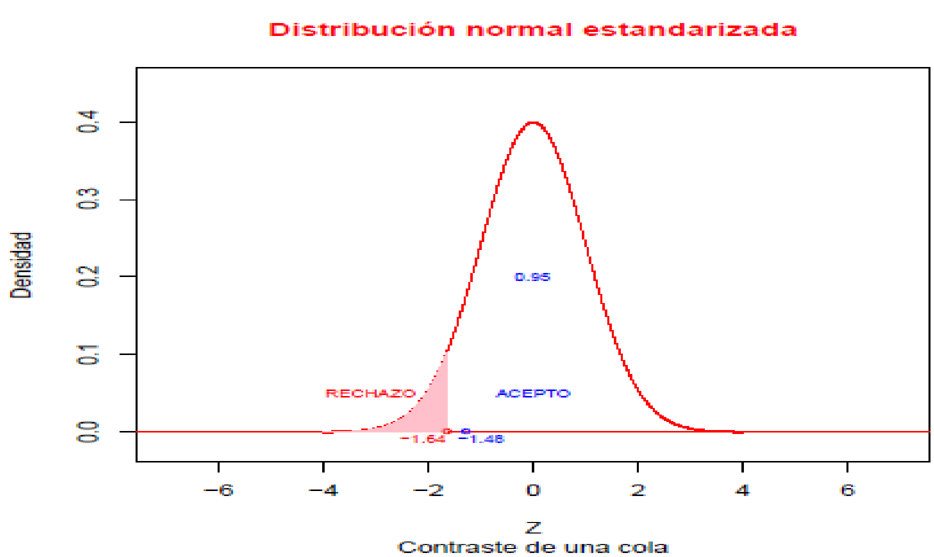
En el modelo normal, el cuantil de orden 0.05 es $-z_{0,05} = -1{,}64$.

Sustituyendo los datos en la expresión del estadístico de contraste, tenemos:

$$z_c = \frac{173{,}47 - 175}{\frac{4}{\sqrt{15}}} = -1{,}48$$

El valor del estadístico de contraste está en la zona de aceptación. Por lo que no se puede rechazar la hipótesis nula que establece una talla media igual o mayor a 175 cm.

Gráficamente la situación es la siguiente:



Ejemplo de Contraste de Hipótesis para la media μ con σ desconocida (una cola)

Supongamos que se desconoce la desviación típica de las tallas en la población del ejemplo anterior. Se desea contrastar la hipótesis nula siguiente a un nivel de significación del 5 %.

$$H_0 : \mu \leq 168$$

frente a la alternativa:

$$H_1 : \mu > 168$$

En este caso es necesario estimar la desviación típica de la población con los datos de la muestra.

Dado el valor muestral de la media $\overline{X}$ y la cuasidesviación típica de la muestra,s, se determina el estadístico de contraste:

$$t_c = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

En este ejemplo el contraste es de una cola, tal que el nivel de significación α que determina el valor t_α es tal que

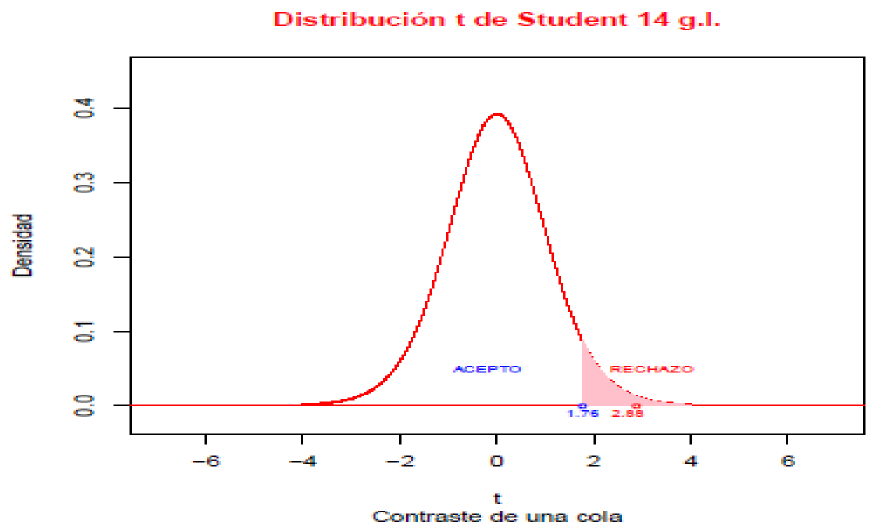
$$\alpha = P(t_{n-1} \geq t_\alpha)$$

Para el nivel de significación dado, $\alpha = 0{,}05$, es necesario determinar el cuantil en la distribución t de Student con n-1 grados de libertad y el valor s (cuasidesviación típica) de la muestra:

Para los datos de la muestra se obtiene que $\overline{X} = 137{,}47$ y $s = 7{,}36$. El cuantil en la distribución t_{14} es $t_\alpha = 1{,}762$. De modo que sustituyendo en la expresión del estadístico de contraste, tenemos:

$$t_c = \frac{173{,}47 - 168}{\frac{7{,}36}{\sqrt{15}}} = 2{,}88$$

El gráfico siguiente muestra la situación que nos lleva a rechazar la hipótesis nula, dado que el valor del estadístico de contraste cae en la zona de rechazo.


Distribución t de Student 14 g.l.
Densidad
0.0
0.1
0.2
0.3
0.4
ACEPTO
RECHAZO
1.76
2.88
−6
−4
−2
0
2
4
6
t
Contraste de una cola