Contraste de hipótesis
Sumario

- Hipótesis nula y alternativa.
- Prueba de hipótesis de una y dos colas.
- Nivel de significación.
- Errores de Tipo I (α) y Tipo II (β)
- Región Crítica.
- Pasos para Formular un Problema de Pruebas de Hipótesis.

- ¿Cómo es posible tomar decisiones acerca de un nuevo método de tratamiento si los resultados no son comparados con otros métodos?
- ¿Cómo es posible corroborar si una prueba diagnostica es más eficiente que otra?
- ¿Cómo es posible establecer si una enfermedad tiene similar comportamiento en dos lugares diferentes?
- ¿Cómo es posible arribar a la conclusión de que los pacientes tratados tienen similar evolución a uno descrito por otros?

Prueba de Hipótesis

- Procedimiento objetivo que permite, sobre la base de una información muestral, tomar una decisión.
- Tiene como objetivo principal, el establecimiento de conclusiones validas para una población estudiando una muestra aleatoria de esta y midiendo el grado de incertidumbre en términos de probabilidad.

Hipótesis

- Proposición cuya verdad o validez no se cuestiona en un primer momento, pero que permite iniciar una cadena de razonamientos que luego puede ser adecuadamente verificada.
- Toma la forma de un enunciado condicional, que debe seguir determinadas reglas para su admisión como razonamiento válido.

Hipótesis

- Afirmacion o suposicion, relacionada con el valor que puede tener un parametro poblacional, del cual depende el modelo teorico de una variable aleatoria.
- Ej.

  La diferencia entre la media de la poblacion de la cual fue tomada la muestra, es cero.

  Ho: $\mu = \mu_o$

  La diferencia entre la media de dos muestras de poblaciones con la misma media, es cero.

  Ho: $\mu_1 = \mu_2$

La hipótesis estadística

- Siempre estará integrada por la hipótesis nula (H_0) y la hipótesis alternativa (H_1), ambas forman una unidad.
- En la hipótesis nula se afirma o considera la no ocurrencia del resultado esperado, el término nula obedece a que es una hipótesis que contiene diferencias nulas.
- La hipótesis alternativa contradice, en algún sentido, la hipótesis nula, se expresa lo contrario a lo que se desea probar.
- De acuerdo a esto, rechazar H_0 equivale a probar el resultado esperado.

La hipótesis estadística

- **Hipotesis Nula**.

  Es la afirmacion que es probada, puede ser rechazada o no y es lo opuesto de lo que se desea probar.

  Se nombra $\mathbf{H_0}$, plantea que no existen diferencias.

- **Hipotesis Alternativa**.

  Afirmacion de lo que realmente se desea probar.

  Se nombra $\mathbf{H_1}$, plantea que existen diferencias

Hipótesis estadística

Nula (H_0) → **Se plantea la igualdad**

Alternativa (H_1) → **Se plantea la diferencia**

Ejemplo

- Se conoce que en una determinada población los niveles de colesterol en sangre se distribuyen normal con media 5,52 mmol/L y desviación estándar 0,16 mmol/L.
- Un grupo de investigadores de un Instituto especializado en Enfermedades Cardiovasculares realizó un estudio en una muestra aleatoria de 150 vegetarianos.

Ejemplo

- A partir de la muestra tomada se determinó el nivel promedio muestral del colesterol en dicho conjunto de personas.
- La hipótesis científica a ser probada estadísticamente es si individuos cuya ingesta está compuesta sólo por vegetales, tienen los mismos niveles de colesterol que la población general.

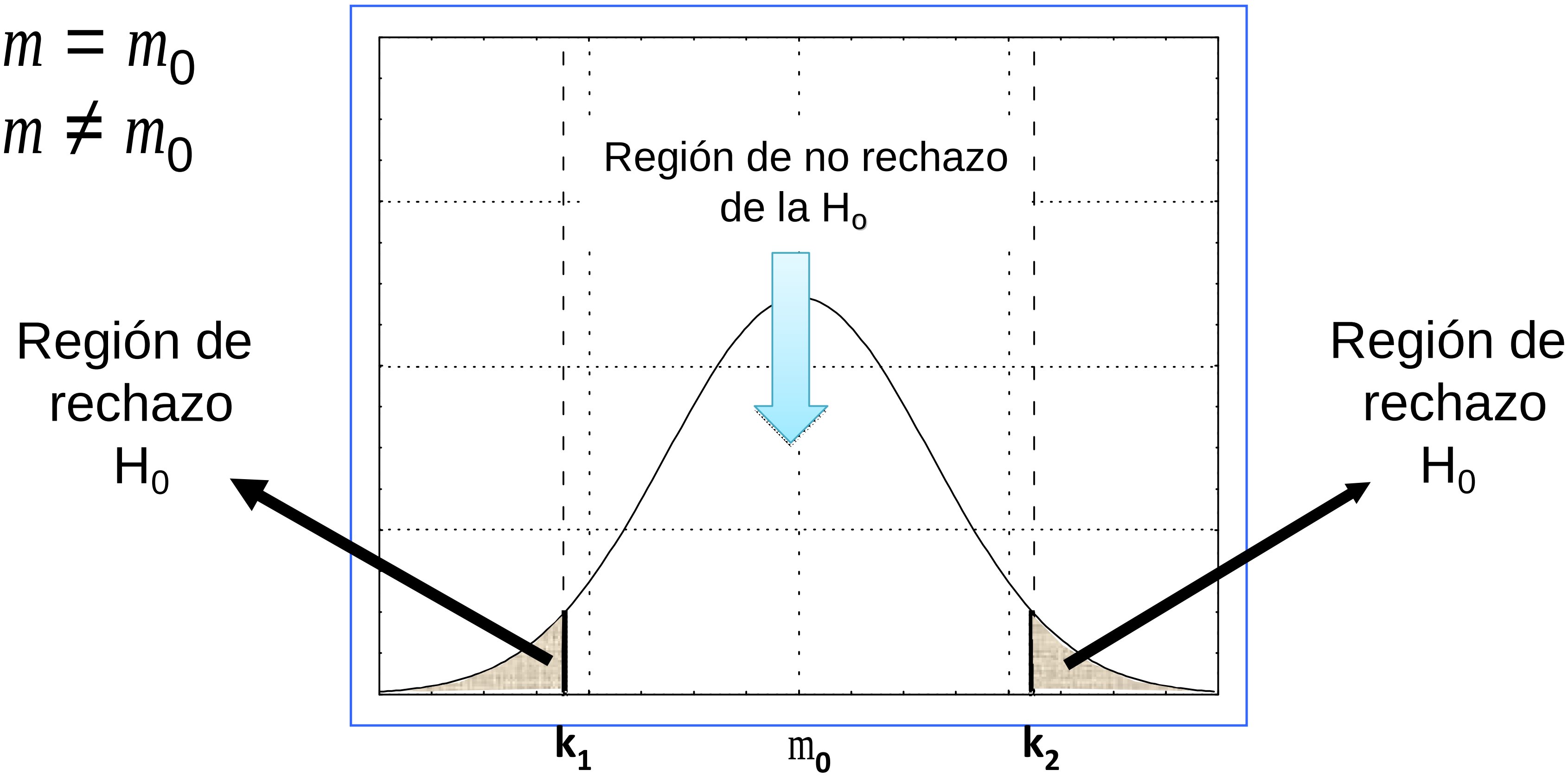
Ejemplo

- La hipótesis nula sería: H_0: $\mu = \mu_0 = 5{,}52$, donde $\mu_0 = 5{,}52$ representa lo general conocido.
- El sentido de nula, viene dado, por que no se espera encontrar ninguna variación en el colesterol medio de los vegetarianos, salvo la debida al azar.
- La hipótesis alternativa sería: H_1: $\mu \neq \mu_0 = 5{,}52$. En este caso, los investigadores, sólo están interesados en probar que existen diferencias entre el promedio de colesterol de vegetarianos y el de la población total.

Ejemplo

- Ellos estarán de acuerdo, en rechazar H_0, tanto si el nivel medio de colesterol en vegetarianos es menor que un valor k_1, como si es mayor que un valor k_2.

H_0: $m = m_0$
H_1: $m \neq m_0$



Ejemplo

- Ahora, se impone, encontrar alguna forma de poder arribar a una conclusión con respecto a, rechazar o no H_0, incluyendo, al igual que en el caso de los intervalos de confianza, la información contenida en una única muestra.
- En general, si el valor del estimador, digamos la media o la proporción muestral o cualquier otro, es tal que hay que rechazar la hipótesis nula, se dice que se ha obtenido un resultado significativo con un **nivel de significación** dado.

Nivel de significación

- Es posible que a causa de, la aleatoriedad de las observaciones muestrales, la estimación obtenida se desvíe tanto de lo esperado, que se tome, la decisión de rechazar H_0, siendo sin embargo H_0 cierta o verdadera.
- Es lógico o conveniente por tanto, que la probabilidad de que esto suceda, sea pequeña.
- Dentro de esta metodología, esta probabilidad, recibe el nombre de **nivel de significación**.

Nivel de significación

- El nivel de significación es un valor arbitrario seleccionado a priori por el investigador de acuerdo a su experiencia y deseo.
- Los valores que con más frecuencia se utilizan son 0,05 y 0,01. Se acostumbra a denotar este valor por la letra griega α.
- El uso del término, significación, es debido a que la diferencia entre, el valor hipotético o teórico y el hallado en la muestra, se considera lo, suficientemente grande, como para que no sea solamente atribuible al azar; es decir, que el concepto se refiere al estado de ser, estadísticamente significativo.

Errores de Tipo I (α) y Tipo II (β)

- Al tomar la decisión con respecto a la hipótesis nula podemos incurrir en dos errores diferentes.

- Cuando, $\mathbf{H_0}$ es **verdadera** y **se rechaza**, se produce el **error de tipo I**, con probabilidad de ocurrencia α, es decir, el nivel de significación de la prueba de hipótesis.

Errores de Tipo I (α) y Tipo II (β)

- Al tomar la decisión con respecto a la hipótesis nula podemos incurrir en dos errores diferentes.
- **Si no se rechaza** la hipótesis nula **H_0 y es falsa** se incurre entonces en el llamado **error de tipo II**.
- La probabilidad de cometer este tipo de error se conoce como β.
- Por su complejidad, no es objetivo de nuestro Curso ocuparnos de este error.

Errores de Tipo I (α) y Tipo II (β)

Si H_0 es	Desición sobre H_0	
	Rechazar	No rechazar
Verdadera	**Error de tipo I**	Acción correcta
Falsa	Acción correcta	**Error de tipo II**

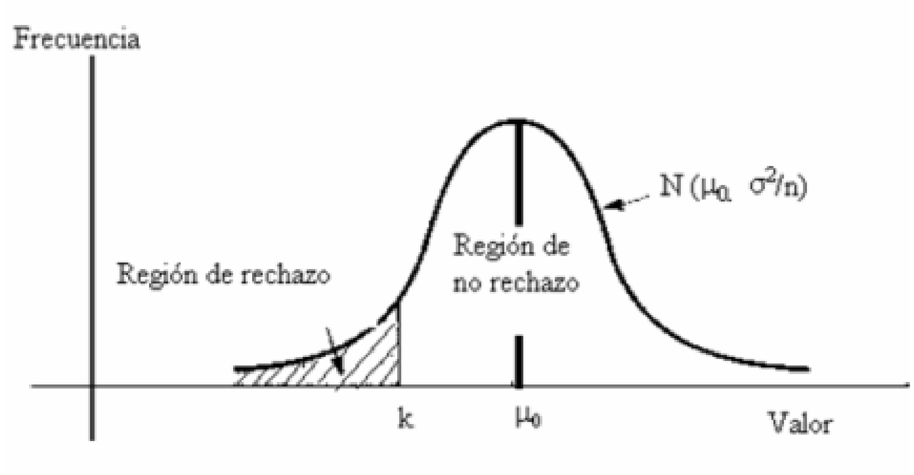
Estadígrafos de la prueba

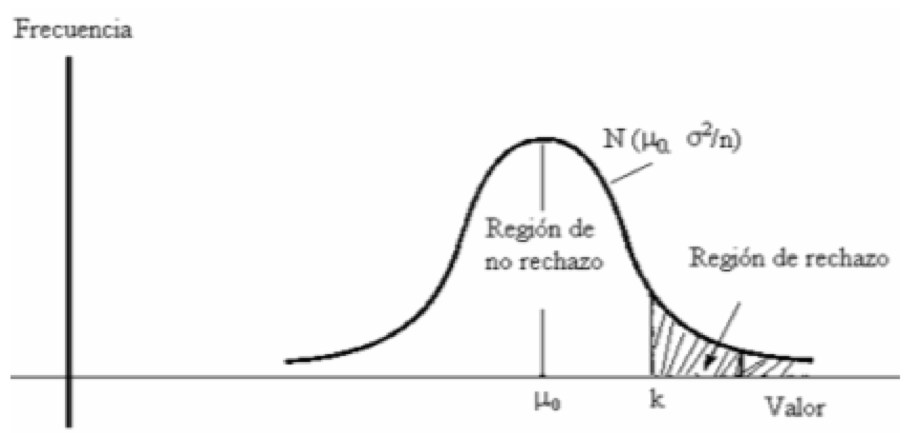
Situación	Distribución	Estadígrafo
Para μ σ conocida	Normal estándar	$Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$
σ desconocida $n > 30$	Normal estándar	$Z=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$
σ desconocida $n \leq 30$	*t*-Student	$t=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$
Proporciones para muestras grandes	Normal estándar	$Z=\dfrac{p-P}{\sqrt{p.(1-p)/n}}$

Segun la forma que H_0 es planteada, las pruebas de hipótesis se clasifica en:

- Prueba de Hipotesis de <u>una cola</u> o <u>unilatera</u>l

H_0: $m \leq m_0$,	H_0: $m \geq m_0$,
H_1: $m > m_0$	H_1: $m < m_0$



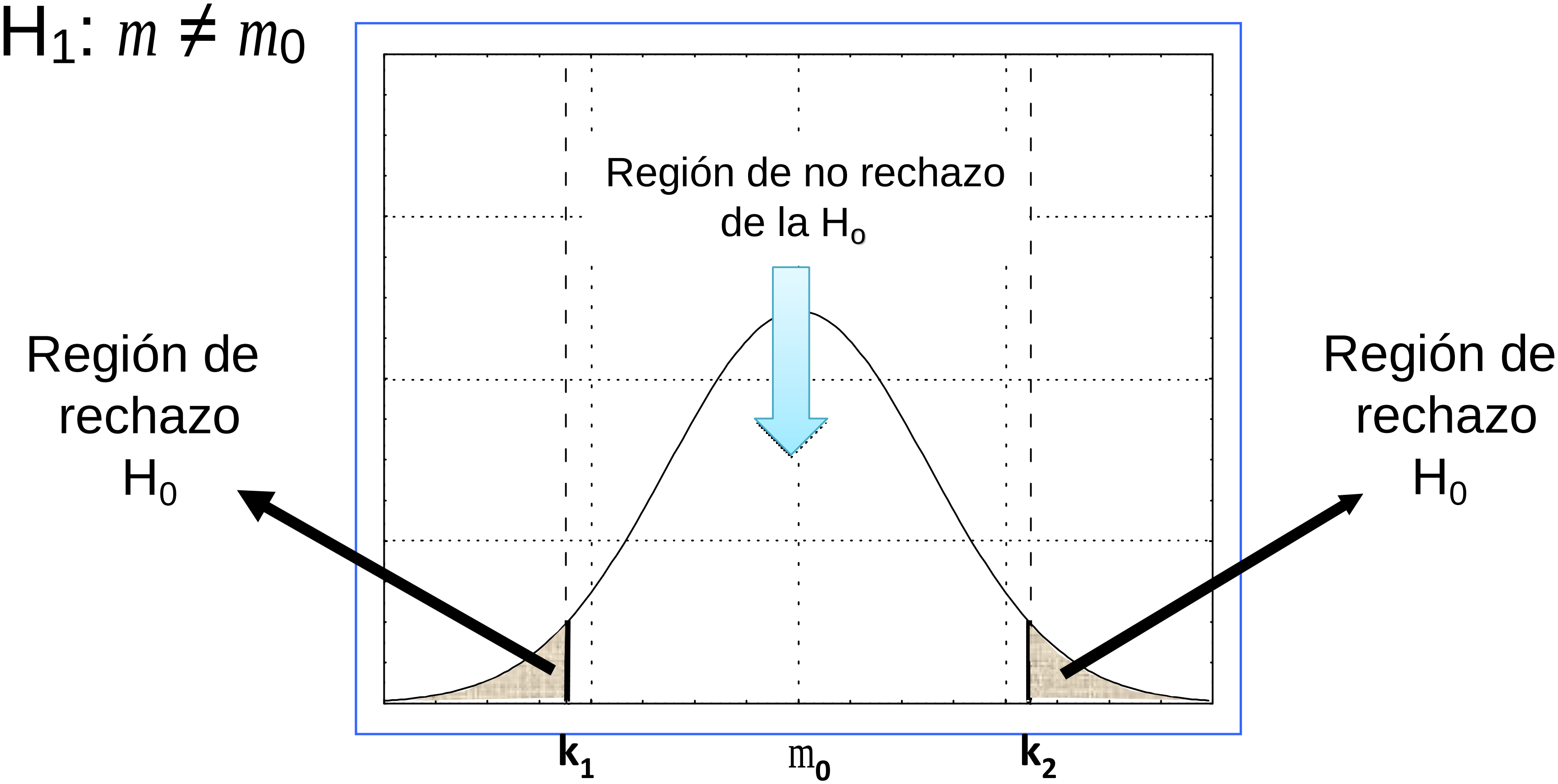


Segun la forma que H_0 es planteada, las pruebas de hipótesis se clasifica en:

- Prueba de Hipotesis <u>dos colas</u> o <u>bilateral</u>

  H_0: $m = m_0$,

  H_1: $m \neq m_0$



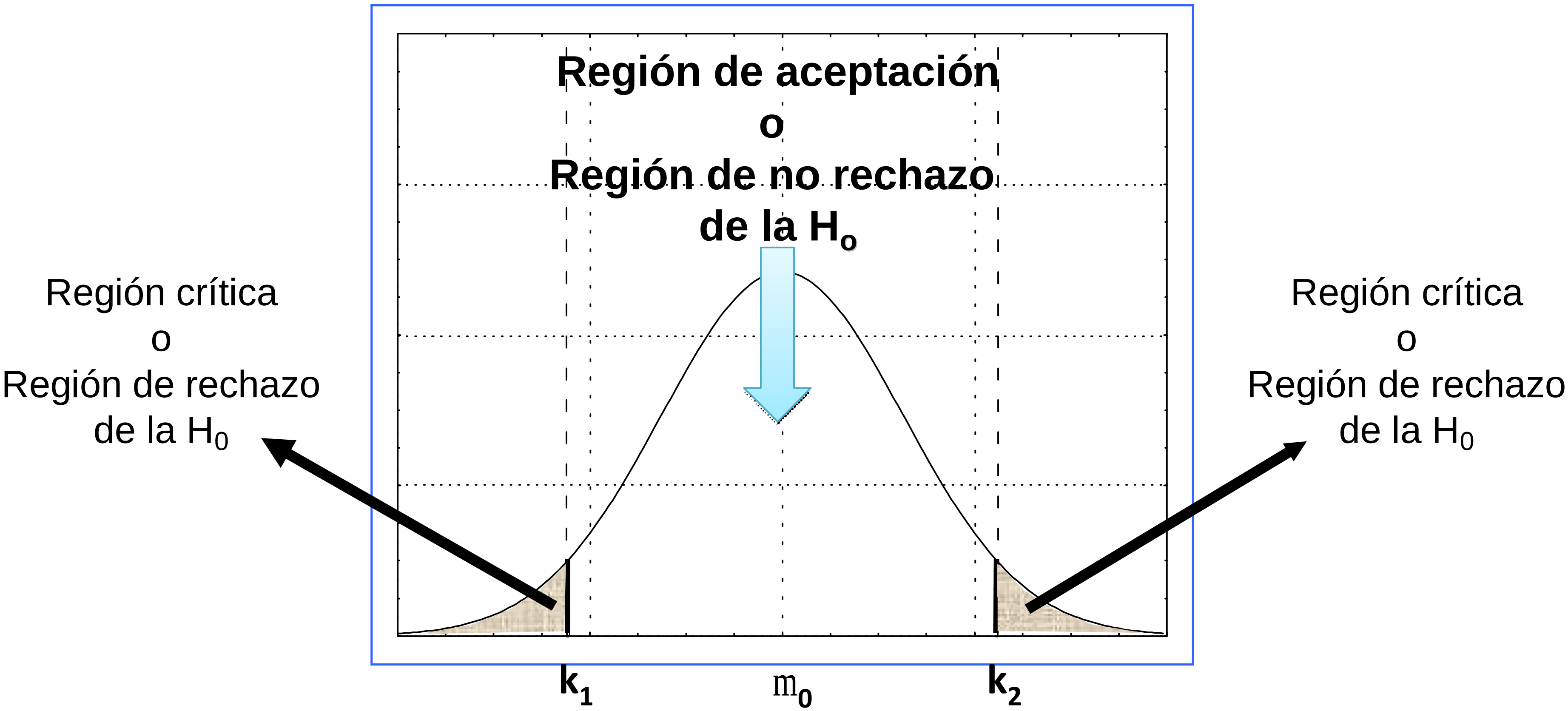
Región crítica

- La región crítica es aquella que se encuentra fuera de la **región de aceptación o de no rechazo de la H_0**
- En la **región de aceptación** se encuentran los valores que determinan la región de no rechazo entre los que esta concentrada el 95% (α=0.05) o el 99% (α=0.01) de la información. Es un intervalo de confianza para el 95 0 el 99% de confianza.

Región crítica

- Los datos que se ubican fuera de ese intervalo de confianza, en las llamadas colas de la distribución (Región crítica o de rechazo) ocurren menos frecuentemente que el resto.
- Se tomará la decisión de **rechazar H_0**, con un nivel de significación α, si el valor estimado del parámetro está en la **región crítica**
- y de **no rechazar H_0** si este valor **no está en la región crítica**.

Región crítica



Pasos a seguir

1. Formulación de las hipótesis estadísticas.

2. Escoger el nivel de significación.

3. Definir el estadígrafo de prueba.

4. Definir la región crítica.

5. Tomar la decisión.

6. Interpretación de la decisión.

Prueba de hipótesis acerca de m con s conocida

- Se conoce que el nivel de glucosa en cierta población sigue una distribución normal con media de 5,88 mmol/L y desviación estándar de 0,22 mmol/L.
- Una muestra de 36 pacientes con Diabetes mellitus Tipo II presenta un nivel promedio de 7,1 mmol/L.
- ¿Será posible concluir que el nivel promedio en esos pacientes es similar al de la población normal?

Prueba de hipótesis acerca de m con s conocida

Población

- μ = 5.88 mmol/L
- σ = 0.22 mmol/L.

Muestra

- n = 36
- $\bar{X}$ = 7.1

Prueba de hipótesis acerca de m con s conocida

- **Formulación de las hipótesis estadísticas.**

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \quad \mu_0 = 5{,}88 \text{ mmol /L}$$

H_0: μ = 5,88 mmol/L

H_1: $\mu \neq$ 5,88 mmol/L

H_0: el nivel promedio de glucosa no difiere de 5,88 mmol/L

H_1: el nivel promedio de glucosa difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s conocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \qquad Z = \frac{7.1 - 5.88}{0.22 / \sqrt{36}} = 33.27$$

Prueba de hipótesis acerca de m con s conocida

- **Definir la región crítica.**

  $\alpha = 0.05$

  $] \; Z_{1-\alpha/2} \, , \, Z_{\alpha/2} \; [$

  $Z_{1-\alpha/2} = -1.96$ $\qquad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

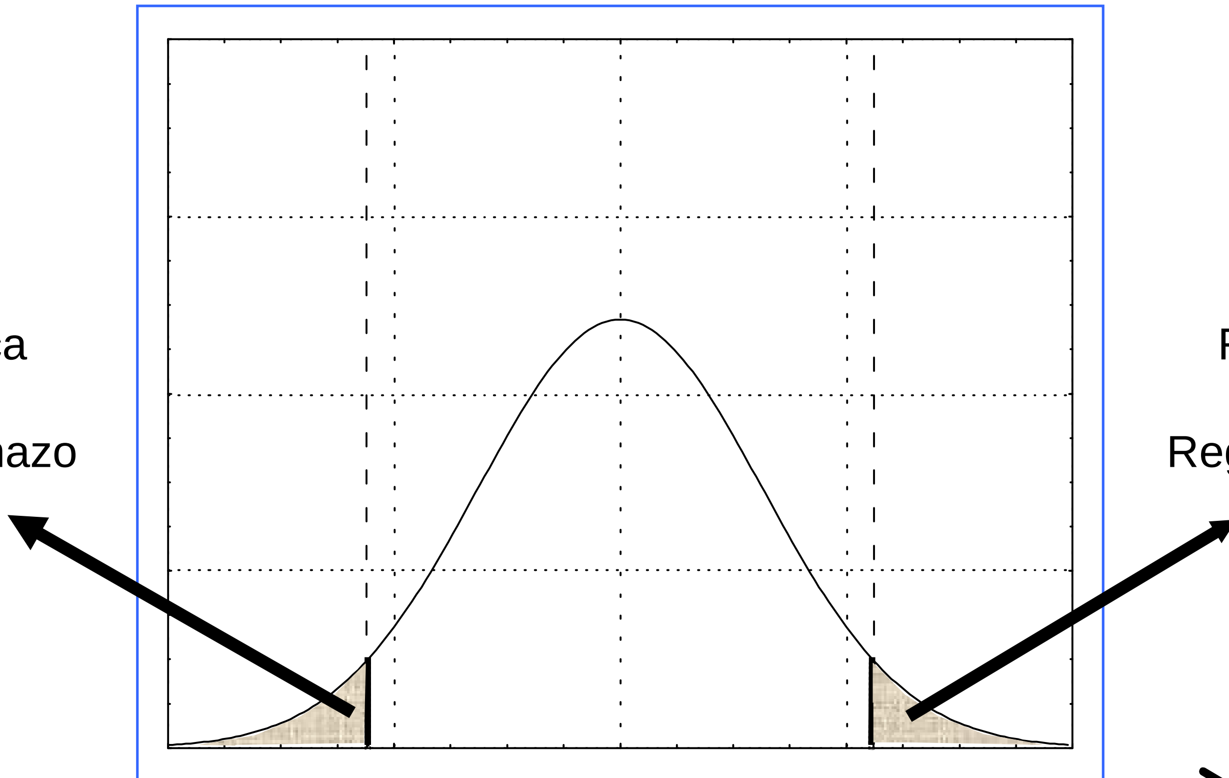
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de m con s conocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -1.96 y > 1.96



Prueba de hipótesis acerca de μ con σ conocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -1.96 y > 1.96
Como el valor del estadígrafo Z de prueba es 33.27,
muy superior a 1.96, este cae en la región crítica y por tanto,
se rechaza H_0

- **Interpretación de la decisión.**

El nivel promedio de glucosa en los pacientes con Diabetes mellitus no es similar al de la población normal
para un nivel de significación de 0,05 o
con un nivel de confiabilidad del 95 % .

H_0: l nivel promedio de glucosa no difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

$n > 30$

- Se conoce que el nivel de glucosa en cierta población sigue una distribución normal con media de 5,88 mmol/L y se desconoce su desviación estándar.Una muestra de 36 pacientes con Diabetes mellitus Tipo II presenta un nivel promedio de 7,1 mmol/L y una desviación estándar de 0,16 mmol/L.

- ¿Será posible concluir que el nivel promedio en esos pacientes es similar al de la población normal?

Prueba de hipótesis acerca de m con s desconocida

Población

- μ = 5.88 mmol/L

Muestra

- n = 36
- $\bar{X}$ = 7.1
- S = 0.16

Prueba de hipótesis acerca de m con s desconocida

- **Formulación de las hipótesis estadísticas.**

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \quad \mu_0 = 5{,}88 \text{ mmol /L}$$

H_0: μ = 5,88 mmol/L

H_1: $\mu \neq$ 5,88 mmol/L

H_0: el nivel promedio de glucosa no difiere de 5,88 mmol/L

H_1: el nivel promedio de glucosa difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N\,(0,1) \qquad Z = \frac{7.1 - 5.88}{0.16/\sqrt{36}} = 45.75$$

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

  $\alpha = 0.05$

  $]\, Z_{1-\alpha/2}\,,\, Z_{\alpha/2}\, [$

  $Z_{1-\alpha/2} = -1.96$ $\qquad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

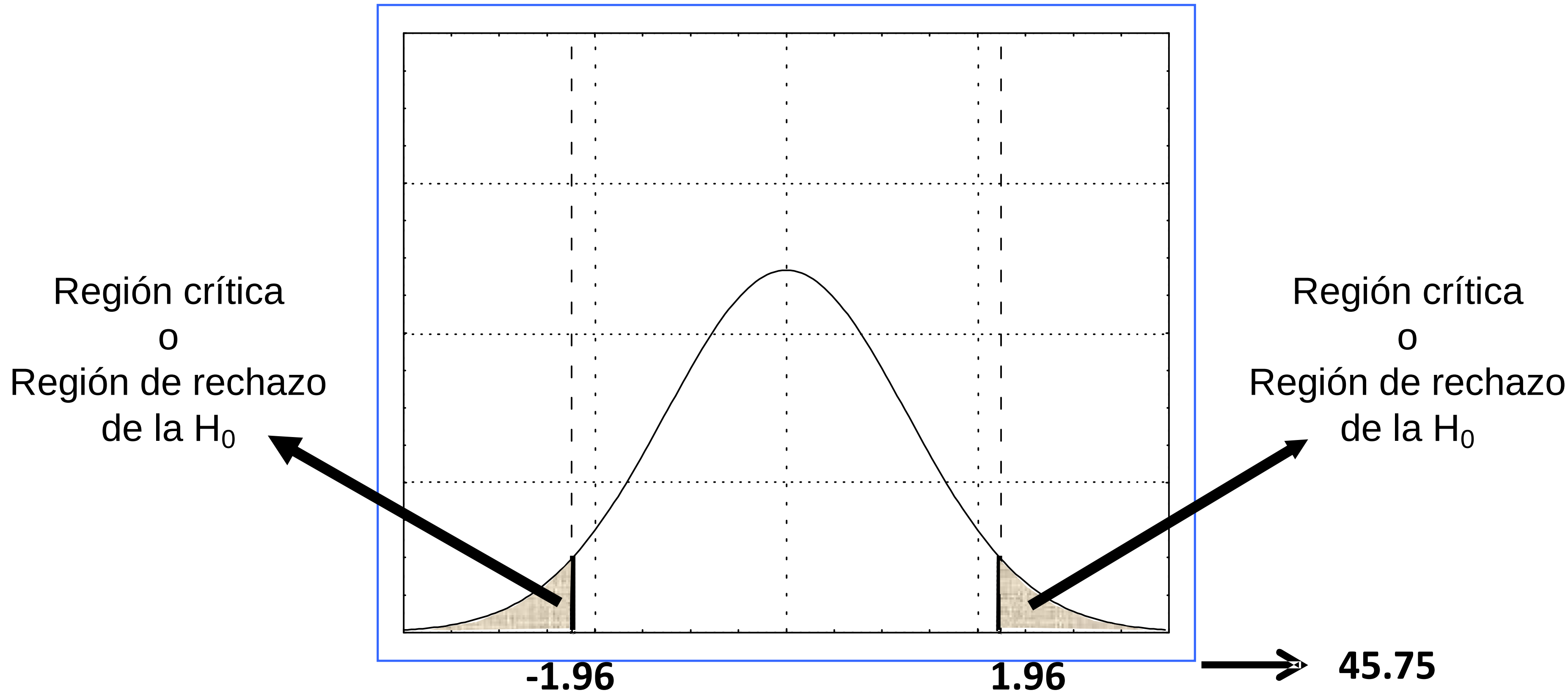
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -1.96 y > 1.96



Prueba de hipótesis acerca de m con s desconocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -1.96 y > 1.96

Como el valor del estadígrafo Z de prueba es 45.75,

muy superior a 1.96, este cae en la región crítica y por tanto,

se rechaza H_0

- **Interpretación de la decisión.**

El nivel promedio de glucosa en los pacientes con Diabetes mellitus no es similar al de la población normal

para un nivel de significación de 0,05 o

con un nivel de confiabilidad del 95 % .

H_0: l nivel promedio de glucosa no difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

$n < 30$

- Se han tomado las determinaciones de uratos en suero de 16 pacientes con determinada enfermedad.
- Si se conoce que el nivel de uratos en suero se distribuye normal. ¿Es posible considerar que el nivel medio de uratos en suero difiere significativamente de 258mmol/L.?
- El tamaño de la muestra es 16 pacientes. A partir de esta información ¿será posible concluir que el nivel medio de uratos en suero de esos pacientes es similar al de la población normal?

340	285	275	275	260	270	235	210
235	210	250	230	250	282	205	285

Prueba de hipótesis acerca de m con s desconocida

Población

- μ = 258 mmol/L

Muestra

- n = 16

$$\overline{X} = \frac{340+285+275+275+260+270+235+210+235+210+250+230+250+282+205+285}{16}$$

$$\overline{X} = 256.06\ mmol/l$$

$$S = \sqrt{\frac{(X_1-\overline{X})+(X_2-\overline{X})+...+(X_n-\overline{X})}{n-1}}$$

$$S = \sqrt{\frac{(X_1-\overline{X})+(X_2-\overline{X})+...+(X_{16}-\overline{X})}{16-1}} = 35.41\ mmol/l$$

Prueba de hipótesis acerca de m con s desconocida

- **Formulación de las hipótesis estadísticas.**

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \quad \mu_0 = 258 \text{ mmol/L}$$

H_0: μ = 258 mmol/L

H_1: $\mu \neq$ 258 mmol/L

H_0: el nivel promedio de glucosa no difiere de 258 mmol/L

H_1: el nivel promedio de glucosa difiere de 258 mmol/L

Prueba de hipótesis acerca de m con s desconocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

$$t = \frac{256.06 - 258}{35.41 / \sqrt{16}} = -0.22$$

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**
  $\alpha = 0.05$
  $] \, t_{n-1\,,\,\alpha/2} \, , \, t_{n-1,\,1-\alpha/2} \, [$
  $t_{15,\,0.025} = -\,2.13$ $t_{15,\,0.75} = 2.13$

Para definir la región crítica que es el paso siguiente se compara el valor del estadígrafo de prueba t con los valores de los percentiles teóricos fijados a priori, es decir, con el conocimiento del valor de α, en este caso $t_{n-1,\alpha/2}$ y $t_{n-1,\,1-\alpha/2}$.

Para un nivel de significación del 5% los valores que determinan la región crítica en una t-Student con 15 grados de libertad son –2.13 y 2.13,

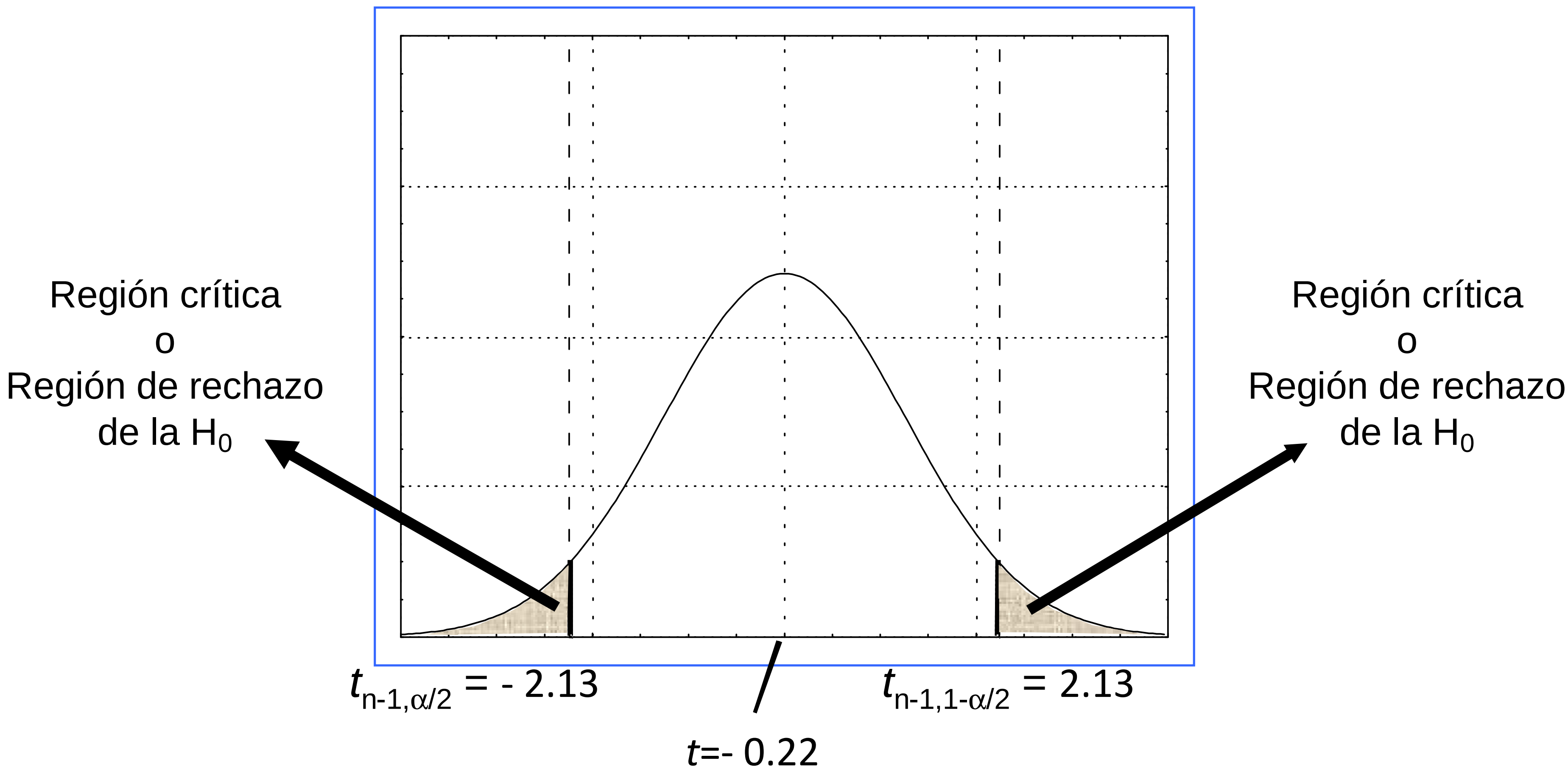
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -2.13 y > 2.13

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -2.13 y > 2.13



Prueba de hipótesis acerca de m con s desconocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -2.13 y > 2.13
Como el valor del estadígrafo *t* de prueba es -0-22,
se encuentra en la región de no rechazo,
no se rechaza H_0

- **Interpretación de la decisión.**

No podemos concluir que el nivel medio de uratos en suero difiera significativamente del valor teórico 258 µmol/L, para un nivel de significación de 0,05 o con un nivel de confiabilidad del 95 % .

H_0: el nivel promedio de glucosa no difiere de 258 mmol/L

Prueba de hipótesis acerca de la proporción poblacional P

De cierta enfermedad del aparato respiratorio se conoce que un 0,30 es su proporción habitual durante muchos años en una población infantil.

En el presente año se toma una muestra representativa de 3000 niños y se registran 600 con dicha afectación.

¿Puede considerarse con un nivel de significación del 0,05 que este año la proporción es como la habitual?

Prueba de hipótesis acerca de la proporción poblacional P

1. Formulación de las hipótesis estadísticas.

$H_0: P = P_0$ $\qquad\qquad$ $P_0 = 0.30$

$H_1: P \neq P_0$

H_0: Proporción de la enfermedad del aparato respiratorio no difiere de 0,30.

H_1: Proporción de la enfermedad del aparato respiratorio difiere de 0,30.

2. Escoger el nivel de significación.

$\alpha = 0.05$

Prueba de hipótesis acerca de la proporción poblacional P

3. Definir el estadígrafo de prueba.

p = 600 / 3000 = 0.20 $\quad\quad$ P_0 = 0.30

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$$

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.20 - 0.30}{\sqrt{\frac{0.30(0.70)}{3000}}} = -11.9$$

Z = -11,9

Prueba de hipótesis acerca de la proporción poblacional P

4. Definir la región crítica.

$\alpha = 0.05$

$]\, Z_{1-\alpha/2}\, ,\, Z_{\alpha/2}\, [$

$Z_{1-\alpha/2} = -1.96$ $\qquad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

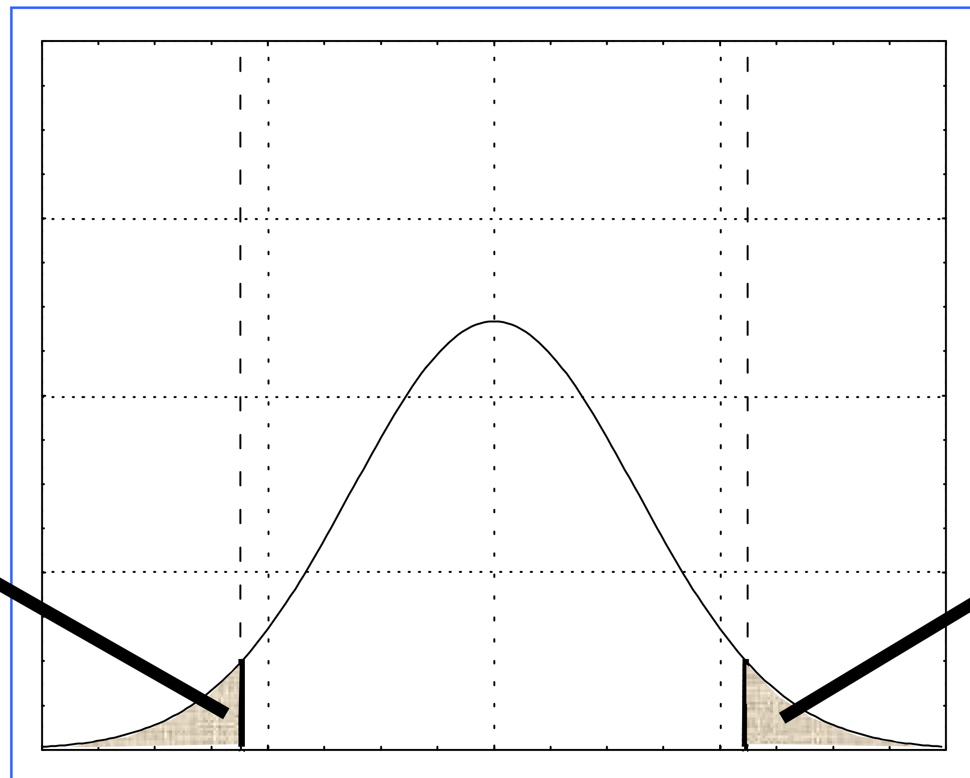
La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de la proporción poblacional P

5. Definir la región crítica.

La región de rechazo se encuentra en < -1.96 y > 1.96

Región crítica
o
Región de rechazo
de la H_0



Región crítica
o
Región de rechazo
de la H_0

$Z_{\alpha/2}$=-1.96

$Z_{1-\alpha/2}$=1.96

Z = -11,9

El valor del estadígrafo Z de prueba es inferior a –1,96,
Cae en la región crítica y por tanto, se rechaza H_0 ,

Concluyendo que:

La proporción de esta enfermedad este año,

si difiere significativamente de 0.30 para el nivel de

significación del 0.05 o un nivel de confiabilidad de

95%.

CONCLUSIONES

$$\sum_{i=1}^{n} X_i$$

H_0: $m = m_0$

H_1: $m \neq m_0$

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$Z = \frac{\overline{X} - m}{s/\sqrt{n}} \sim N(0,1)$$

$$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

- Las pruebas estadísticas encaminadas a verificar la veracidad de una prueba estadística se denominan pruebas de hipótesis.

- La hipótesis a contrastar se denomina Hipótesis nula y se simboliza por H_0. La hipótesis complementaria se nombra Hipótesis alternativa y se simboliza por $H_{1.}$

- Las pruebas de hipótesis son construídas en función de la Hipótesis nula, deberemos expresar sus conclusiones en términos de **rechazo o de no rechazo de** H_0, nunca se debe decir que se acepta H_1.

bvs

Policlínico Universitario Rampa

Biblioteca Virtual

Contraste de hipótesis
Sumario

- Hipótesis nula y alternativa.
- Prueba de hipótesis de una y dos colas.
- Nivel de significación.
- Errores de Tipo I (α) y Tipo II (β)
- Región Crítica.
- Pasos para Formular un Problema de Pruebas de Hipótesis.

- ¿Cómo es posible tomar decisiones acerca de un nuevo método de tratamiento si los resultados no son comparados con otros métodos?
- ¿Cómo es posible corroborar si una prueba diagnostica es más eficiente que otra?
- ¿Cómo es posible establecer si una enfermedad tiene similar comportamiento en dos lugares diferentes?
- ¿Cómo es posible arribar a la conclusión de que los pacientes tratados tienen similar evolución a uno descrito por otros?

Prueba de Hipótesis

- Procedimiento objetivo que permite, sobre la base de una información muestral, tomar una decisión.
- Tiene como objetivo principal, el establecimiento de conclusiones validas para una población estudiando una muestra aleatoria de esta y midiendo el grado de incertidumbre en términos de probabilidad.

Hipótesis

- Proposición cuya verdad o validez no se cuestiona en un primer momento, pero que permite iniciar una cadena de razonamientos que luego puede ser adecuadamente verificada.
- Toma la forma de un enunciado condicional, que debe seguir determinadas reglas para su admisión como razonamiento válido.

Hipótesis

- Afirmacion o suposicion, relacionada con el valor que puede tener un parametro poblacional, del cual depende el modelo teorico de una variable aleatoria.
- Ej.

  La diferencia entre la media de la poblacion de la cual fue tomada la muestra, es cero.

  Ho: $\mu = \mu_o$

  La diferencia entre la media de dos muestras de poblaciones con la misma media, es cero.

  Ho: $\mu_1 = \mu_2$

La hipótesis estadística

- Siempre estará integrada por la hipótesis nula (H_0) y la hipótesis alternativa (H_1), ambas forman una unidad.
- En la hipótesis nula se afirma o considera la no ocurrencia del resultado esperado, el término nula obedece a que es una hipótesis que contiene diferencias nulas.
- La hipótesis alternativa contradice, en algún sentido, la hipótesis nula, se expresa lo contrario a lo que se desea probar.
- De acuerdo a esto, rechazar H_0 equivale a probar el resultado esperado.

La hipótesis estadística

- **Hipotesis Nula**.

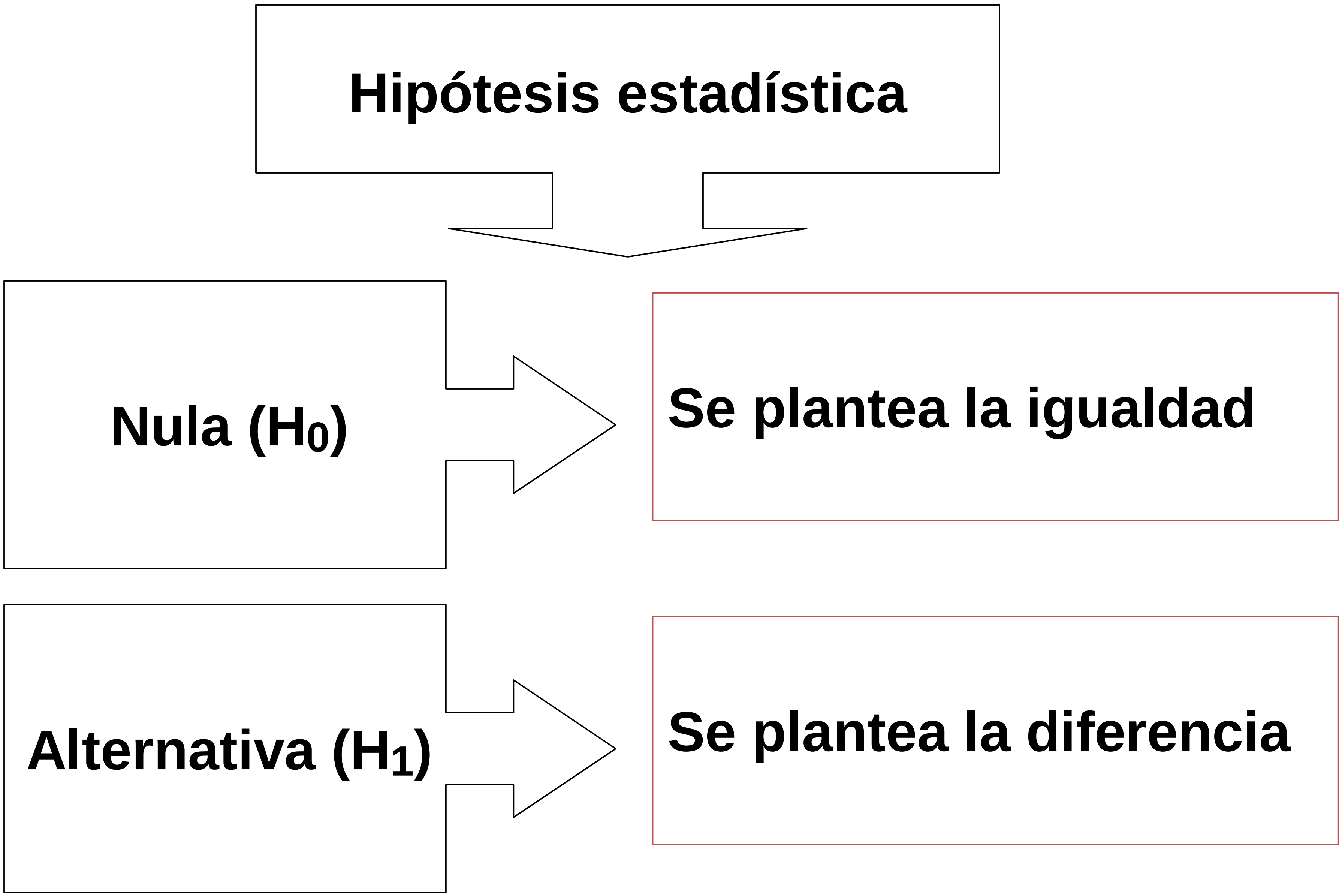
  Es la afirmacion que es probada, puede ser rechazada o no y es lo opuesto de lo que se desea probar.

  Se nombra $\mathbf{H_0}$, plantea que no existen diferencias.

- **Hipotesis Alternativa**.

  Afirmacion de lo que realmente se desea probar.

  Se nombra $\mathbf{H_1}$, plantea que existen diferencias


Hipótesis estadística
Nula (H_0)
Se plantea la igualdad
Alternativa (H_1)
Se plantea la diferencia

Ejemplo

- Se conoce que en una determinada población los niveles de colesterol en sangre se distribuyen normal con media 5,52 mmol/L y desviación estándar 0,16 mmol/L.
- Un grupo de investigadores de un Instituto especializado en Enfermedades Cardiovasculares realizó un estudio en una muestra aleatoria de 150 vegetarianos.

Ejemplo

- A partir de la muestra tomada se determinó el nivel promedio muestral del colesterol en dicho conjunto de personas.
- La hipótesis científica a ser probada estadísticamente es si individuos cuya ingesta está compuesta sólo por vegetales, tienen los mismos niveles de colesterol que la población general.

Ejemplo

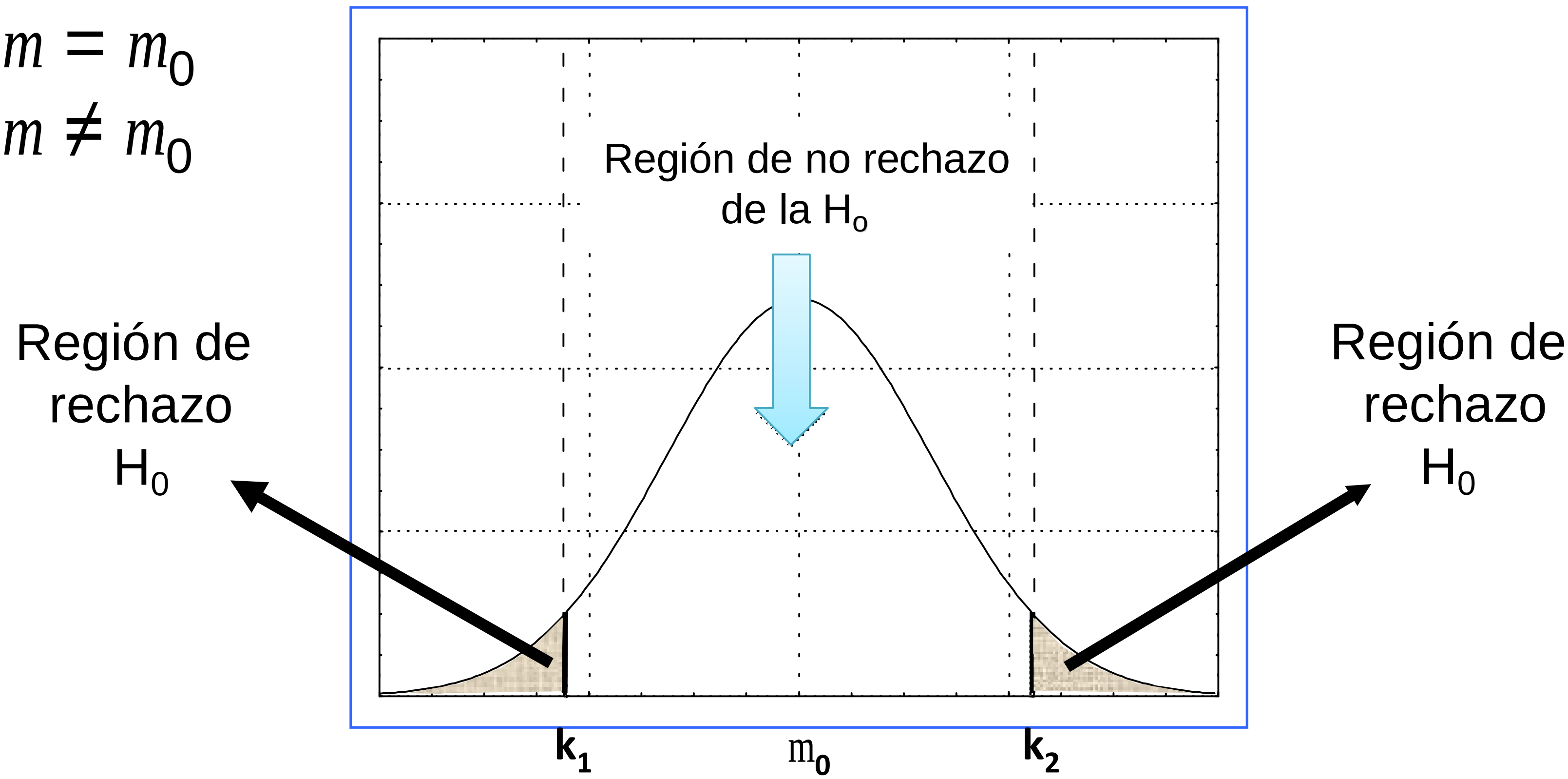
- La hipótesis nula sería: H_0: $\mu = \mu_0 = 5{,}52$, donde $\mu_0 = 5{,}52$ representa lo general conocido.
- El sentido de nula, viene dado, por que no se espera encontrar ninguna variación en el colesterol medio de los vegetarianos, salvo la debida al azar.
- La hipótesis alternativa sería: H_1: $\mu \neq \mu_0 = 5{,}52$. En este caso, los investigadores, sólo están interesados en probar que existen diferencias entre el promedio de colesterol de vegetarianos y el de la población total.

Ejemplo

- Ellos estarán de acuerdo, en rechazar H_0, tanto si el nivel medio de colesterol en vegetarianos es menor que un valor k_1, como si es mayor que un valor k_2.

H_0: $m = m_0$

H_1: $m \neq m_0$



Ejemplo

- Ahora, se impone, encontrar alguna forma de poder arribar a una conclusión con respecto a, rechazar o no H_0, incluyendo, al igual que en el caso de los intervalos de confianza, la información contenida en una única muestra.
- En general, si el valor del estimador, digamos la media o la proporción muestral o cualquier otro, es tal que hay que rechazar la hipótesis nula, se dice que se ha obtenido un resultado significativo con un **nivel de significación** dado.

Nivel de significación

- Es posible que a causa de, la aleatoriedad de las observaciones muestrales, la estimación obtenida se desvíe tanto de lo esperado, que se tome, la decisión de rechazar H_0, siendo sin embargo H_0 cierta o verdadera.
- Es lógico o conveniente por tanto, que la probabilidad de que esto suceda, sea pequeña.
- Dentro de esta metodología, esta probabilidad, recibe el nombre de **nivel de significación**.

Nivel de significación

- El nivel de significación es un valor arbitrario seleccionado a priori por el investigador de acuerdo a su experiencia y deseo.
- Los valores que con más frecuencia se utilizan son 0,05 y 0,01. Se acostumbra a denotar este valor por la letra griega α.
- El uso del término, significación, es debido a que la diferencia entre, el valor hipotético o teórico y el hallado en la muestra, se considera lo, suficientemente grande, como para que no sea solamente atribuible al azar; es decir, que el concepto se refiere al estado de ser, estadísticamente significativo.

Errores de Tipo I (α) y Tipo II (β)

- Al tomar la decisión con respecto a la hipótesis nula podemos incurrir en dos errores diferentes.

- Cuando, **H_0** es **verdadera** y **se rechaza**, se produce el **error de tipo I**, con probabilidad de ocurrencia α, es decir, el nivel de significación de la prueba de hipótesis.

Errores de Tipo I (α) y Tipo II (β)

- Al tomar la decisión con respecto a la hipótesis nula podemos incurrir en dos errores diferentes.

- **Si no se rechaza** la hipótesis nula **H_0 y es falsa** se incurre entonces en el llamado **error de tipo II**.

- La probabilidad de cometer este tipo de error se conoce como β.

- Por su complejidad, no es objetivo de nuestro Curso ocuparnos de este error.

Errores de Tipo I (α) y Tipo II (β)

Si H_0 es	Desición sobre H_0	
	Rechazar	No rechazar
Verdadera	**Error de tipo I**	Acción correcta
Falsa	Acción correcta	**Error de tipo II**

Estadígrafos de la prueba

Situación	Distribución	Estadígrafo
Para μ σ conocida	Normal estándar	$Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$
σ desconocida $n > 30$	Normal estándar	$Z=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$
σ desconocida $n \leq 30$	*t*-Student	$t=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$
Proporciones para muestras grandes	Normal estándar	$Z=\dfrac{p-P}{\sqrt{p.(1-p)/n}}$

Segun la forma que H_0 es planteada, las pruebas de hipótesis se clasifica en:

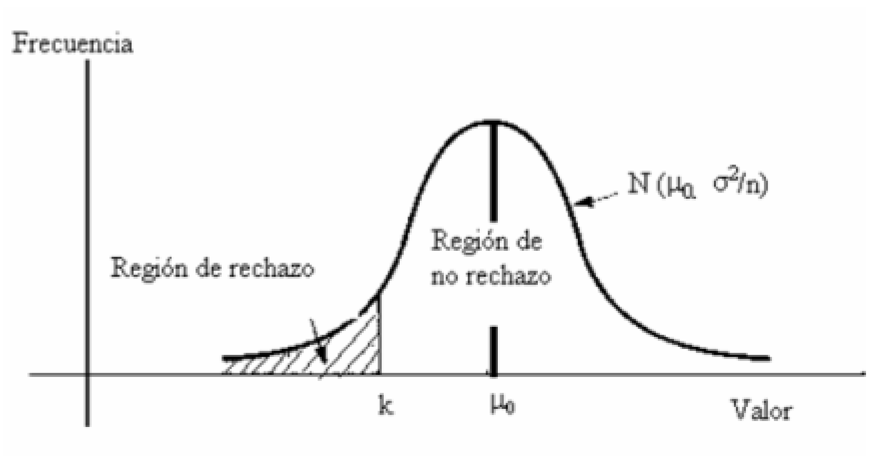
- Prueba de Hipotesis de <u>una cola</u> o <u>unilatera</u>l

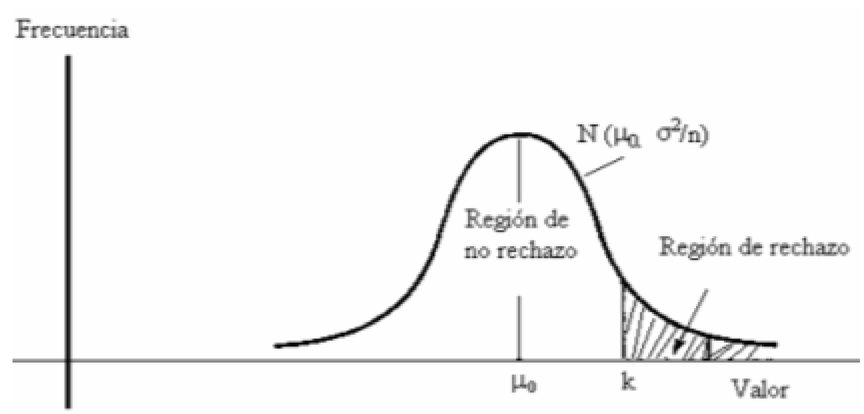
H_0: $m \leq m_0$,

H_1: $m > m_0$

H_0: $m \geq m_0$,

H_1: $m < m_0$





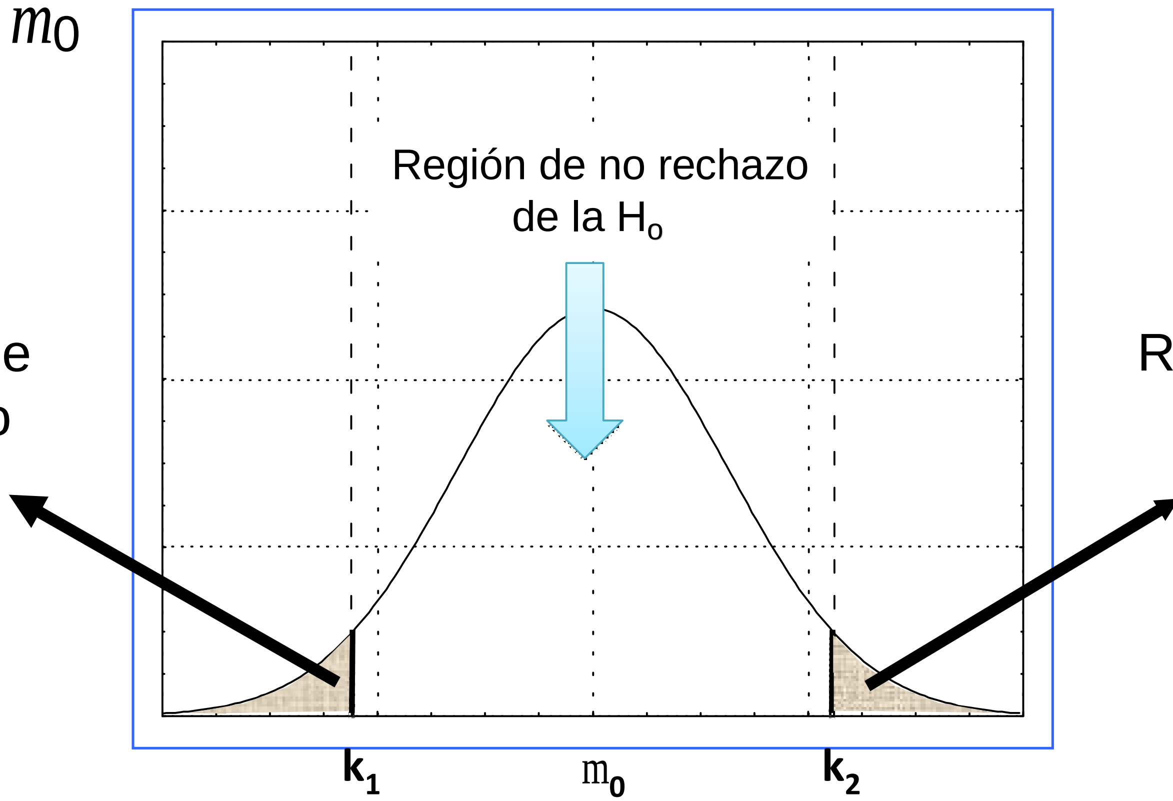
Segun la forma que H_0 es planteada, las pruebas de hipótesis se clasifica en:

- Prueba de Hipotesis <u>dos colas</u> o <u>bilateral</u>

H_0: $m = m_0$,

H_1: $m \neq m_0$

Región de rechazo H_0



Región de rechazo H_0

Región crítica

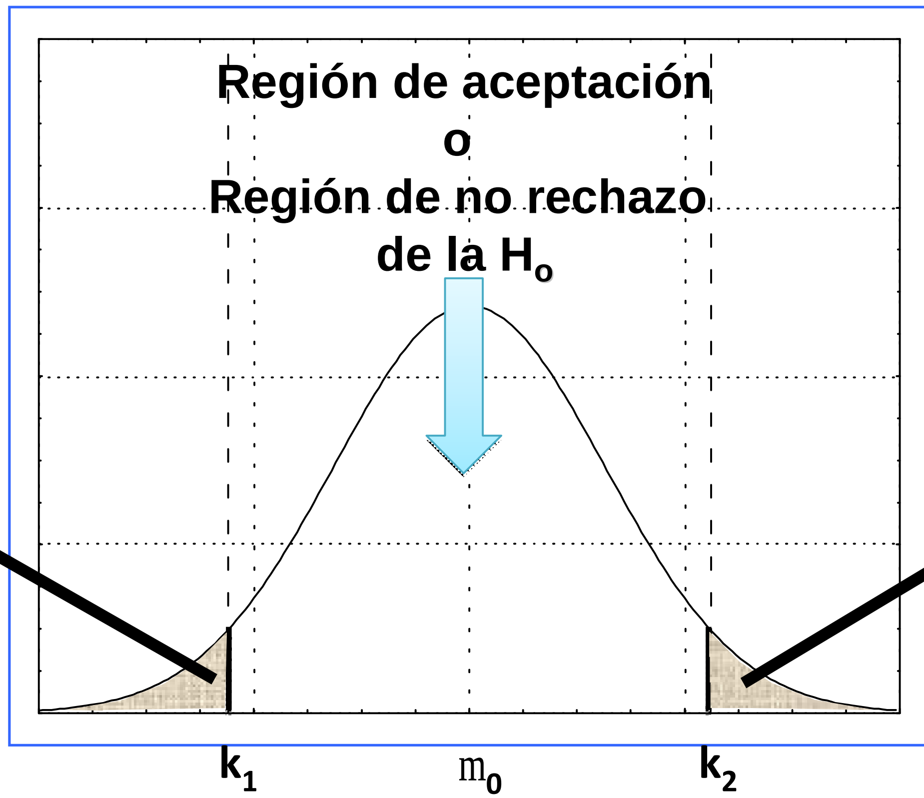
- La región crítica es aquella que se encuentra fuera de la **región de aceptación o de no rechazo de la H_0**
- En la **región de aceptación** se encuentran los valores que determinan la región de no rechazo entre los que esta concentrada el 95% (α=0.05) o el 99% (α=0.01) de la información. Es un intervalo de confianza para el 95 0 el 99% de confianza.

Región crítica

- Los datos que se ubican fuera de ese intervalo de confianza, en las llamadas colas de la distribución (Región crítica o de rechazo) ocurren menos frecuentemente que el resto.
- Se tomará la decisión de **rechazar H_0**, con un nivel de significación α, si el valor estimado del parámetro está en la **región crítica**
- y de **no rechazar H_0** si este valor **no está en la región crítica**.

Región crítica

Región crítica
o
Región de rechazo
de la H_0



Región crítica
o
Región de rechazo
de la H_0

Pasos a seguir

1. Formulación de las hipótesis estadísticas.
2. Escoger el nivel de significación.
3. Definir el estadígrafo de prueba.
4. Definir la región crítica.
5. Tomar la decisión.
6. Interpretación de la decisión.

Prueba de hipótesis acerca de μ con σ conocida

- Se conoce que el nivel de glucosa en cierta población sigue una distribución normal con media de 5,88 mmol/L y desviación estándar de 0,22 mmol/L.
- Una muestra de 36 pacientes con Diabetes mellitus Tipo II presenta un nivel promedio de 7,1 mmol/L.
- ¿Será posible concluir que el nivel promedio en esos pacientes es similar al de la población normal?

Prueba de hipótesis acerca de m con s conocida

Población

- μ = 5.88 mmol/L
- σ = 0.22 mmol/L.

Muestra

- n = 36
- $\bar{X}$ = 7.1

Prueba de hipótesis acerca de μ con σ conocida

- **Formulación de las hipótesis estadísticas.**

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \quad \mu_0 = 5{,}88 \text{ mmol/L}$$

H_0: μ = 5,88 mmol/L

H_1: $\mu \neq$ 5,88 mmol/L

H_0: el nivel promedio de glucosa no difiere de 5,88 mmol/L

H_1: el nivel promedio de glucosa difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s conocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \qquad Z = \frac{7.1 - 5.88}{0.22 / \sqrt{36}} = 33.27$$

Prueba de hipótesis acerca de m con s conocida

- **Definir la región crítica.**

  $\alpha = 0.05$

  $]\, Z_{1-\alpha/2} , Z_{\alpha/2} \,[$

  $Z_{1-\alpha/2} = -1.96$ $\qquad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

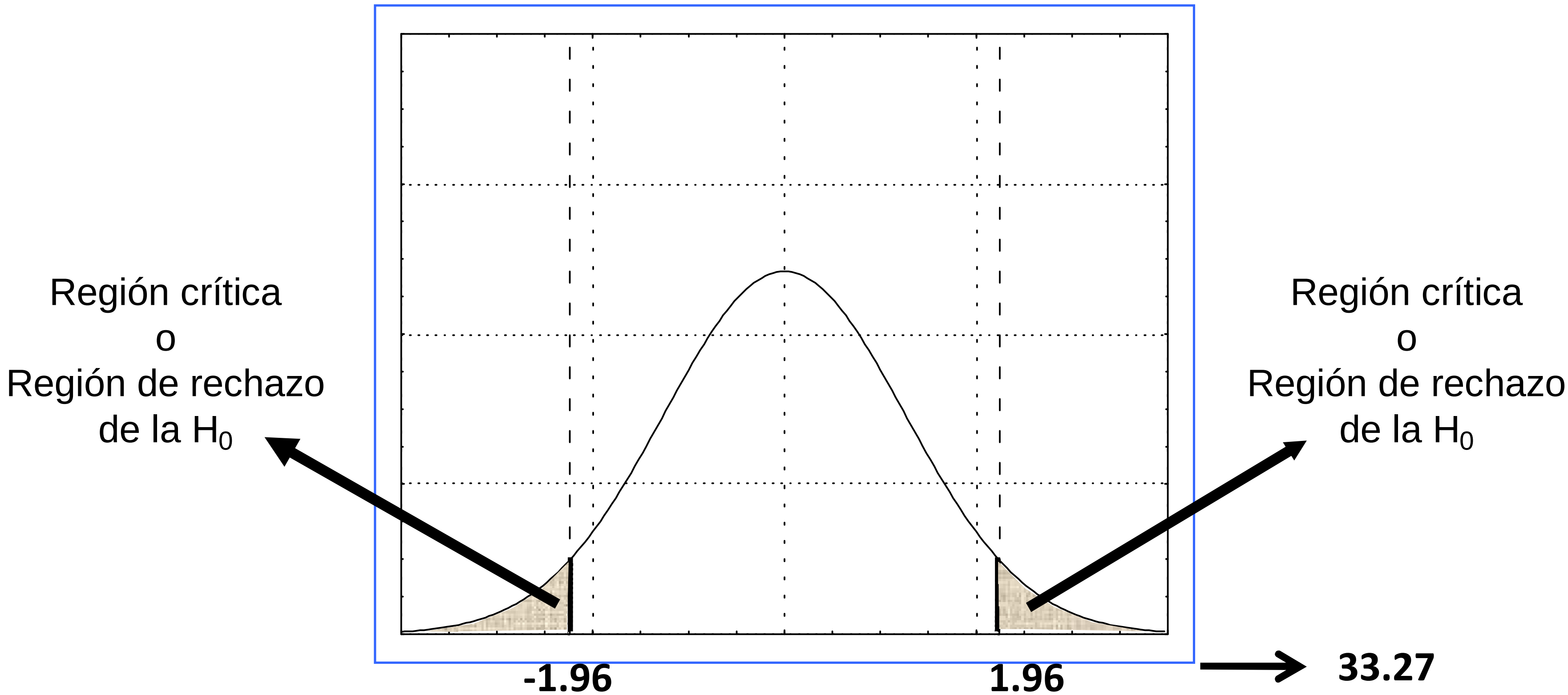
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de m con s conocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -1.96 y > 1.96



Prueba de hipótesis acerca de μ con σ conocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -1.96 y > 1.96
Como el valor del estadígrafo Z de prueba es 33.27,
muy superior a 1.96, este cae en la región crítica y por tanto,
se rechaza H_0

- **Interpretación de la decisión.**

El nivel promedio de glucosa en los pacientes con Diabetes mellitus no es similar al de la población normal
para un nivel de significación de 0,05 o
con un nivel de confiabilidad del 95 % .

H_0: l nivel promedio de glucosa no difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

n > 30

- Se conoce que el nivel de glucosa en cierta población sigue una distribución normal con media de 5,88 mmol/L y se desconoce su desviación estándar.Una muestra de 36 pacientes con Diabetes mellitus Tipo II presenta un nivel promedio de 7,1 mmol/L y una desviación estándar de 0,16 mmol/L.
- ¿Será posible concluir que el nivel promedio en esos pacientes es similar al de la población normal?

Prueba de hipótesis acerca de m con s desconocida

Población

- μ = 5.88 mmol/L

Muestra

- n = 36
- $\bar{X}$ = 7.1
- S = 0.16

Prueba de hipótesis acerca de μ con σ desconocida

- **Formulación de las hipótesis estadísticas.**

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

$\mu_0 = 5{,}88$ mmol /L

H_0: $\mu = 5{,}88$ mmol/L

H_1: $\mu \neq 5{,}88$ mmol/L

H_0: el nivel promedio de glucosa no difiere de 5,88 mmol/L

H_1: el nivel promedio de glucosa difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1) \qquad Z = \frac{7.1 - 5.88}{0.16/\sqrt{36}} = 45.75$$

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

  $\alpha = 0.05$

  $] \; Z_{1-\alpha/2} \, , \, Z_{\alpha/2} \; [$

  $Z_{1-\alpha/2} = -1.96$ $\qquad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

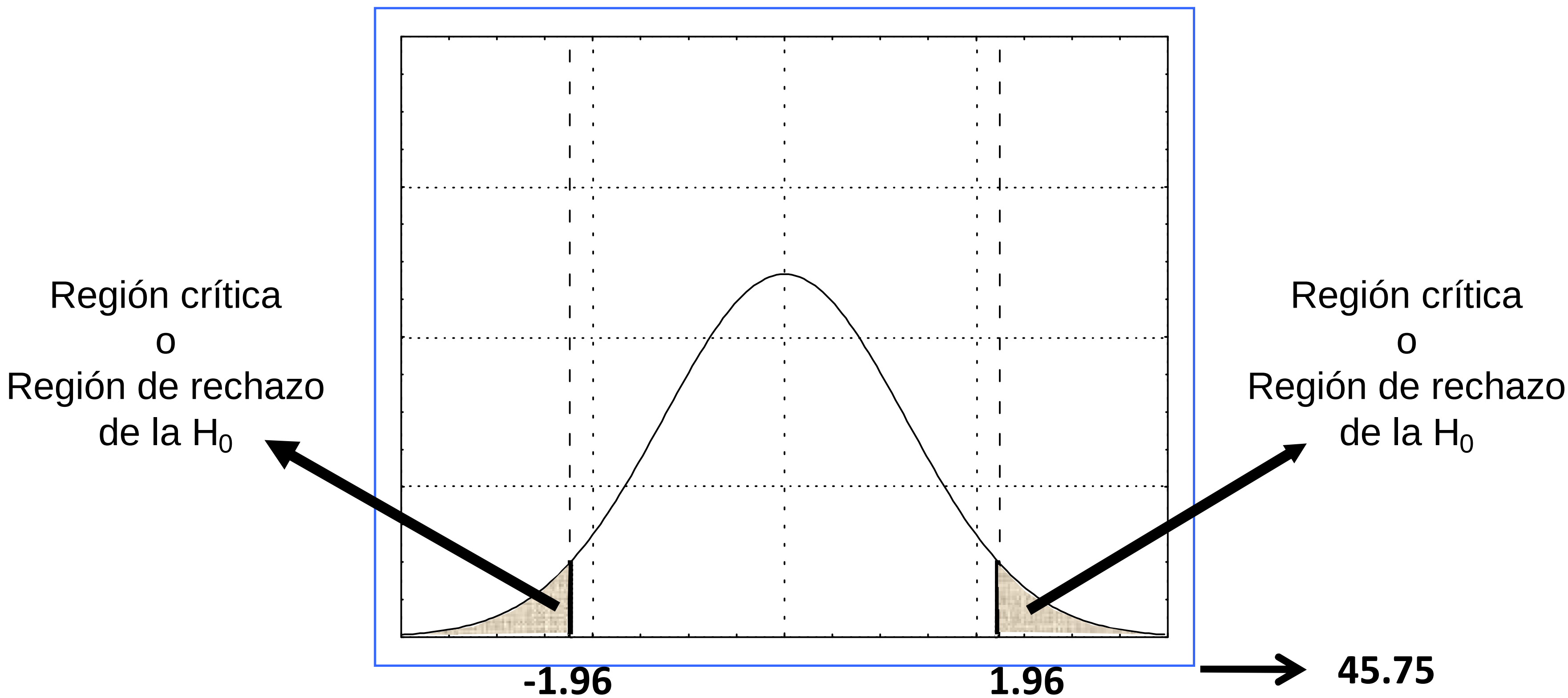
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -1.96 y > 1.96



Prueba de hipótesis acerca de m con s desconocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -1.96 y > 1.96
Como el valor del estadígrafo Z de prueba es 45.75,
muy superior a 1.96, este cae en la región crítica y por tanto,
se rechaza H_0

- **Interpretación de la decisión.**

El nivel promedio de glucosa en los pacientes con Diabetes mellitus no es similar al de la población normal
para un nivel de significación de 0,05 o
con un nivel de confiabilidad del 95 % .

H_0: l nivel promedio de glucosa no difiere de 5,88 mmol/L

Prueba de hipótesis acerca de m con s desconocida

n < 30

- Se han tomado las determinaciones de uratos en suero de 16 pacientes con determinada enfermedad.
- Si se conoce que el nivel de uratos en suero se distribuye normal. ¿Es posible considerar que el nivel medio de uratos en suero difiere significativamente de 258mmol/L.?
- El tamaño de la muestra es 16 pacientes. A partir de esta información ¿será posible concluir que el nivel medio de uratos en suero de esos pacientes es similar al de la población normal?

340	285	275	275	260	270	235	210
235	210	250	230	250	282	205	285

Prueba de hipótesis acerca de m con s desconocida

Población

- µ = 258 mmol/L

Muestra

- n = 16

$$\overline{X} = \frac{340+285+275+275+260+270+235+210+235+210+250+230+250+282+205+285}{16}$$

$$\overline{X} = 256.06\ mmol/l$$

$$S = \sqrt{\frac{(X_1-\overline{X})+(X_2-\overline{X})+...+(X_n-\overline{X})}{n-1}}$$

$$S = \sqrt{\frac{(X_1-\overline{X})+(X_2-\overline{X})+...+(X_{16}-\overline{X})}{16-1}} = 35.41\ mmol/l$$

Prueba de hipótesis acerca de μ con σ desconocida

- **Formulación de las hipótesis estadísticas.**

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \quad \mu_0 = 258 \text{ mmol/L}$$

H_0: μ = 258 mmol/L

H_1: $\mu \neq$ 258 mmol/L

H_0: el nivel promedio de glucosa no difiere de 258 mmol/L

H_1: el nivel promedio de glucosa difiere de 258 mmol/L

Prueba de hipótesis acerca de μ con σ desconocida

- **Escoger el nivel de significación.**

$\alpha = 0.05$

- **Definir el estadígrafo de prueba.**

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

$$t = \frac{256.06 - 258}{35.41 / \sqrt{16}} = -0.22$$

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

  $\alpha = 0.05$

  $]\ t_{n-1\ ,\ \alpha/2}\ ,\ t_{n-1,\ 1-\ \alpha/2}\ [$

  $t_{15,\ 0.025} = -\ 2.13 \quad t_{15,\ 0.75} = 2.13$

Para definir la región crítica que es el paso siguiente se compara el valor del estadígrafo de prueba t con los valores de los percentiles teóricos fijados a priori, es decir, con el conocimiento del valor de α, en este caso $t_{n-1,\alpha/2}$ y $t_{n-1,\ 1-\alpha/2}$.

Para un nivel de significación del 5% los valores que determinan la región crítica en una t-Student con 15 grados de libertad son –2.13 y 2.13,

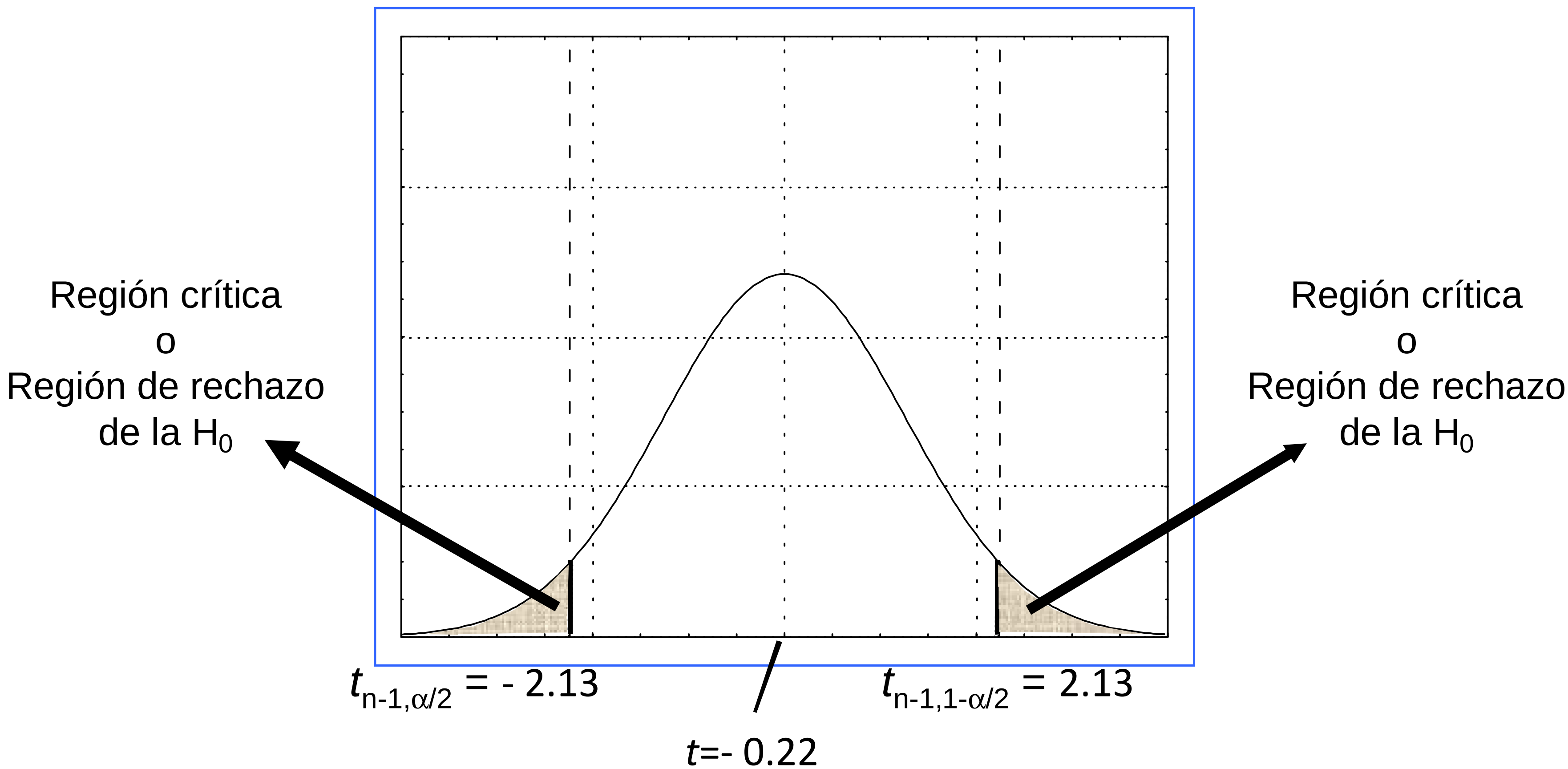
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -2.13 y > 2.13

Prueba de hipótesis acerca de m con s desconocida

- **Definir la región crítica.**

La región de rechazo se encuentra en < -2.13 y > 2.13



Prueba de hipótesis acerca de m con s desconocida

- **Tomar la decisión.**

La región de rechazo se encuentra en < -2.13 y > 2.13
Como el valor del estadígrafo *t* de prueba es -0-22,
se encuentra en la región de no rechazo,
no se rechaza H_0

- **Interpretación de la decisión.**

No podemos concluir que el nivel medio de uratos en suero difiera significativamente del valor teórico 258 μmol/L, para un nivel de significación de 0,05 o con un nivel de confiabilidad del 95 % .

H_0: el nivel promedio de glucosa no difiere de 258 mmol/L

Prueba de hipótesis acerca de la proporción poblacional P

De cierta enfermedad del aparato respiratorio se conoce que un 0,30 es su proporción habitual durante muchos años en una población infantil.

En el presente año se toma una muestra representativa de 3000 niños y se registran 600 con dicha afectación.

¿Puede considerarse con un nivel de significación del 0,05 que este año la proporción es como la habitual?

Prueba de hipótesis acerca de la proporción poblacional P

1. Formulación de las hipótesis estadísticas.

$H_0: P = P_0$ $\qquad\qquad$ $P_0 = 0.30$

$H_1: P \neq P_0$

H_0: Proporción de la enfermedad del aparato respiratorio no difiere de 0,30.

H_1: Proporción de la enfermedad del aparato respiratorio difiere de 0,30.

2. Escoger el nivel de significación.

$\alpha = 0.05$

Prueba de hipótesis acerca de la proporción poblacional P

3. Definir el estadígrafo de prueba.

p = 600 / 3000 = 0.20 $\qquad$ P_0 = 0.30

$$Z = \frac{p-P}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1)$$

$$Z = \frac{p-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.20-0.30}{\sqrt{\frac{0.30(0.70)}{3000}}} = -11.9$$

Z = -11,9

Prueba de hipótesis acerca de la proporción poblacional P

4. Definir la región crítica.

$\alpha = 0.05$

$] \, Z_{1-\alpha/2} \, , \, Z_{\alpha/2} \, [$

$Z_{1-\alpha/2} = -1.96$ $\quad\quad$ $Z_{\alpha/2} = 1.96$

Para definir la región crítica se compara el valor del estadígrafo de prueba Z con los valores de los percentiles teóricos fijados a priori para el valor de α, en este caso $Z_{\alpha/2}$ y $Z_{1-\alpha/2}$.

Para α= 0,05, alrededor de la media más/menos dos desviaciones estándar está concentrada aproximadamente el 95% de la información.

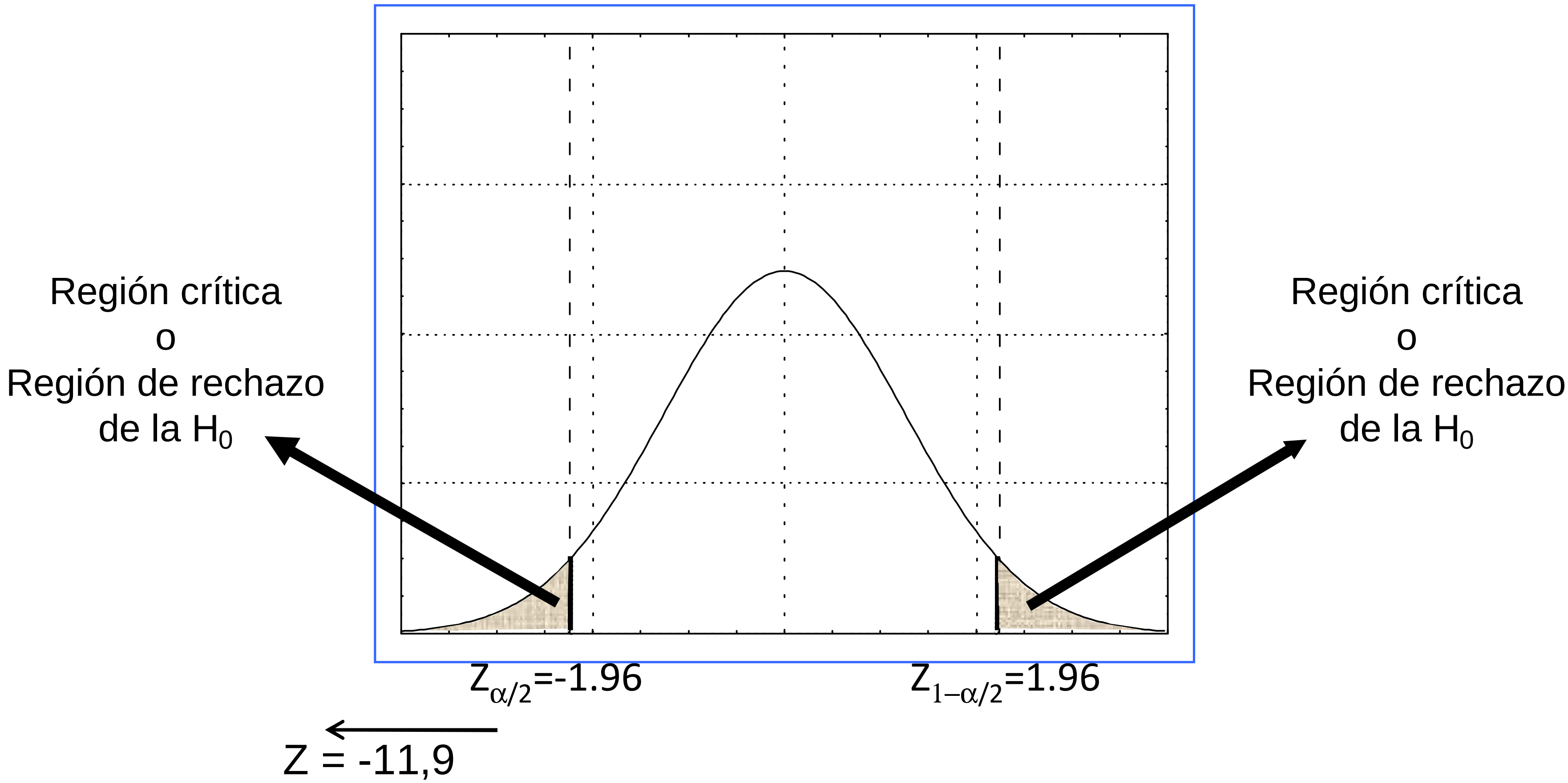
Se espera entonces que los datos que se ubican en las llamadas colas de la distribución ocurran menos frecuentemente que el resto.

La región de rechazo se encuentra en < -1.96 y > 1.96

Prueba de hipótesis acerca de la proporción poblacional P

5. Definir la región crítica.

La región de rechazo se encuentra en < -1.96 y > 1.96



El valor del estadígrafo Z de prueba es inferior a –1,96,
Cae en la región crítica y por tanto, se rechaza H_0 ,

Concluyendo que:

La proporción de esta enfermedad este año,

si difiere significativamente de 0.30 para el nivel de

significación del 0.05 o un nivel de confiabilidad de

95%.

CONCLUSIONES

$$\sum_{i=1}^{n} X_i$$

$H_0: m = m_0$

$H_1: m \neq m_0$

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

$$Z = \frac{\overline{X} - m}{s / \sqrt{n}} \sim N\,(0,1)$$

$$Z = \frac{p - P}{\sqrt{\dfrac{P\,(1 - P\,)}{n}}}$$

- Las pruebas estadísticas encaminadas a verificar la veracidad de una prueba estadística se denominan pruebas de hipótesis.

- La hipótesis a contrastar se denomina Hipótesis nula y se simboliza por H_0. La hipótesis complementaria se nombra Hipótesis alternativa y se simboliza por $H_{1.}$

- Las pruebas de hipótesis son construídas en función de la Hipótesis nula, deberemos expresar sus conclusiones en términos de **rechazo o de no rechazo de** H_0, nunca se debe decir que se acepta H_1.

bvs

Policlínico Universitario Rampa

Biblioteca Virtual

http://www.policlinicarampa.sld.cu/recursos.html