Clase de Repaso

Segundo Parcial

Variables aleatorias continuas

- Si $f(x)$ es una **función de densidad de probabilidad** entonces:
  - $f(x) \geq 0$ para cualquier valor de x
  - $\int_{-\infty}^{\infty} f(x)dx = 1$ equivale a $\sum_i p(x_i) = 1$ en discretas.

Si X es cualquier v.a., la función de distribución acumulada de X, se escribe
$F(x) = P(X < x) = P(X \leq x)$

Variables aleatorias continuas

Si $F(x)$ es la función de distribución de una v.a. continua X, entonces $f(x)$ está dada por:

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

siempre y cuando exista la derivada.

De lo anterior se tiene que:

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt$$

Entonces:

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dt$$

Variables aleatorias continuas

El valor esperado de una v.a. continua es

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \mu$$

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

También es válida la expresión alternativa:

$$Var(X) = \sigma^2 = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x) dx \; - \mu^2$$

Distribución Uniforme

Una variable aleatoria posee distribución uniforme en el intervalo de 6 a 10.

$$E(Y) = \frac{(a+b)}{2}$$

$$\sigma^2 = Var(Y) = \frac{(b-a)^2}{12}$$

Distribución Normal

10) Suponga que la variable aleatoria X sigue una distribución normal con $\mu = 80$ y $\sigma^2 = 100$

.

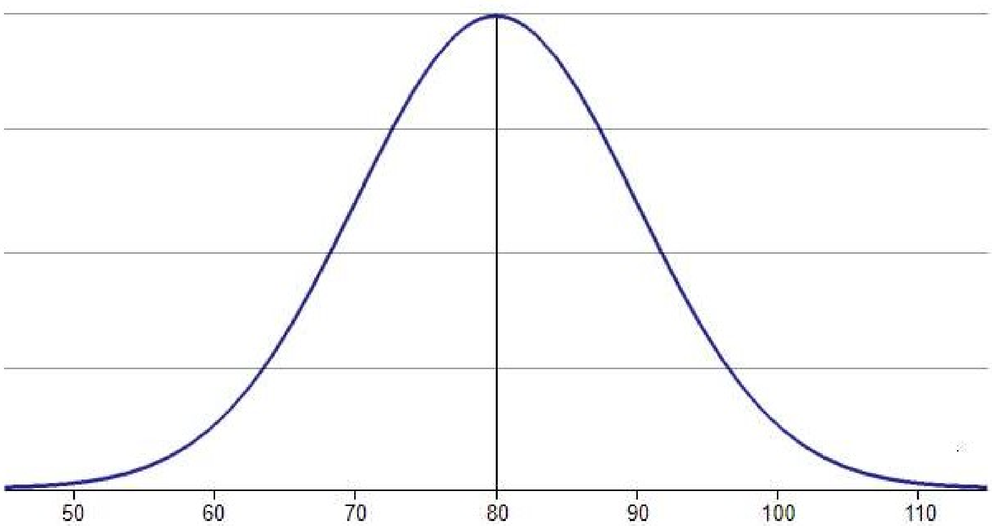
a) Halle la probabilidad de que X sea superior a 60.

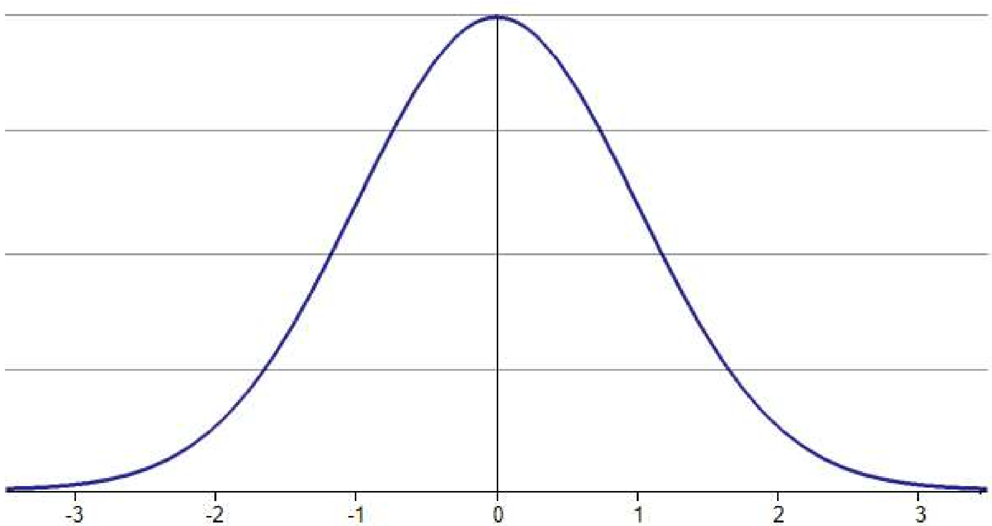
b) Halle la probabilidad de que X sea superior a 72 e inferior a 82.

c) Halle la probabilidad de que X sea inferior a 60.

d) La probabilidad de que X sea superior a ______ es 0,1.

$$Z = \frac{X - \mu}{\sigma}$$





Distribución Media Muestral

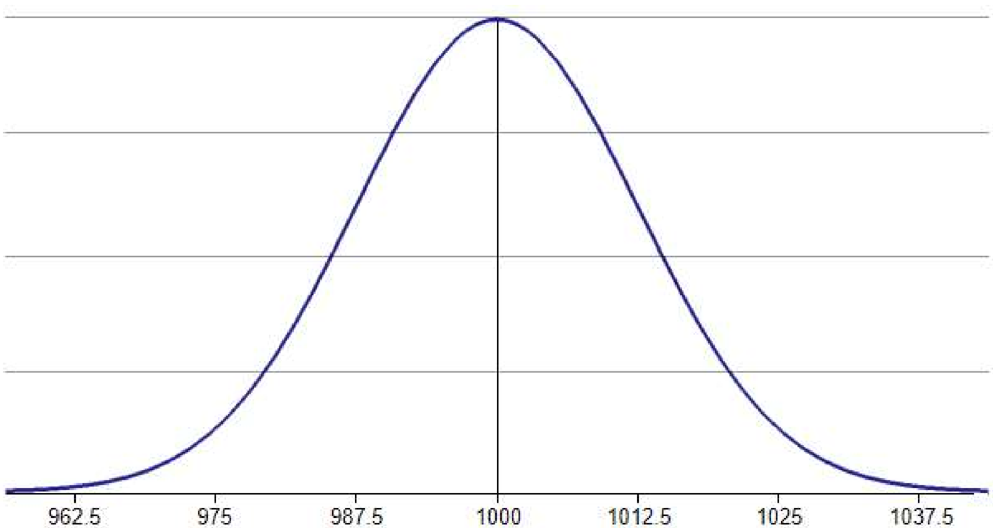
6) Una empresa decide realizar una prueba de aptitud a sus futuros empleados. Sabe, en base a estudios anteriores, que la cantidad media de aciertos es de 1000 con una desviación estándar de 125. Si se aplica la prueba a 100 individuos seleccionados al azar:

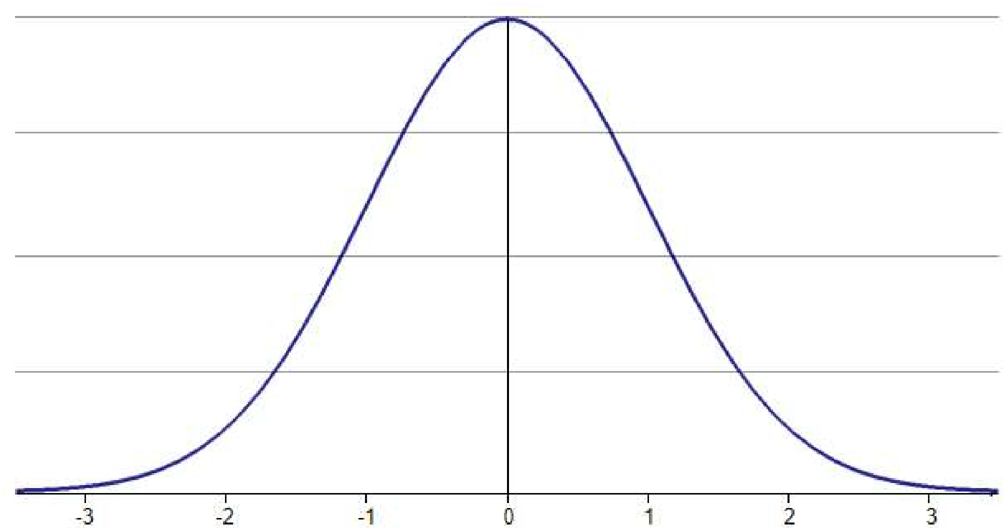
$$Z = \frac{\overline{X}_n - \mu_x}{\left(\frac{\sigma_x}{\sqrt{n}}\right)}$$

a) ¿Cuál es la probabilidad de que la media muestral se encuentre entre 985 y 1015 aciertos?

b) ¿Cuál es la probabilidad de que la media muestral sea de como mínimo 1020 aciertos?

c) ¿Cuál es la probabilidad de que la media muestral se encuentre entre 960 y 1040 aciertos?





Distribución Proporción Muestral

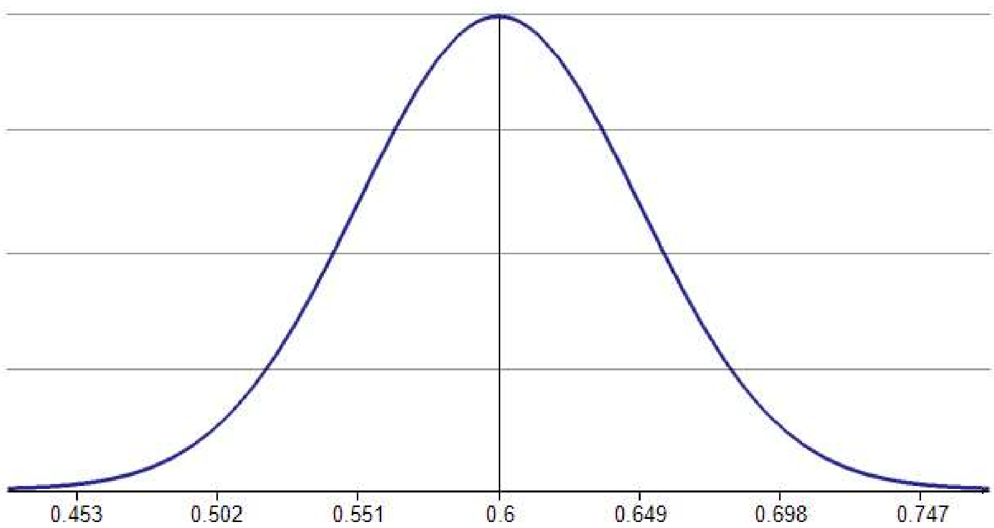
17) Suponga que tenemos una población con una proporción $p = 0{,}60$ y una muestra aleatoria de tamaño $n = 100$ extraída de la población.

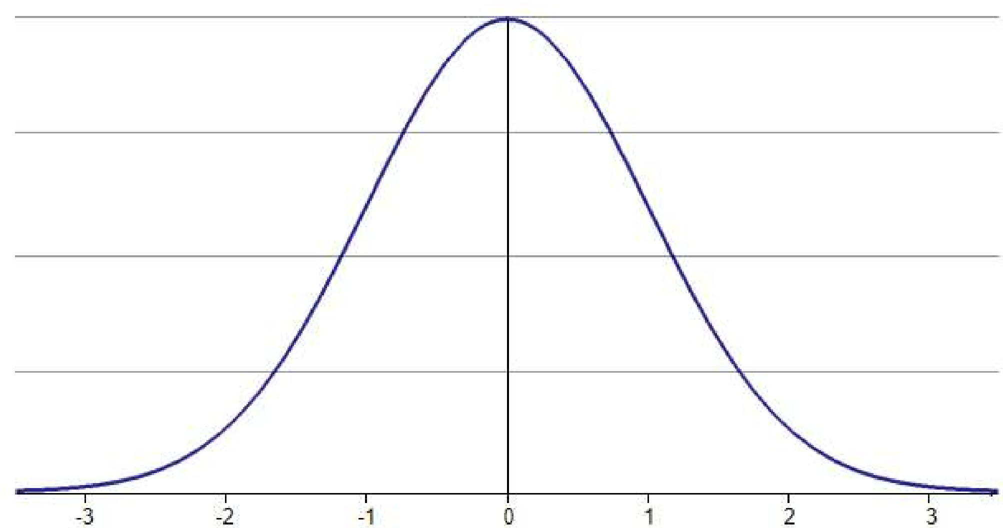
a) ¿Cuál es la probabilidad de que la proporción muestral sea superior a 0,66?

b) ¿Cuál es la probabilidad de que la proporción muestral sea inferior a 0,48?

c) ¿Cuál es la probabilidad de que la proporción muestral esté comprendida entre 0,52 y 0,66?

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





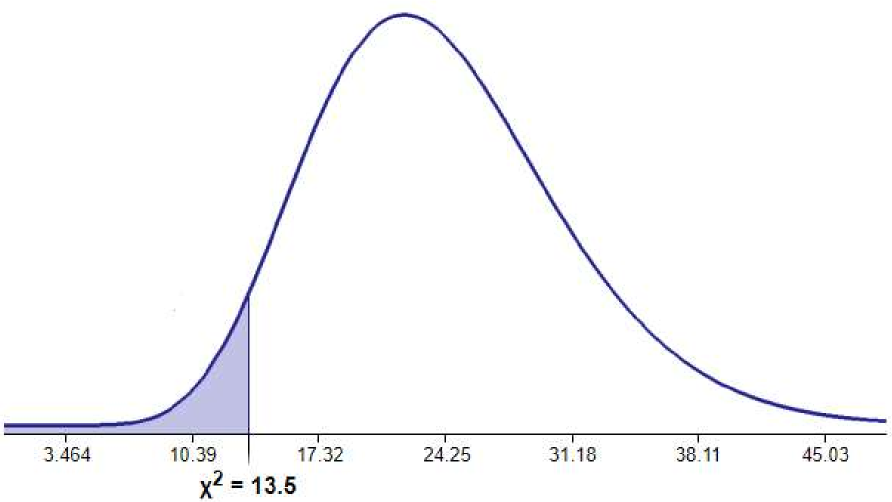
Distribución Varianza Muestral

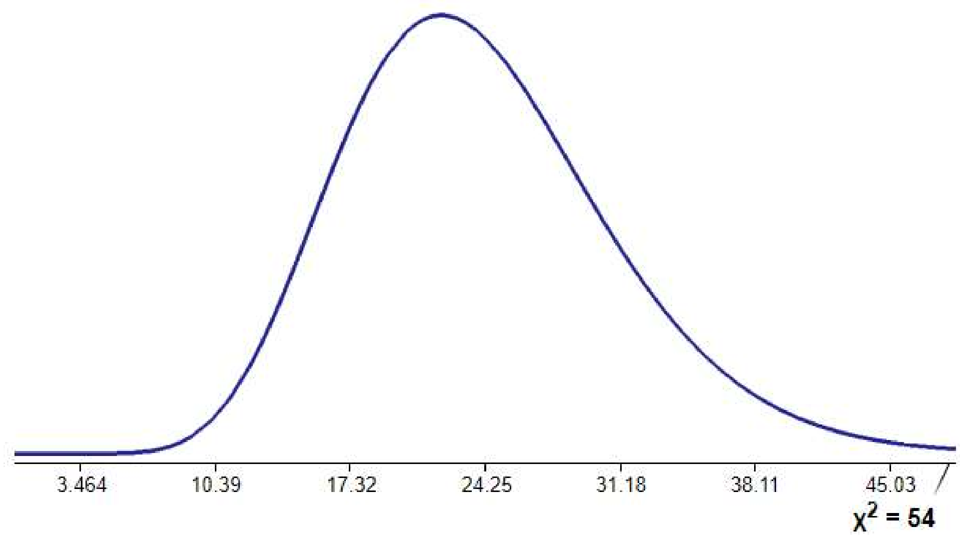
25) En una gran ciudad se ha observado que durante el verano las facturas del consumo de electricidad siguen una distribución normal que tiene una desviación típica de \$100. Se ha tomado una muestra aleatoria de 25 facturas.

a) Halle la probabilidad de que la varianza muestral sea inferior a 5625.

b) Halle la probabilidad de que la desviación típica muestral sea superior a \$150.

$$\frac{(n-1)s_X^2}{\sigma_X^2} \text{ se distribuye según una } \chi^2_{n-1}$$





Propiedades de los Estimadores

- **Estimador insesgado**: $E\left(\hat{\theta}\right)=\theta$
- **Eficiencia:** $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si $\mathrm{var}\left(\hat{\theta}_1\right)<\mathrm{var}\left(\hat{\theta}_2\right)$
- **Consistencia:** equivale a decir que el estimador se vuelve insesgado cuando n crece y que la varianza del estimador tiende a cero cuando $n \to \infty$.

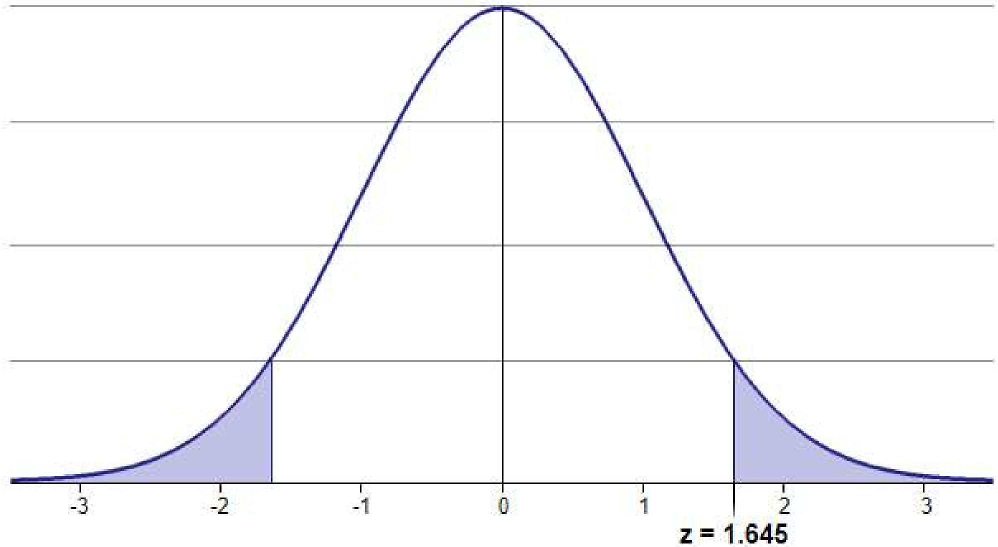
Propiedades de los Estimadores

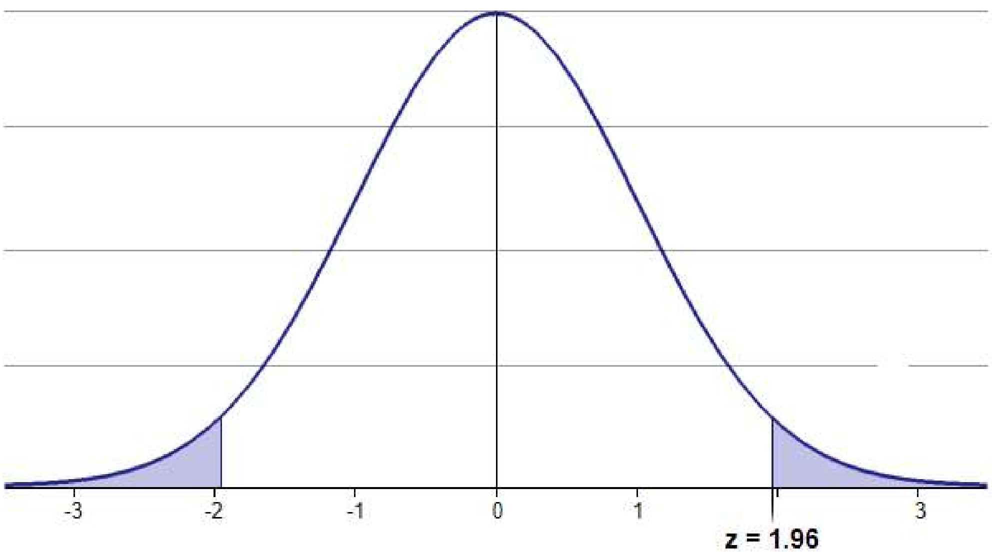
27) Suponga que se tiene una muestra aleatoria de tamaño $2n$ tomada de una población X, con $E[X]=\mu$ y $Var[X]=\sigma^2$. Sean:

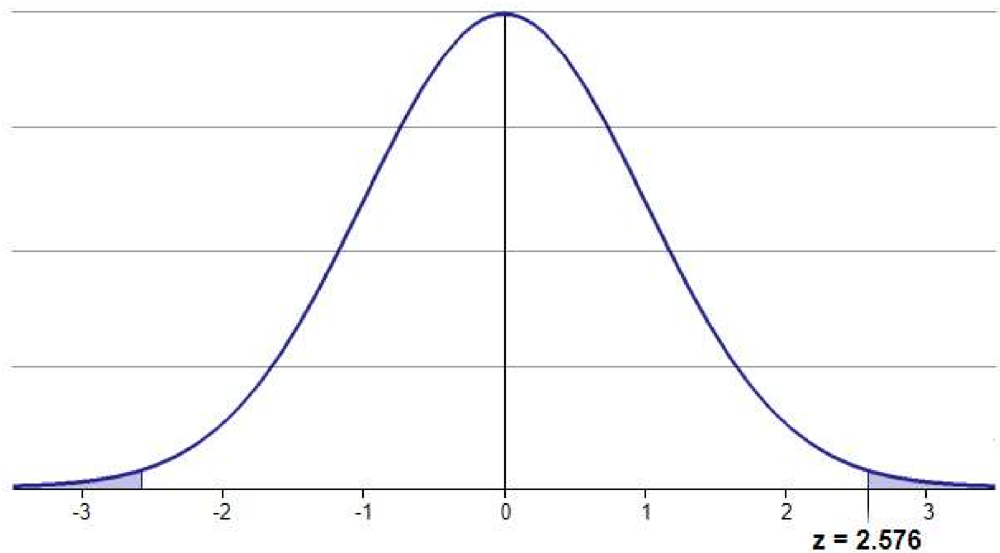
$$\overline{X}_1=\frac{1}{2n}\sum_{i=1}^{2n}X_i \qquad y \qquad \overline{X}_2=\frac{1}{n}\sum_{i=1}^{n}X_i$$

Intervalos de Confianza

Región de confianza $(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
90%	0.10	0.05	1,645
95%	0.05	0.025	1.96
99%	0.01	0.005	2,576







Intervalos de Confianza

Parámetro	Muestra	Distribución poblacional	σ	Intervalo de confianza
Media	n grande ($n \geq 30$?)	Cualquiera	Conocida	$\bar{X} \pm Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
Media	n grande ($n \geq 30$?)	Cualquiera	Desconocida	$\bar{X} \pm Z_{\alpha/2} \dfrac{S}{\sqrt{n}}$
Media	$n \leq 30$	Debe ser aprox. normal	Conocida	$\bar{X} \pm Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
Media	$n \leq 30$	Debe ser aprox. normal	Desconocida	$\bar{X} \pm t_{n-1,\alpha/2} \dfrac{S}{\sqrt{n}}$
Proporción	$np \geq 10$ $n(1-p) \geq 10$			$\hat{p} \pm Z_{\alpha/2} \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
Varianza		Normal		$\dfrac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$

Intervalos de Confianza

Inferencia sobre µ

$¿n \geq 30?$

SI

Uso la Tabla Normal

Si conozco σ, uso σ $\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Si no conozco σ, uso s $\bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$

NO

¿La Población es Normal?

NO

No puedo hacer nada

SI

¿Conozco σ?

SI

$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ Uso la Tabla Normal

NO

Uso t de Student con s en lugar de σ $\bar{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$

Intervalos de Confianza

Inferencia sobre p

$$\left.\begin{array}{l} ¿\,np \geq 10? \\ ¿\,nq \geq 10? \end{array}\right\}$$ Uso la Tabla Normal $$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Inferencia sobre σ o σ^2

¿La Población es Normal? $\xrightarrow{\text{SI}}$ Uso χ^2_{n-1} $$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$$

Intervalos de Confianza

- **¡Ojo con la interpretación del intervalo!**
- Hay una probabilidad del 95% de que la verdadera media poblacional este dentro del intervalo [……,……] **INCORRECTO**
- Hay una probabilidad del 95% de que cualquier intervalo de confianza generado a partir de una muestra aleatoria de tamaño n contenga a la verdadera media poblacional. **CORRECTO**

$$\uparrow \text{nivel de confianza} \Rightarrow \uparrow \text{ancho} \Rightarrow \uparrow \text{certeza} \;\; \text{pero} \;\; \downarrow \text{precision}$$

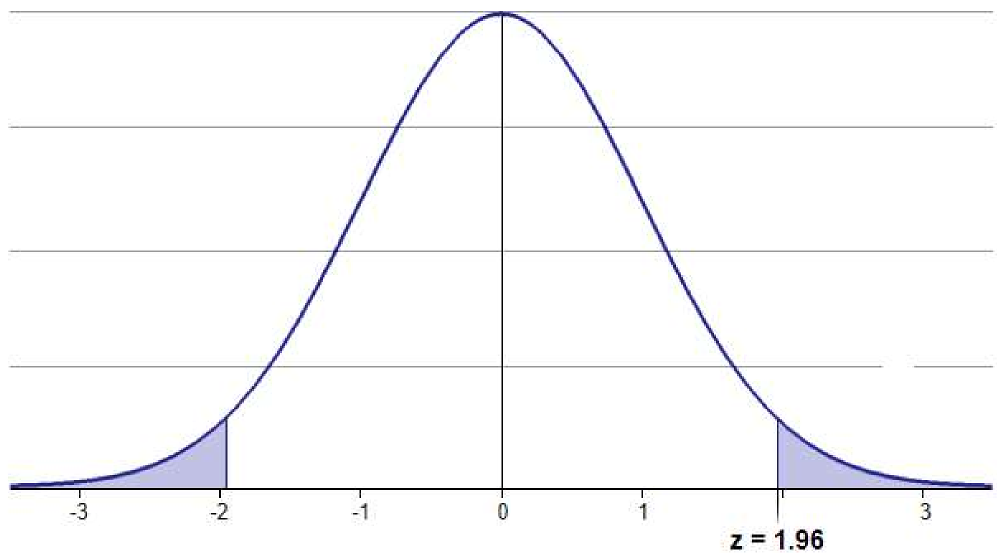
Intervalos de Confianza

11) Una muestra de 352 suscriptores de la revista Wired indicó que el tiempo medio invertido en el uso de internet es de 13,4 hs a la semana con una desviación estándar de 6,8 hs. Determinar un intervalo de confianza del 95% para el tiempo medio que pasan los suscriptores en internet.

$$\bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

$$13{,}4 \ \pm 1{,}96 * \frac{6{,}8}{\sqrt{352}} = 13{,}4 \pm 0{,}7104$$

$$(12{,}6897, \qquad 14{,}1104)$$



Intervalos de Confianza

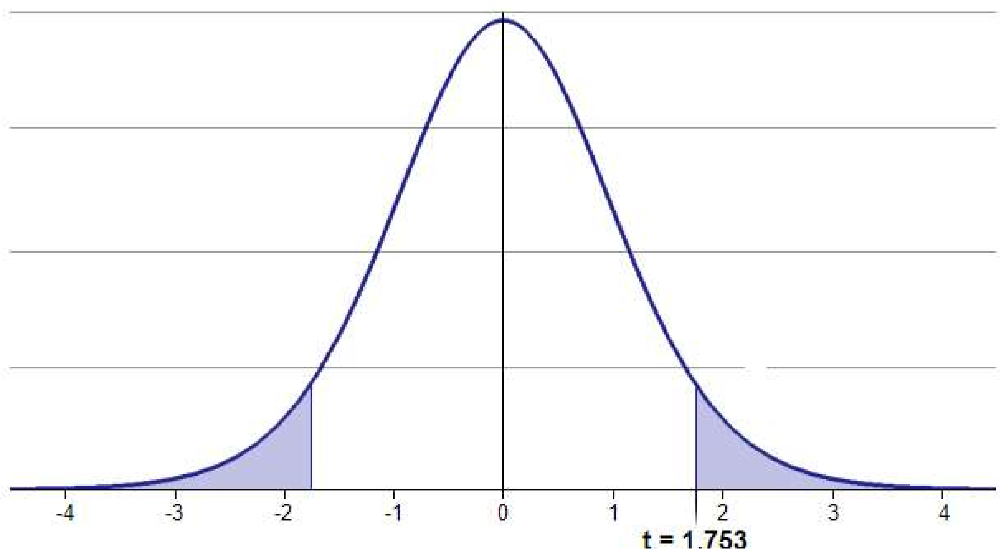
7) La asociación de productores de azúcar desea estimar el consumo medio de azúcar por año. Una muestra de 16 personas revela que el consumo medio anual es de 60 libras con una desviación estándar de 20 libras.

c) Construir un intervalo de confianza del 90% para la media poblacional

$$\overline{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$$

$$60 \ \pm 1{,}753 * \frac{20}{\sqrt{16}} = 60 \pm 8{,}765$$

$$(51{,}235, \qquad 68{,}765)$$



Intervalos de Confianza

12) De un proceso productivo del que se obtienen piezas seriadas se seleccionó una muestra de 350 piezas encontrándose 18 defectuosas:

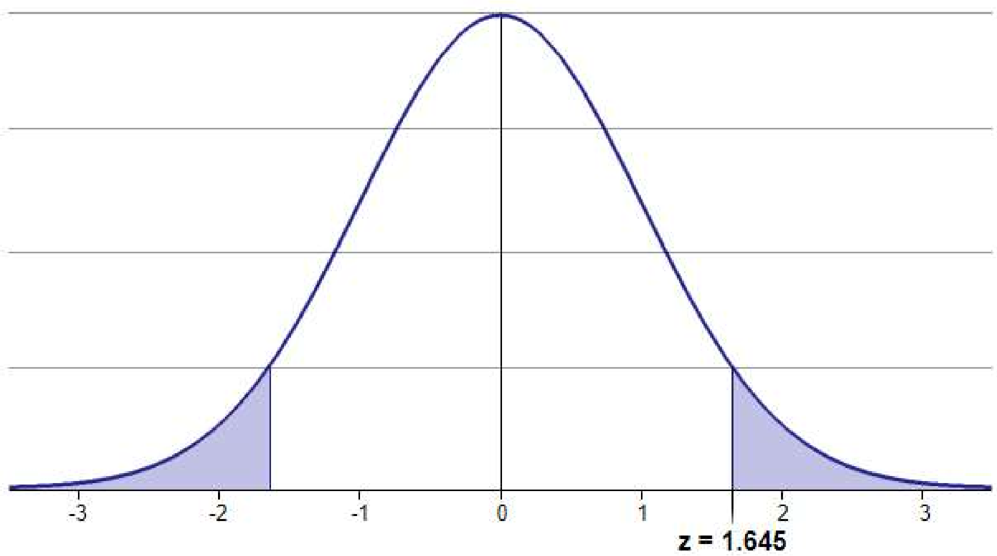
a) Compruebe los supuestos que deben cumplirse y estime el porcentaje defectuoso del proceso con una confianza del 90%.

$n\hat{p} = 350 * {}^{18}/_{350} = 18 > 10$ $\qquad$ $n(1-\hat{p}) = 350 * \left(1 - {}^{18}/_{350}\right) = 332 > 10$

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{18}{350} \pm 1{,}645 * \sqrt{\frac{\frac{18}{350}\left(1-\frac{18}{350}\right)}{350}} = \frac{18}{350} \pm 0{,}0194$$

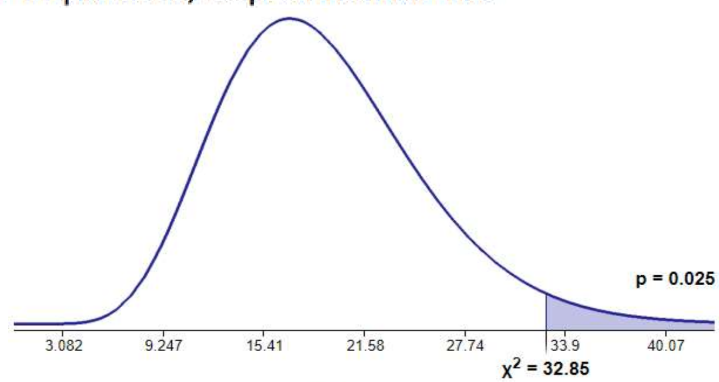
$$(0{,}032, \qquad 0{,}071)$$



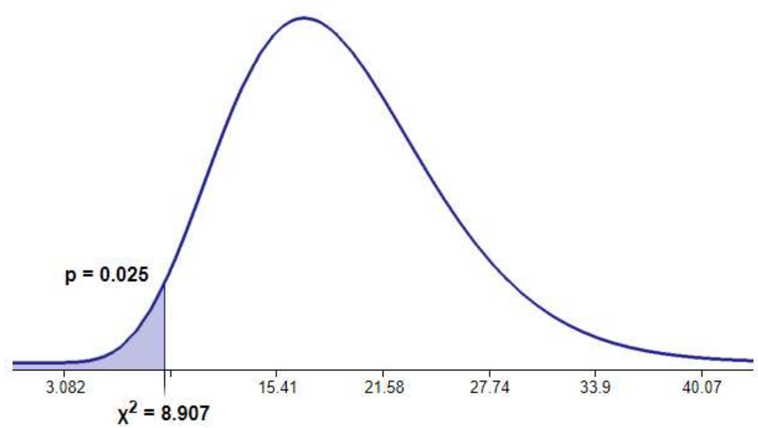
Intervalos de Confianza

16) El director de control de calidad de una empresa química ha extraído una muestra aleatoria de veinte sacos de fertilizantes de 100 kilos para estimar la varianza de los kilos de impurezas. Se ha observado que la varianza muestral es de 6,62. Halle el intervalo de confianza al 95% de la varianza poblacional de los kilos de impurezas, especificando los supuestos necesarios para construir este intervalo.

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$$



$$\frac{(20-1)*6{,}62}{32{,}85} < \sigma^2 \frac{(20-1)*6{,}62}{8{,}907}$$



$(3{,}8289, \qquad 14{,}1215)$

Test de Hipótesis

- El rendimiento promedio de determinado auto es de 14 km por litro de combustible. El grupo de I&D dice haber mejorado el rendimiento de un nuevo carburador. A tal fin fabricarán varios carburadores y los pondrán a prueba. ¿Cual es la hipótesis nula y alternativa que se desea testear?

Test Unilateral Derecho

H_0: $\mu \leq 14$ vs. H_A: $\mu > 14$

H_0: $\mu = 14$ vs. H_A: $\mu > 14$

- Una planta embotelladora afirma que sus botellas de 2 litros contienen un promedio mínimo de 1997 ml. Se selecciona una muestra de botellas y se medirán sus contenidos para investigar la afirmación del fabricante. Plantear H_0 y H_A

Test Unilateral Izquierdo

$H_0\!: \mu \geq 1997 \qquad H_A\!: \mu < 1997$

$H_0\!: \mu = 1997 \qquad H_A\!: \mu < 1997$

- Se acaba de recibir un embarque y se debe someter a un control de calidad, para aceptar todo el embarque o regresarlo al proveedor porque no cumple con las especificaciones. Supongamos que se tratan de piezas que deben medir 5 cm. Si la longitud promedio de las piezas es mayor o menor a la norma de 5 cm se rechaza el embarque. Plantear H_0 y H_A

Test Bilateral

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

Test de Hipótesis

- El rendimiento promedio de determinado auto es de 14 km por litro de combustible. El grupo de I&D dice haber mejorado el rendimiento de un nuevo carburador. A tal fin fabricarán varios carburadores y los pondrán a prueba. ¿Cual es la hipótesis nula y alternativa que se desea testear?

Test Unilateral Derecho

H_0: $\mu \leq 14$ vs. H_A: $\mu > 14$

H_0: $\mu = 14$ vs. H_A: $\mu > 14$

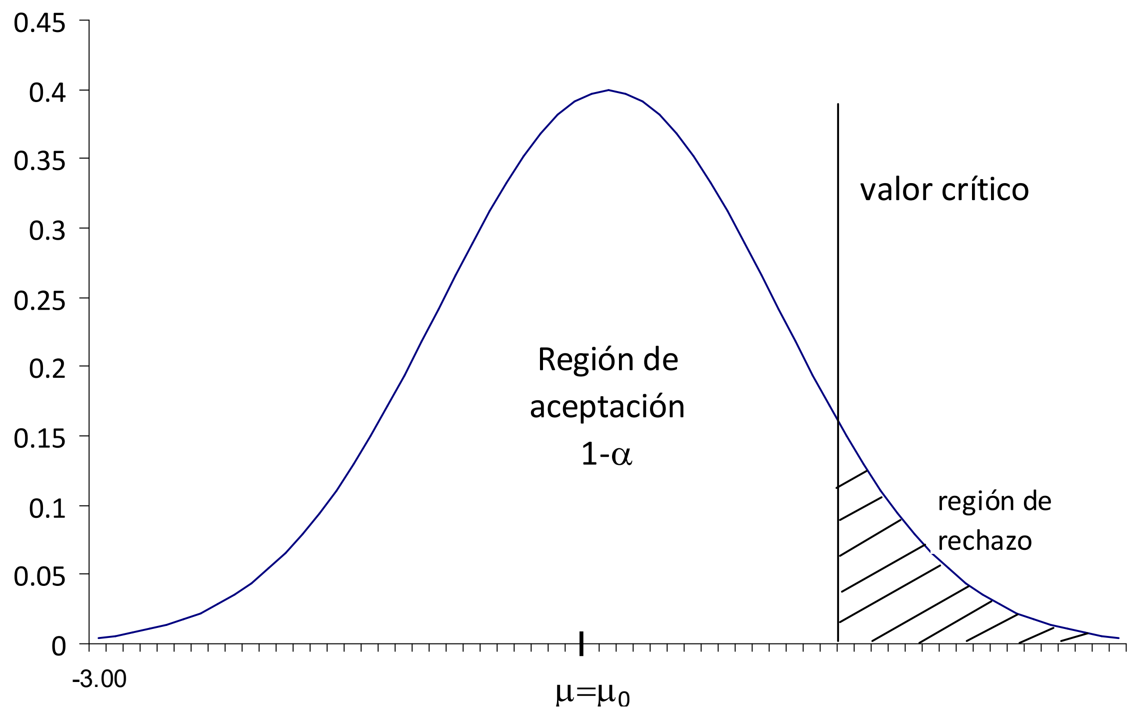
Regla de Decisión:

Se rechaza H_0 si $$Z_{obs} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} > Z_\alpha$$

Alternativa:

Se rechaza H_0 si $$valor\ p < \alpha$$

$$valor\ p = P(Z \geq Z_{obs})$$



Test de Hipótesis

- Una planta embotelladora afirma que sus botellas de 2 litros contienen un promedio mínimo de 1997 ml. Se selecciona una muestra de botellas y se medirán sus contenidos para investigar la afirmación del fabricante. Plantear H_0 y H_A

Test Unilateral Izquierdo

$H_0: \mu \geq 1997$ $\quad$ $H_A: \mu < 1997$

$H_0: \mu = 1997$ $\quad$ $H_A: \mu < 1997$

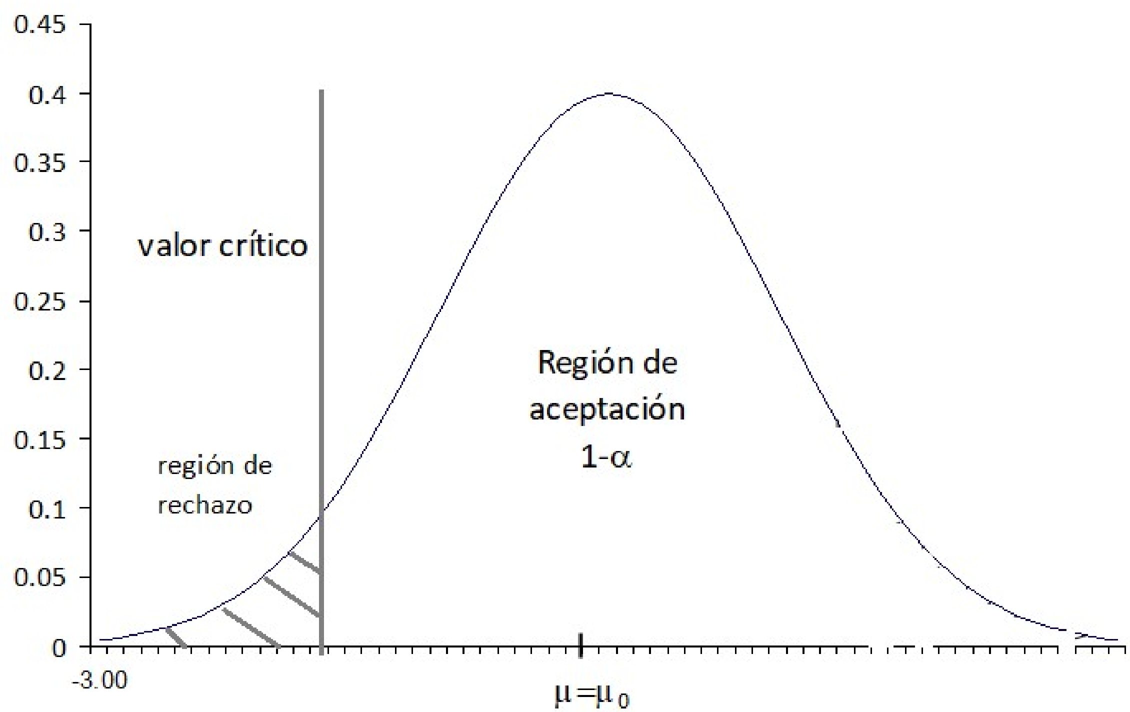
Regla de Decisión:

Se rechaza H_0 si $\quad Z_{obs} = \dfrac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} < Z_\alpha$

Alternativa:

Se rechaza H_0 si $\quad valor\ p < \alpha$

$valor\ p = P(Z \leq Z_{obs})$



Test de Hipótesis

- Se acaba de recibir un embarque y se debe someter a un control de calidad, para aceptar todo el embarque o regresarlo al proveedor porque no cumple con las especificaciones. Supongamos que se tratan de piezas que deben medir 5 cm. Si la longitud promedio de las piezas es mayor o menor a la norma de 5 cm se rechaza el embarque. Plantear H_0 y H_A

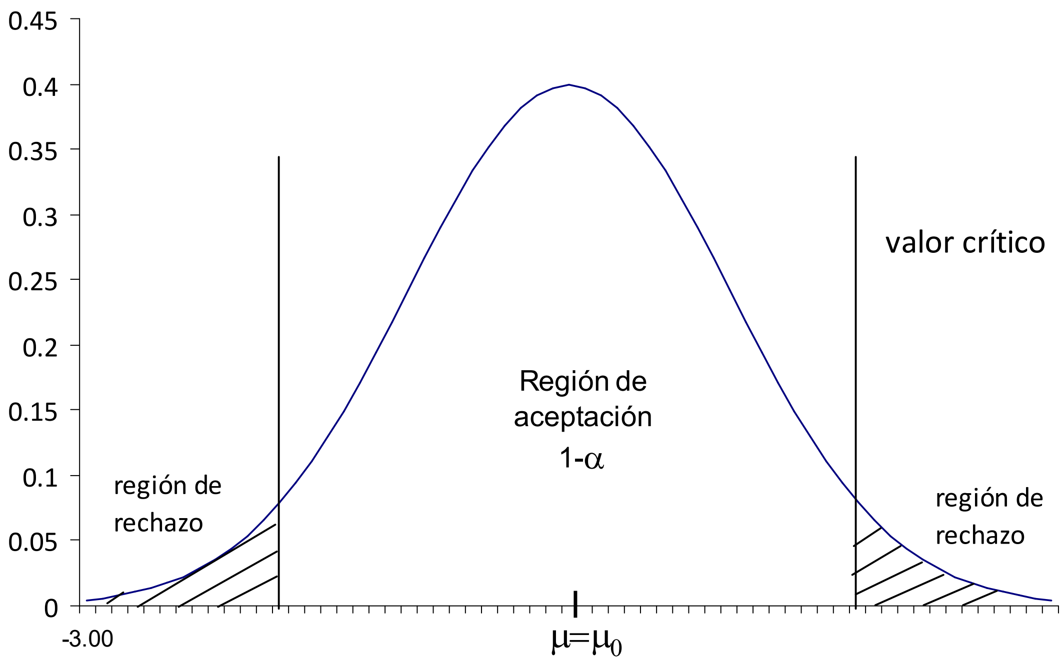
$$H_0\colon\ \mu = 5 \qquad H_A\colon\ \mu \neq 5$$

Test Bilateral

Regla de Decisión:

Se rechaza H_0 si

$$Z_{obs} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} < -Z_{\alpha/2} \quad \textbf{ó} \quad Z_{obs} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} > Z_{\alpha/2}$$



Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$$valor\ p = 2 * P(Z \geq |Z_{obs}|)$$

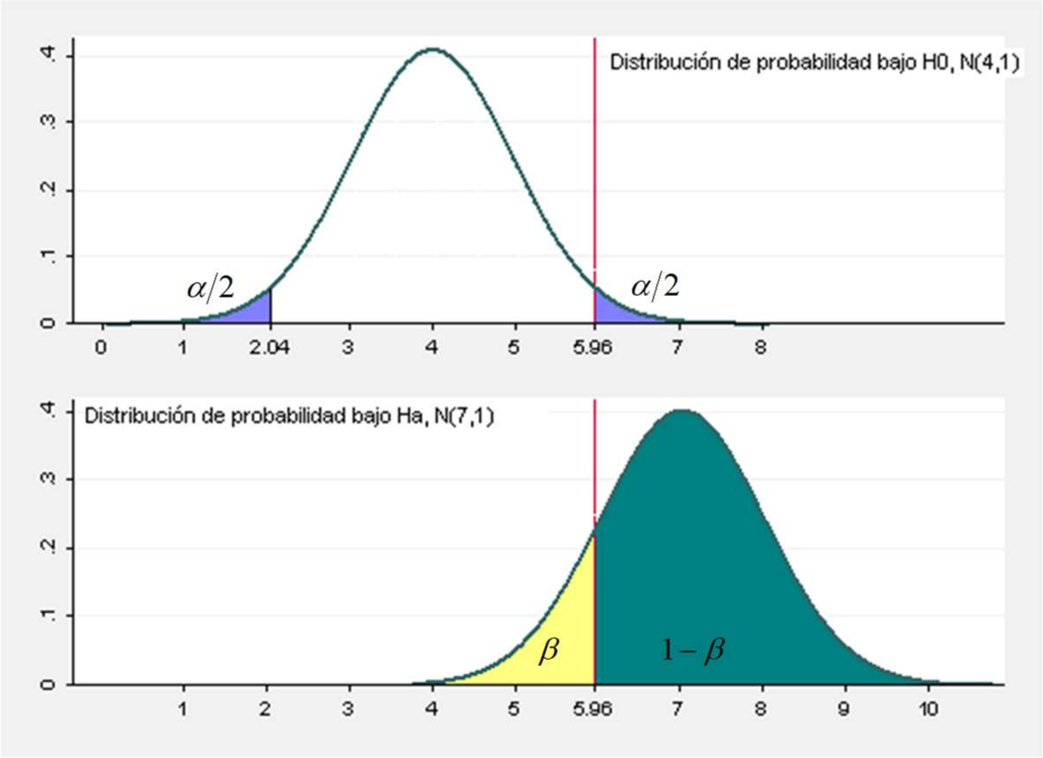
Test de Hipótesis

- Dos son los tipos de error que podemos cometer:
- P(**rechazar H_0** cuando H_0 es verdadera) = **Error de tipo I**
- P(**no rechazar H_0** cuando H_0 es falsa) = **Error de tipo II**

Test de Hipótesis		
	Desición estadística	
Verdadero estado de la Hipótesis Nula	No se rechaza H0	Se rechaza H_0
H_0 es verdadera	**Decisión correcta** 1-α= nivel de confianza (región de aceptación)	**Error de tipo I** = α = nivel de significación del test
H_0 es falsa (H_A es verdadera)	**Error de tipo II** = β	**Decisión Correcta** Probabilidad =1-β = potencia del test

Test de Hipótesis

- Dos son los tipos de error que podemos cometer:
- P(**rechazar H_0** cuando H_0 es verdadera) = **Error de tipo I**
- P(**no rechazar H_0** cuando H_0 es falsa) = **Error de tipo II**



Test de Hipótesis

- **Muestras chicas: n < 30, desvío poblacional desconocido**

Test Unilateral Derecho

H_0: $\mu \leq 14$ vs. H_A: $\mu > 14$

H_0: $\mu = 14$ vs. H_A: $\mu > 14$

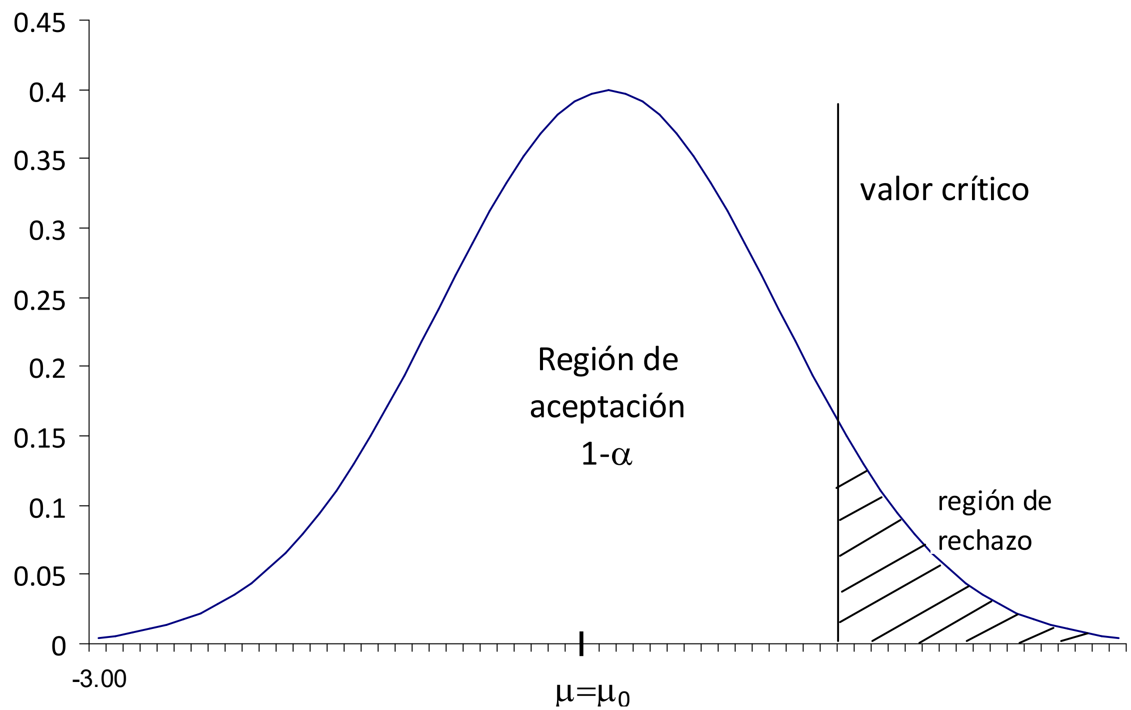
Regla de Decisión:

Se rechaza H_0 si $$t_{obs} = \frac{\bar{X} - \mu_{H_0}}{s/\sqrt{n}} > t_{n-1,\alpha}$$

Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$$valor\ p = P(t_{n-1} \geq t_{obs})$$



Test de Hipótesis

- **Muestras chicas: n < 30, desvío poblacional desconocido**

Test Unilateral Izquierdo

H_0: $\mu \geq 1997$ H_A: $\mu < 1997$

H_0: $\mu = 1997$ H_A: $\mu < 1997$

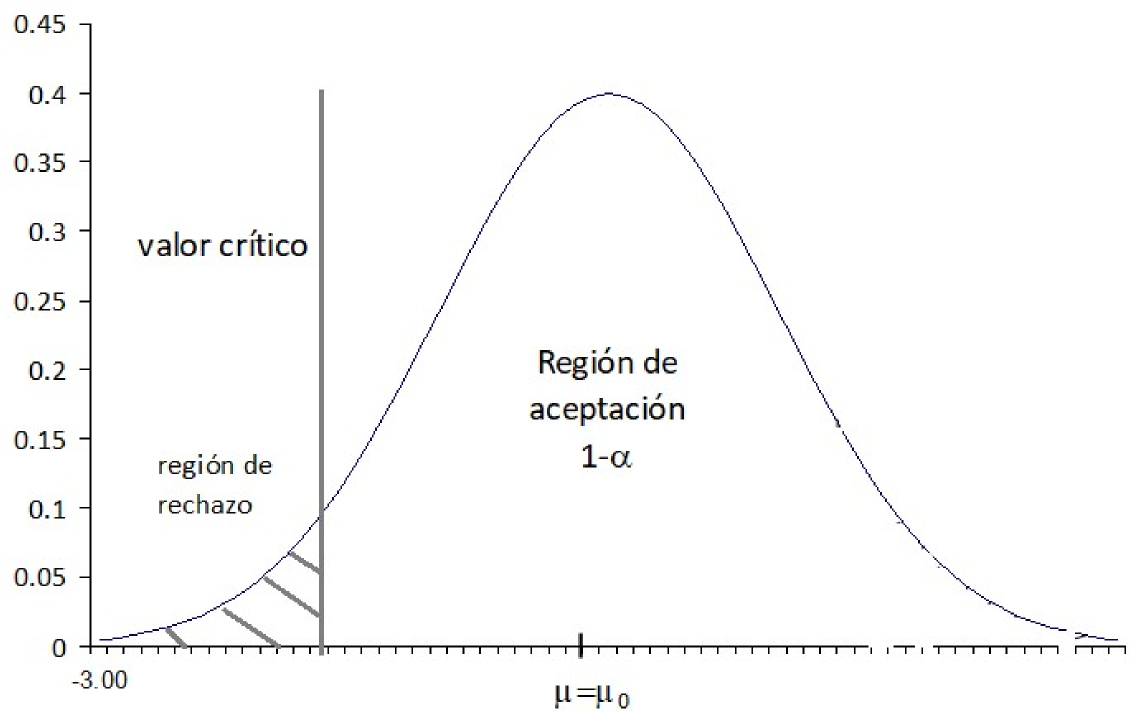
Regla de Decisión:

Se rechaza H_0 si $t_{obs} = \dfrac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} < t_{n-1,\alpha}$

Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$valor\ p = P(t_{n-1} \leq t_{obs})$



Test de Hipótesis

- **Muestras chicas: n < 30, desvío poblacional desconocido**

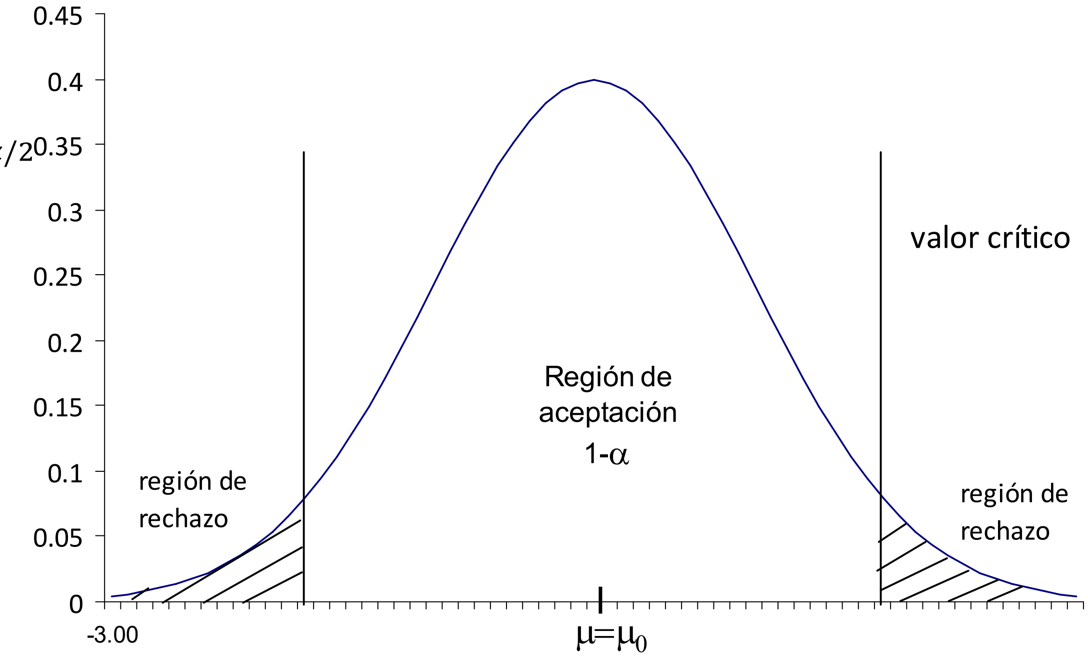
$H_0\!: \mu = 5 \qquad H_A\!: \mu \neq 5$

Test Bilateral

Regla de Decisión:

Se rechaza H_0 si

$$t_{obs} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} < -t_{n-1,\alpha/2} \quad \text{ó} \quad t_{obs} = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}} > t_{n-1,\alpha/2}$$



Alternativa:

Se rechaza H_0 si $\quad valor\ p < \alpha$

$$valor\ p = 2 * P(t_{n-1} \geq |t_{obs}|)$$

Test de Hipótesis

- **Proporción** $np_0 \geq 10$ y $n(1-p_0) \geq 10$

Test Unilateral Derecho

$H_0: p \leq 0{,}4$ vs. $H_A: p > 0{,}4$

$H_0: p = 0{,}4$ vs. $H_A: p > 0{,}4$

Regla de Decisión:

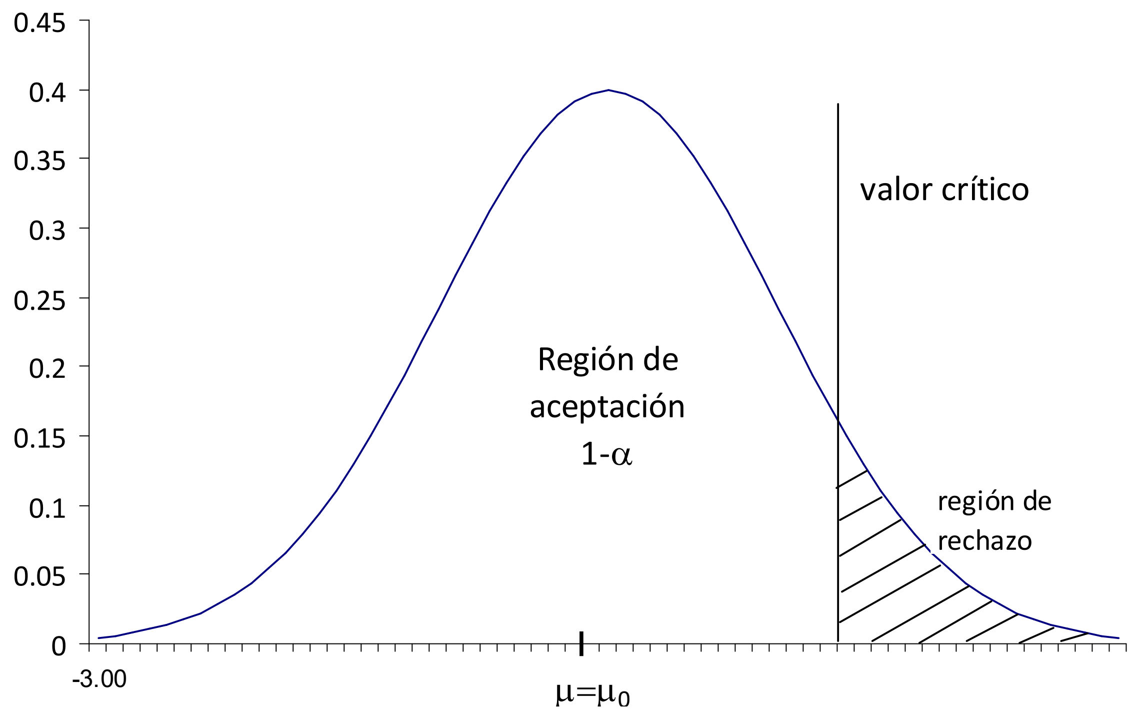
Se rechaza H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > Z_\alpha$$

Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$$valor\ p = P(Z \geq Z_{obs})$$



Test de Hipótesis

- **Proporción** $np_0 \geq 10$ y $n(1-p_0) \geq 10$

Test Unilateral Izquierdo

$H_0: p \geq 0{,}4 \qquad H_A: p < 0{,}4$

$H_0: p = 0{,}4 \qquad H_A: p < 0{,}4$

Regla de Decisión:

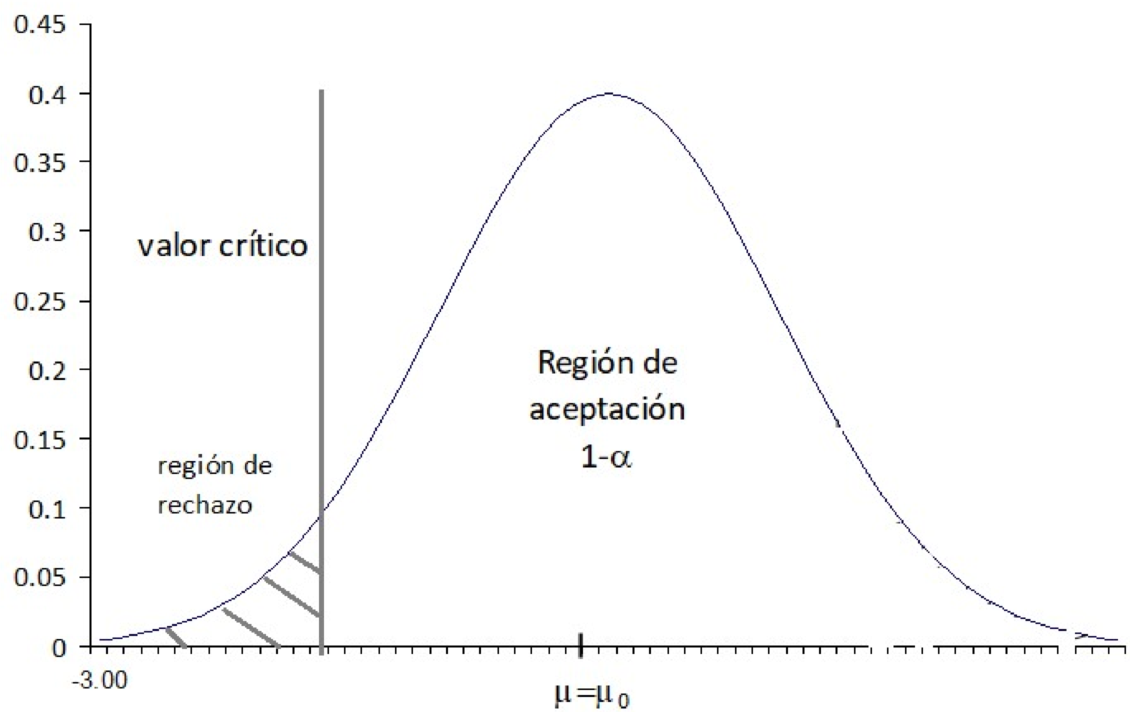
Se rechaza H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -Z_\alpha$$

Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$$valor\ p = P(Z \leq Z_{obs})$$



Test de Hipótesis

- **Proporción** $np_0 \geq 10$ y $n(1-p_0) \geq 10$

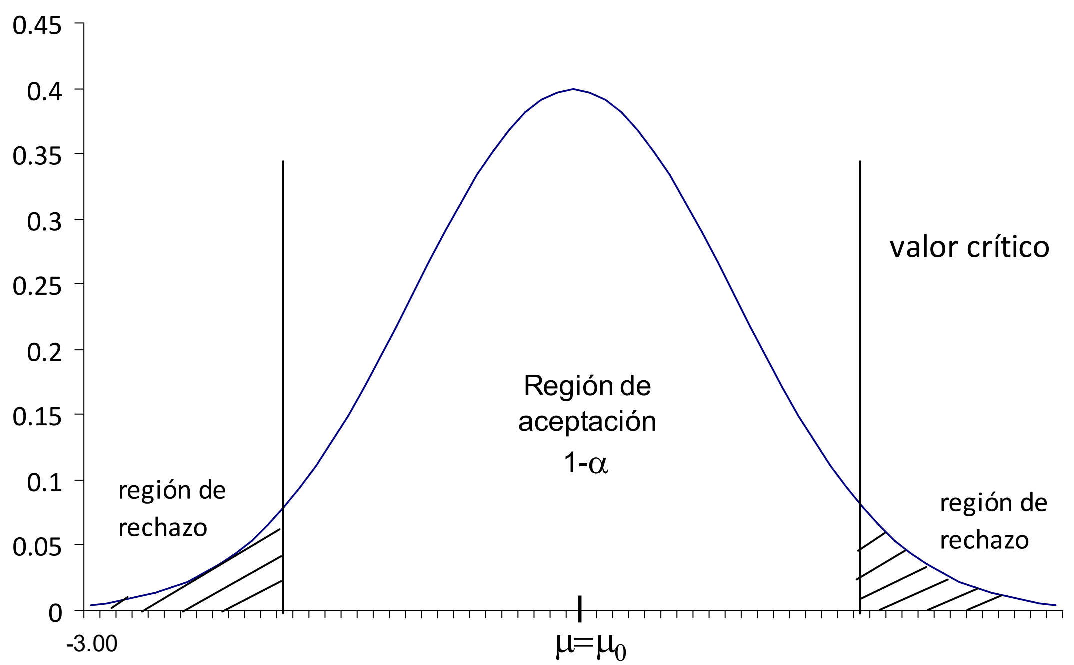
$H_0: p = 0{,}4$ $\quad$ $H_A: p \neq 0{,}4$

Test Bilateral

Regla de Decisión:

Se rechaza H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > Z_{\alpha/2} \quad \text{o} \quad Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -Z_{\alpha/2}$$



Alternativa:

Se rechaza H_0 si $valor\ p < \alpha$

$$valor\ p = 2 * P(Z \geq |Z_{obs}|)$$

Test de Hipótesis

- **Varianza**

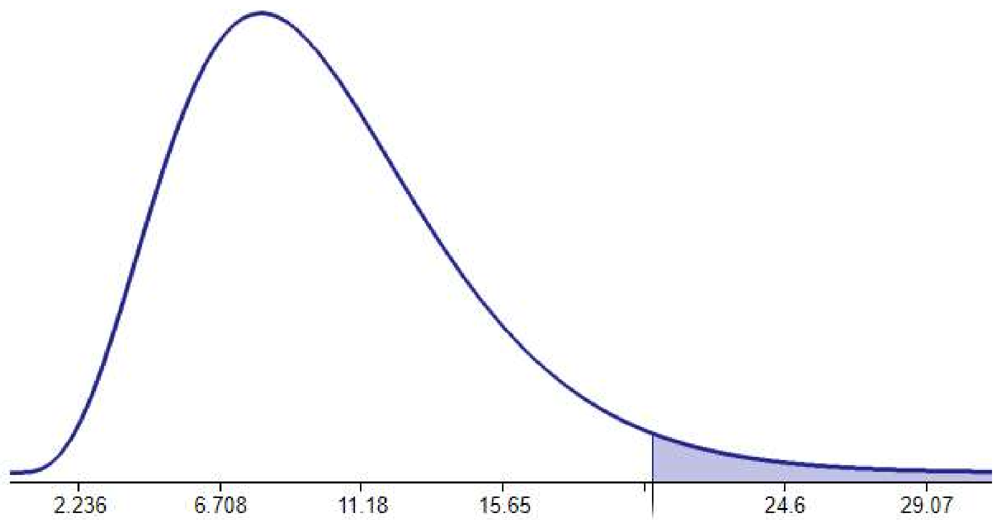
Test Unilateral Derecho

$H_0: \sigma^2 \leq 0{,}4$ vs. $H_A: \sigma^2 > 0{,}4$

$H_0: \sigma^2 = 0{,}4$ vs. $H_A: \sigma^2 > 0{,}4$

Regla de Decisión:

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



Test de Hipótesis

- **Varianza**

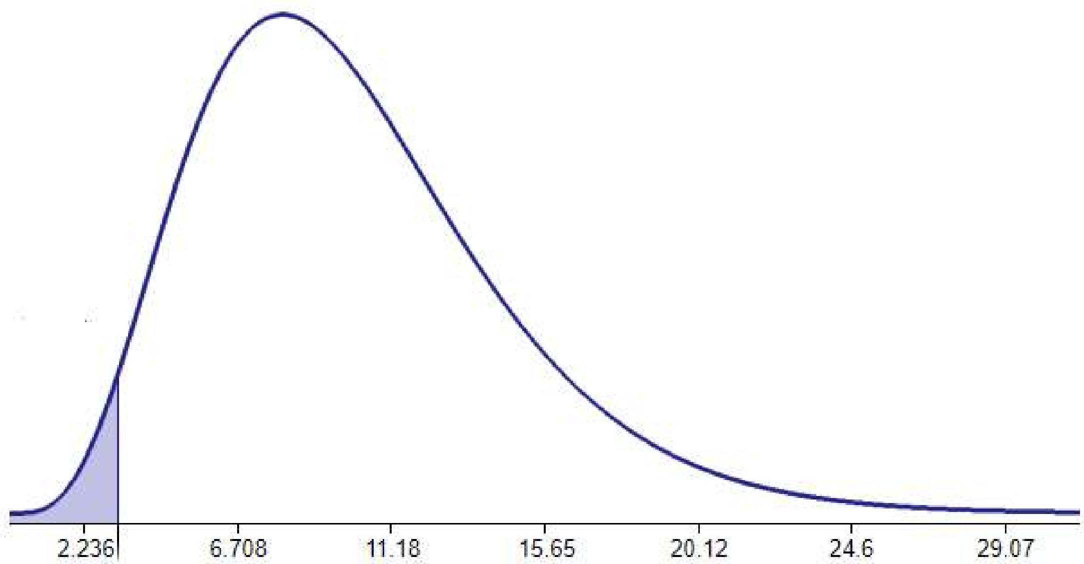
Test Unilateral Izquierdo

$H_0: \sigma^2 \geq 0{,}4$ $\quad H_A: \sigma^2 < 0{,}4$

$H_0: \sigma^2 = 0{,}4$ $\quad H_A: \sigma^2 < 0{,}4$

Regla de Decisión:

Se rechaza H_0 si $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha}$



Test de Hipótesis

- **Varianza**

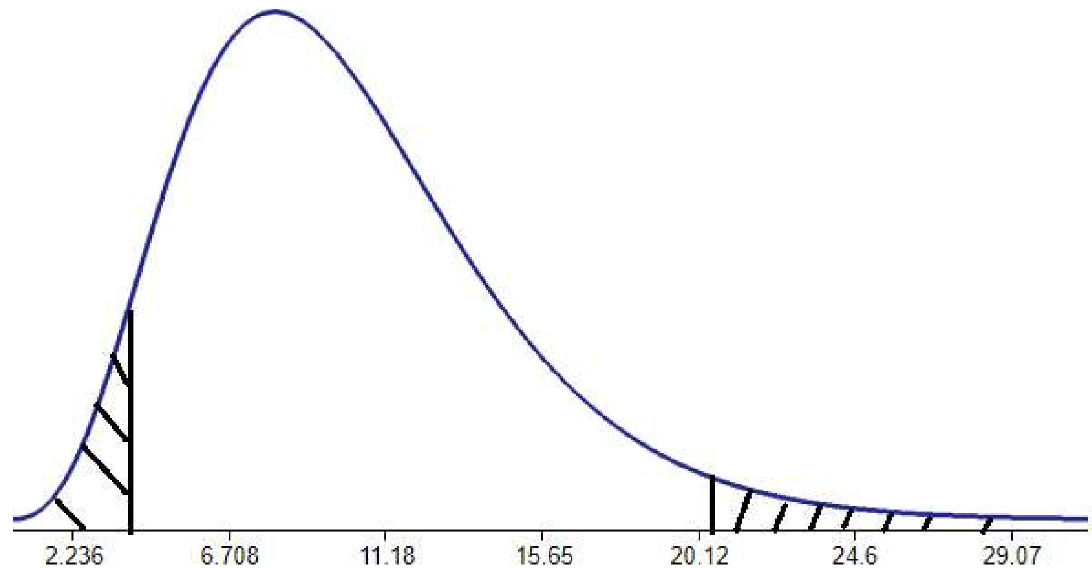
$H_0: \sigma^2 = 0{,}4 \qquad H_A: \sigma^2 \neq 0{,}4$

Test Bilateral

Regla de Decisión:

Se rechaza H_0 si

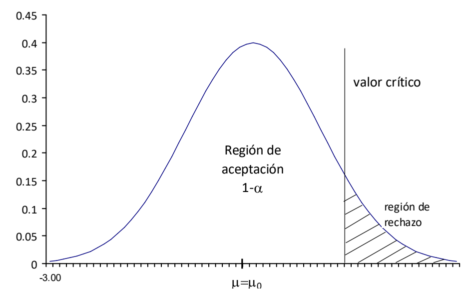
$$\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha/2} \text{ o } \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha/2}$$



Test de Hipótesis diferencia de medias

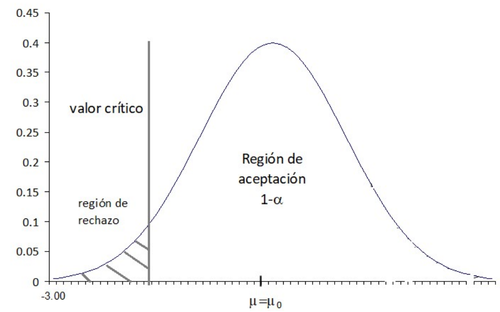
- Muestras independientes (y grandes n>100)
  - $H_0: \mu_x - \mu_y \leq 0 \;\; vs. \;\; H_1: \mu_x - \mu_y > 0$

Rechazo H_0 si $\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > Z_\alpha$



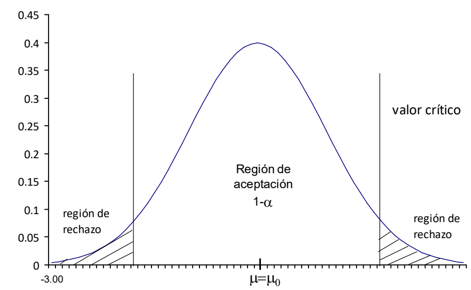
  - $H_0: \mu_x - \mu_y \geq 0 \;\; vs. \;\; H_1: \mu_x - \mu_y < 0$

Rechazo H_0 si $\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} < -Z_\alpha$



  - $H_0: \mu_x - \mu_y = 0 \;\; vs. \;\; H_1: \mu_x - \mu_y \neq 0$

Rechazo H_0 si $\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} < -Z_{\alpha/2}$ o $\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > Z_{\alpha/2}$



Test de Hipótesis diferencia de medias

- Muestras pequeñas (n<100), varianzas desconocidas que se supone iguales

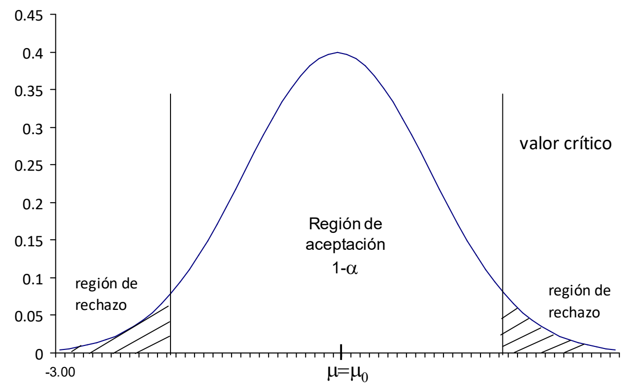
$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{(n_x + n_y - 2)}$$

- $H_0: \mu_x - \mu_y = 0$ vs. $H_1: \mu_x - \mu_y \neq 0$

Rechazo H_0 si

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} < -t_{n_x + n_y - 2, \alpha/2} \quad \text{ó}$$

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} > t_{n_x + n_y - 2, \alpha/2}$$



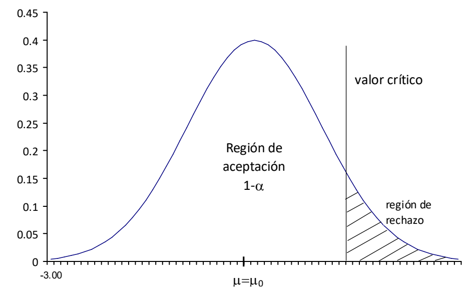
Test de Hipótesis diferencia de medias

- Muestras pequeñas (n<100), varianzas desconocidas que se supone iguales

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

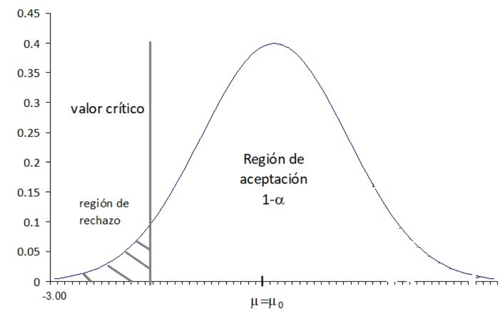
- $H_0: \mu_x - \mu_y \leq 0$ vs. $H_1: \mu_x - \mu_y > 0$

Rechazo H_0 si $t = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} > t_{n_x+n_y-2,\alpha}$



- $H_0: \mu_x - \mu_y \geq 0$ vs. $H_1: \mu_x - \mu_y < 0$

Rechazo H_0 si $t = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} < -t_{n_x+n_y-2,\alpha}$



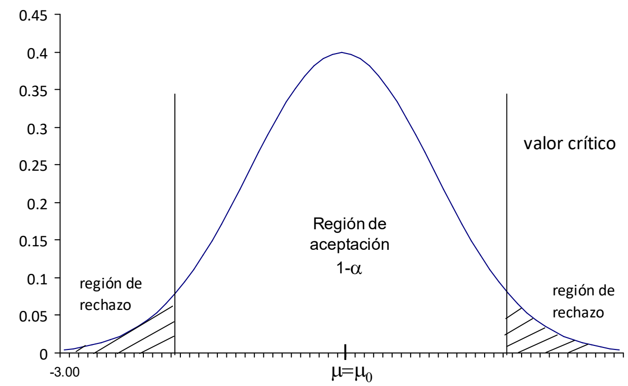
Test de Hipótesis diferencia de medias

- Datos Apareados

$d_i = x_i - y_i$ $\quad\quad \bar{d} = \frac{\sum_i d_i}{n}$ y $S_{\bar{d}} = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$

- $H_0 : \mu_d = 0$ vs. $H_1 : \mu_d \neq 0$

Rechazo H_0 si $t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}} / \sqrt{n}} > t_{n-1, \alpha/2}$ o si



Rechazo H_0 si $t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}} / \sqrt{n}} < -t_{n-1, \alpha/2}$

donde la $P(t_{n-1} > t_{n-1,\alpha}) = \alpha$

t de student con *n-1* grados de libertad

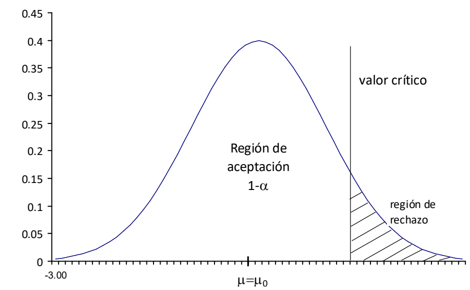
Test de Hipótesis diferencia de medias

- Datos Apareados

$d_i = x_i - y_i$ $\bar{d} = \frac{\sum_i d_i}{n}$ y $S_{\bar{d}} = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$

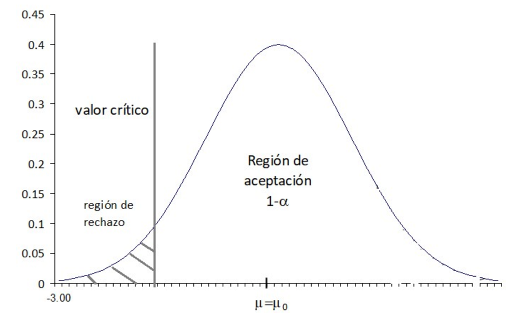
- H_0: $\mu_d = 0$ o $\mu_d \leq 0$ vs. $H_1 : \mu_d > 0$

Rechazo H_0 si $t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}}/\sqrt{n}} > t_{n-1,\alpha}$



- H_0: $\mu_d = 0$ o $\mu_d \geq 0$ vs. $H_1 : \mu_d < 0$

Rechazo H_0 si $t = \frac{\bar{d} - \mu_{\bar{d}}}{S_{\bar{d}}/\sqrt{n}} < -t_{n-1,\alpha}$



Test de Hipótesis diferencia de Proporciones

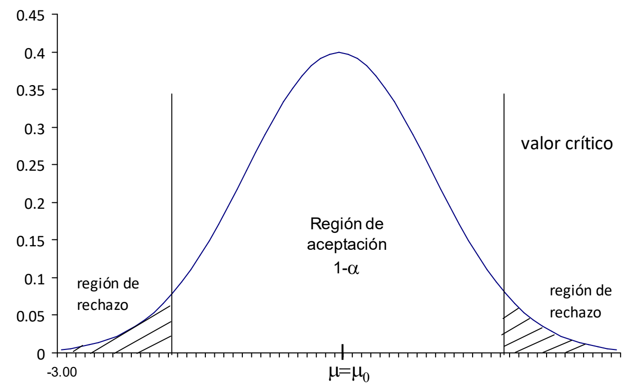
$$\hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y}$$

- $H_0 : p_x - p_y = 0$ vs. $H_1 : p_x - p_y \neq 0$

Rechazo H_0 si $\frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}} < -z_{\alpha/2}$

- o

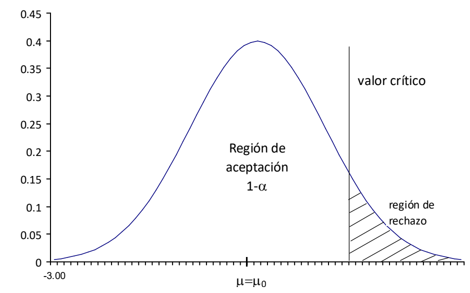
Rechazo H_0 si $\frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}} > z_{\alpha/2}$



Test de Hipótesis diferencia de Proporciones

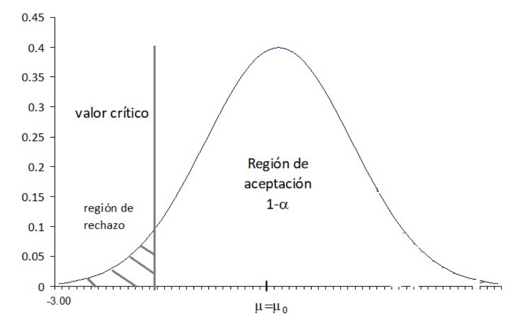
$$\hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y}$$

- $H_0: p_x - p_y = 0$ o $p_x - p_y \leq 0$ vs. $H_1: p_x - p_y > 0$



Rechazo H_0 si $\frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}} > z_\alpha$

- $H_0: p_x - p_y = 0$ o $p_x - p_y \geq 0$ vs. $H_1: p_x - p_y < 0$



Rechazo H_0 si $\frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}} < -z_\alpha$

Regresión Lineal

Variable dependiente	Variable explicativa
⇕	⇕
Variable explicada	Variable independiente
⇕	⇕
Predicha	Predictora
⇕	⇕
Regresada	**Regresora**
⇕	⇕
Respuesta	Estímulo
⇕	⇕
Endógena	Exógena
⇕	⇕
Resultado	Covariante
⇕	⇕
Variable controlada	Variable de control

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$e_i = Y_i - \hat{Y}_i$$

Mínimos Cuadrados Ordinarios (MCO)

Minimizar: $\sum_{i=1}^{n} e_i^2$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_2 = \frac{S_{XY}}{S_X^2}$$

Regresión Lineal

Dependent Variable: GASTOS_M3
Method: Least Squares
Date: 10/08/19 Time: 19:20
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1736.649	484.1497	3.587007	0.0005
INGRESO_M3	0.501499	0.057249	8.759908	0.0000

R-squared	0.439154	Mean dependent var	5013.232
Adjusted R-squared	0.433431	S.D. dependent var	4083.877
S.E. of regression	3073.966	Akaike info criterion	18.91912
Sum squared resid	9.26E+08	Schwarz criterion	18.97123
Log likelihood	-943.9561	Hannan-Quinn criter.	18.94021
F-statistic	76.73599	Durbin-Watson stat	2.228648
Prob(F-statistic)	0.000000		

$$Gastos_i = 1736{,}649 + 0{,}501499 * Ingreso_i$$

Intervalos de confianza:

$$\beta_1: 1736{,}649 \pm t_{n-2,\alpha/2} * 484{,}1497$$

$$\beta_2: 0{,}501499 \pm t_{n-2,\alpha/2} * 0{,}057249$$

Regresión Lineal

```
Call:
lm(formula = rendimientos_exc$GGAL ~ rendimientos_exc$MERVAL)

Residuals:
     Min       1Q   Median       3Q      Max
-21.1557  -5.1457  -0.7833   6.1878  31.1741

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              0.56590    0.60653   0.933    0.352
rendimientos_exc$MERVAL  1.17468    0.05549  21.170   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.383 on 191 degrees of freedom
Multiple R-squared:  0.7012,	Adjusted R-squared:  0.6996
F-statistic: 448.2 on 1 and 191 DF,  p-value: < 2.2e-16
```

$$GGAL_i = 0{,}5659 + 1{,}17468 * MERVAL_i$$

Intervalos de confianza:

$$\beta_1: 0{,}5659 \pm t_{n-2,\alpha/2} * 0{,}60653$$

$$\beta_2: 0{,}05549 \pm t_{n-2,\alpha/2} * 0{,}05549$$