INFERENCIA, ESTIMACIÓN Y PRUEBAS ESTADÍSTICAS

SÉPTIMA Y OCTAVA CLASE

Analytics AoZ

1. Inferencia Estadística – Estimación Puntual.
2. Estimación Interválica.
3. Pruebas de Hipótesis y tipos de errores.
4. Pruebas Chi-cuadrado
5. Test en R
   - ❑ Ejemplo de t-student una muestra y muestras parejas.
6. Diseño experimental – ANOVA
7. Pruebas no Paramétricas
   - ❑ Wilcoxon
   - ❑ U de Mann-Whitney
   - ❑ K de Kruskall-Wallis
8. Casos aplicados de Geología de los puntos 1 al 7 en Rstudio.

INTRODUCCIÓN

Analytics AoZ

- Uno de los objetivos de la Estadística es el de realizar inferencia acerca de las característica de una población.
- Estas inferencias se basan en la información obtenida de una **muestra aleatoria**, de la población en estudio, la cual debe ser representativa de la población de interés.

CONCEPTOS PREVIOS

- Para abordar el tema de Estimación Puntual e Intervalos de Confianza, se debe tener claro los siguientes conceptos:

  - Población y Muestra
  - Parámetro y Estadígrafo o Estimador
  - Inferencia Estadística

Población

Es un conjunto total, finito o infinito, de personas, objetos o cosas, que tienen al menos una característica en común, cuyo estudio nos interesa o acerca de los cuales se desea información. Tenemos aquí, la noción de población como un conjunto universo.

Ejemplos:

- Todas las rocas mayores de 300000 años de antiguedad de la Cuenca San Gabán.
- Los Alumnos de la UNI, que viven en el cono norte de la ciudad de Lima.
- Los animales que viven en Ancash.

Muestra

Subconjunto finito de unidades de análisis (de tamaño "n") seleccionados de la población en estudio. La muestra debe ser **representativa**, es decir debe tener, en lo posible, caracteristicas similares a las de su poblacion.

Ejemplos:

- 50 rocas seleccionadas, de entre las personas que residen en la Ciudad de Arequipa.
- Se seeleccionan 35 colegios de nivel primaria, de entre los colegios ubicados en el cono norte de Lima.
- 80 familias seleccionadas de entre todas las familias residents en el distrito de Jesús María.

Parámetro (θ)

Es una medida de resumen de alguna característica de interés (Variable), que representa a la población y cuyo valor numérico se calcula en base al estudio de toda la población (Censo).

Como el realizar un censo generalmente no es factible por el tiempo que requiere y por el costo que implica, estos valores son desconocidos y por lo tanto deben de ser ESTIMADOS.

Los principales Parámetros son:

Media o Promedio Poblacional : μ
Proporción Poblacional : π
Varianza Poblacional : σ^2

También estaremos interesados en los parámetros:

Diferencia de Medias : $\mu_1 - \mu_2$
Diferencia de Proporciones : $\pi_1 - \pi_2$
Cociente de Varianzas : σ_1^2 / σ_2^2

Analytics AoZ

POBLACIÓN

μ

σ^2

σ

π

Parámetros

Muestreo

Inferencia Estadística

MUESTRA

$X_1, X_2, \ldots, X_n$

$\overline{X}$

S^2

S

p

Estimadores

Estimador ($\hat{\Theta}$)

Es una medida de resumen que representa los datos de una muestra aleatoria de tamaño n X_1, $X_2,\ldots,X_n$; tomados de una determinada población

Es decir:

$$\text{Estimador} = \hat{\Theta} = f(X_1, X_2, \ldots, X_n)$$

Esto nos indica que el estimador solo depende de los valores muestrales.

Los estimadores o estadígrafos correspondientes a los parámetros definidos anteriormente son:

Media Muestral:

- Se denotado por $\overline{x}$
- Es el estimador puntual de la media poblacional μ
- Se calcula de la siguiente manera:

  si tenemos una muestra de tamaño "n"

  $X_1, X_2, \ldots, X_n$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Proporción muestral:

- Se denotado por "p"
- Es el estimador puntual de la proporción poblacional π
- Se calcula de la siguiente manera:

  si tenemos una muestra de tamaño "n"

  $X_1, X_2, \ldots, X_n$

$$p = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \textit{donde: } \mathrm{x_i} = \begin{cases} 1 & \text{exito} \\ 0 & \text{fracaso} \end{cases}$$

Analytics AoZ

Varianza muestral:

- Denotado por S^2
- Es un estimador de la Varianza poblacional σ^2
- Se calcula de la siguiente forma:

  si tenemos una muestra aleatoria de tamaño "n"

  $X_1, X_2, \ldots, X_n$

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

o también

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}{n-1}$$

ERROR DE ESTIMACIÓN

Analytics AoZ

Definimos como error de estimación a la diferencia, términos absolutos, entre el Estimador y el Parámetro.

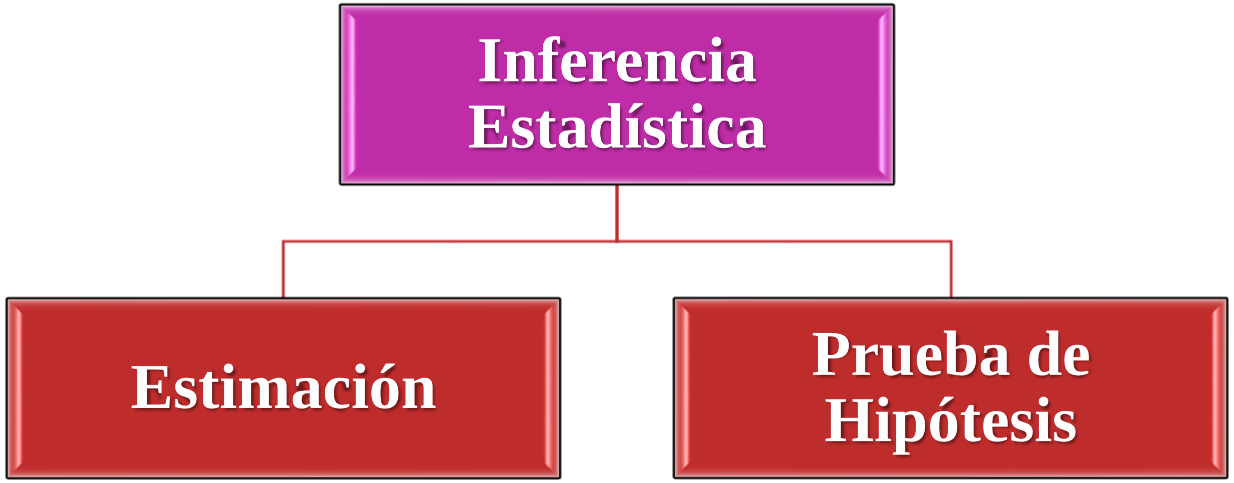
Este error se debe a que una muestra no da la información completa con respecto de la población.

Este error puede ser medido y controlado, de cierta forma, utilizando técnicas estadísticas adecuadas.

$$\mathbf{E} = \left| \hat{\theta} - \theta \right|$$

1. INFERENCIA ESTADÍSTICA-ESTIMACIÓN

Es una parte de la estadística que permite hacer afirmaciones sobre los parámetros de la población bajo estudio con base a las observaciones de una muestra. La inferencia provee los procedimientos inductivos y las mediciones de incertidumbre para efectuar estas afirmaciones.
Los pilares de los procedimientos de inferencia estadística son: La Estimación y la Prueba de Hipótesis.



ESTIMACIÓN ESTADÍSTICA

Analytics AoZ

La estimación estadística consiste en utilizar datos muestrales para determinar los valores de los parámetros desconocidos de una población. Este proceso puede adoptar la forma de un sólo punto o de un intervalo.

Esto quiere decir que se tiene:

* Estimación puntual
* Estimación por intervalos

Estimación puntual

Si se tiene una muestra de tamaño n de la población X; se dice que una estadística T de la muestra es un estimador puntual del parámetro θ de la población, si es una función de las observaciones de la muestra.

PROPIEDADES DE UN BUEN ESTIMADOR PUNTUAL

Una estadística muestral que cumple las siguientes propiedades, se puede considerar como un buen estimador puntual:

Insesgamiento	Consistencia	Eficiencia	Suficiencia
Sí su valor esperado es igual al parámetro.	Sí tienen una distancia mínima respecto al valor del parámetro.	Sí y sólo sí tiene una varianza mínima.	Si es capaz de sustraer de la muestra toda la información que ésta contenga acerca del parámetro.

PROPIEDADES DE UN BUEN ESTIMADOR

- Si utilizamos el valor de un estimador o estadígrafo para calcular un parámetro de una población, este valor es una estimación puntual del parámetro. Estas estimaciones reciben el nombre de estimación puntual porque son números únicos, o puntos situados en el eje real.

  El estadígrafo cuyo valor se utiliza para la estimación puntual del parámetro se llama Estimador, y el valor de este estimador será el valor estimado del parámetro.

Así, por ejemplo:

$\bar{x}$: Es un estimador puntual de μ y el valor numérico de $\bar{x}$ será el valor estimado de μ

p: Es un estimador puntual de la proporción poblacional π, y el valor numérico de p será el valor estimado de π

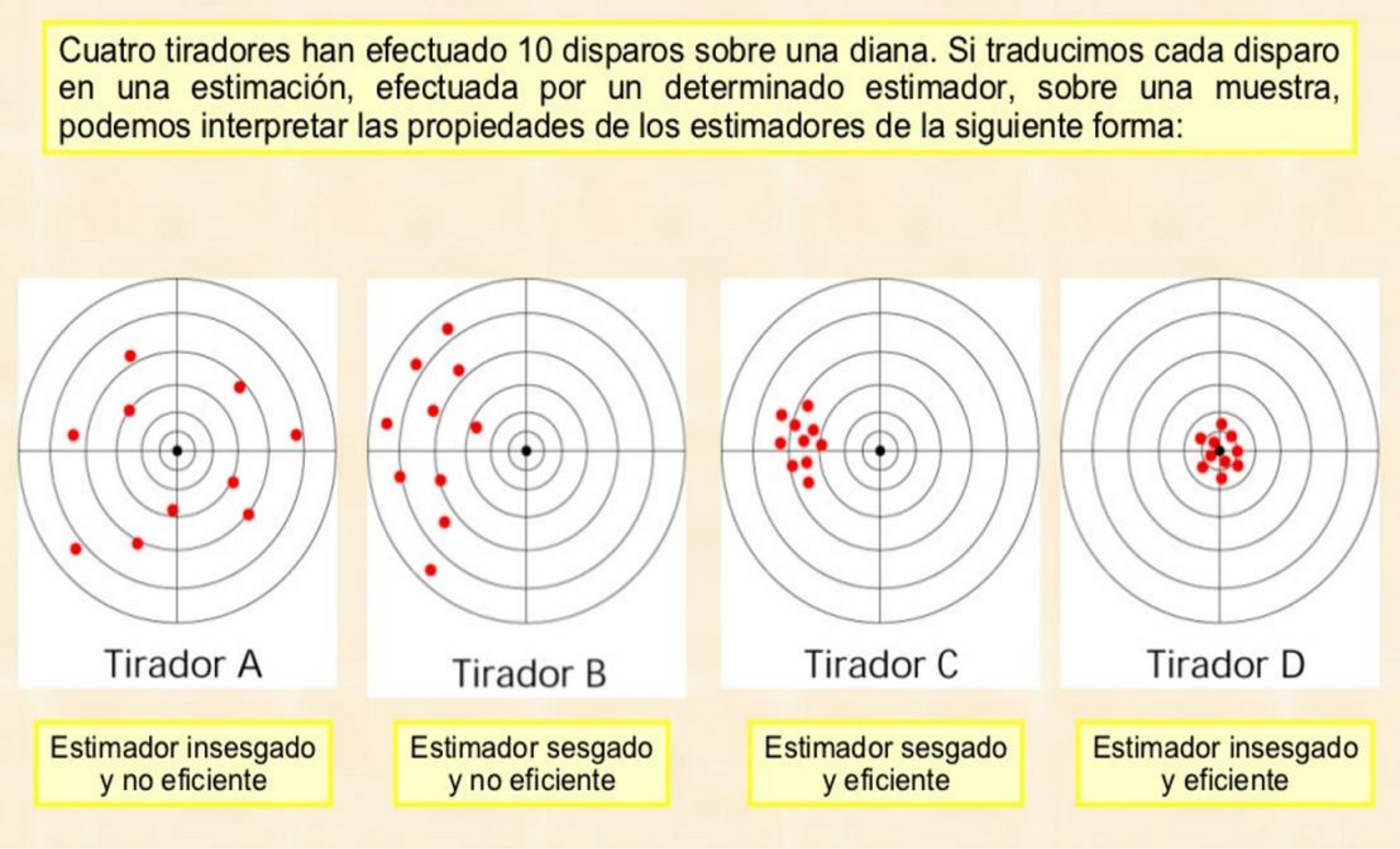
S^2: Es un estimador puntual de σ^2 y el valor numérico de S^2 será el valor estimado de σ^2

En general diremos que $\hat{\theta}$ será el estimador puntual de θ, y el valor de $\hat{\theta}$ será valor estimado de θ

- Como los estimadores son variables aleatorias Estamos interesados en estudiar ciertas propiedades de los estimadores para decidir cual de los estimadores es el más apropiado, para un determinado parámetro, en una situación dada.
- Las propiedades específicas, de los estimadores, que estudiaremos a continuación son:
  - Insesgabilidad
  - Consistencia
  - Eficiente
  - Suficiencia

UN BUEN ESTIMADOR PUNTUAL

Analytics AoZ



DEFINICIÓN DE UN ESTIMADOR INSESGADO

Se dice que un estadígrafo $\hat{\theta}$ es un estimador insesgado del parámetro $\boldsymbol{\theta}$, si y solo sí:

$$E(\hat{\theta}) = \theta$$

Ejemplo 1.-

La media de una muestra $\bar{x}$ es un estimador insesgado de la media de la población μ.

$$E(\bar{x}) = E\left(\frac{\sum x}{n}\right) = \frac{\sum E(x)}{n} = \frac{nE(x)}{n} = \frac{n\mu}{n} = \mu$$

Ejemplo 2.-

La varianza de una muestra s^2 es un estimador insesgado de la varianza de la población σ^2.

$$E(s^2) = E\left(\frac{\sigma^2(n-1)}{\sigma^2(n-1)} s^2\right) = \frac{\sigma^2}{n-1} E\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{\sigma^2}{n-1} E(\chi^2_{n-1}) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

ESTIMADOR INSESGADO

Se denomina sesgo de un estimador a la diferencia entre la esperanza del estimador y el verdadero parámetro a estimar. Es deseable que un estimador sea insesgado o centrado, esto es, que el sesgo sea nulo para que la esperanza del estimador sea igual al parámetro que se desea estimar.

Propiedades de la esperanza y varianza

a) $E[aX_1 \pm bX_2] = E[aX_1] \pm E[bX_2] = aE[X_1] \pm bE[X_2]$

b) $Var[aX_1 \pm bX_2] = Var[aX_1] + Var[bX_2] = a^2\,Var[X_1] + b^2\,Var[X_2]$

DEFINICIÓN DE UN ESTIMADOR CONSISTENTE

Analytics AoZ

Se dice que un estadígrafo es consistente sí tienen una distancia mínima respecto al valor del parámetro. Esto significa que debe cumplir con las siguientes condiciones:

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta \text{ y } \lim_{n\to\infty} V(\hat{\theta}) = 0$$

Ejemplo .-

La varianza de una muestra s^2 es un estimador consistente de la varianza de la población σ^2

$$\text{Ya probamos que:} \quad E(s^2) = \sigma^2 \qquad \lim_{n\to\infty} E(s^2) = \lim_{n\to\infty} \sigma^2 = \sigma^2$$

$$V(s^2) = V\left(\frac{\sigma^2(n-1)}{\sigma^2(n-1)} s^2\right) = \frac{\sigma^4}{(n-1)^2} V\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{\sigma^4}{(n-1)^2} V(\chi^2_{n-1}) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$$

$$\lim_{n\to\infty} V(s^2) = \lim_{n\to\infty} \frac{2\sigma^4}{n-1} = 0$$

DEFINICIÓN DE UN ESTIMADOR EFICIENTE

Analytics AoZ

Se dice que un estadígrafo es eficiente sí y sólo sí tiene una varianza mínima. Esto es, sean $\hat{\theta}_1$ y $\hat{\theta}_2$ estimadores del parámetro θ, se dice que $\hat{\theta}_1$ es más eficiente $\hat{\theta}_2$ si y solo si:

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

Ejemplo .-

Sea X_1, X_2, X_3, X_4, X_5 una muestra aleatoria extraída de una población N(μ, σ²) y sean las estadísticas para estimar μ. $T_1 = \bar{x}$ y $T_2 = \dfrac{X_1 + X_2 + 2X_3 + X_4 + X_5}{6}$

Primero, hay que comprobar si son estimadores insesgados

$$E(T_1) = E(\bar{x}) = \mu$$

$$E(T_2) = E\left(\frac{X_1 + X_2 + 2X_3 + X_4 + X_5}{6}\right) = \frac{1}{6}(\mu + \mu + 2\mu + \mu + \mu) = \mu$$

DEFINICIÓN DE UN ESTIMADOR EFICIENTE

Analytics AoZ

Segundo, hay que obtener la varianza de cada estimador, para conocer cuál de ellos la varianza mínima:

$$V(T_1) = V(\bar{x}) = V\left(\frac{\sum X}{n}\right) = \frac{\sum V(X)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{5} = 0,2\sigma^2$$

$$V(T_2) = V\left(\frac{X_1 + X_2 + 2X_3 + X_4 + X_5}{6}\right) = \frac{1}{36}(\sigma^2 + \sigma^2 + 4\sigma^2 + \sigma^2 + \sigma^2) = 0,222\sigma^2$$

Finalmente, se tiene que $V(T_1) < V(T_2)$, entonces se puede afirmar que el estimador T_1 es más eficiente que el estimador T_2

DEFINICIÓN DE UN ESTIMADOR SUFICIENTE

Analytics AoZ

Se dice que un estadístico es estimador suficiente sí y sólo si es capaz de sustraer de la muestra toda la información que ésta contenga acerca del parámetro.

Para hallar un estimador suficiente se utiliza el criterio de factorización de Newman-Fisher, el cual está dado por el siguiente teorema.

Teorema.-

El estadístico $\hat{\theta}$ será un estimador suficiente del parámetro $\boldsymbol{\theta}$ si y solo si la densidad conjunta la muestra aleatoria se puede factorizar de la siguiente manera:

$$\prod_{i=1}^{n} f(x_i, \theta) = g(\hat{\theta}, \theta)\, h(x_1, x_2, \ldots, x_n)$$

EJEMPLO DE UN ESTIMADOR SUFICIENTE

Analytics AoZ

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria extraída de una población Poisson cuyo parámetro es λ. Si define a $T = \sum X_i$ ¿se podrá afirmar que es un estimador suficiente para λ?

Por el Teorema de la Factorización, en este caso hallaremos primero la función de densidad conjunta y trataremos de descomponerla factorizándola en dos factores:

$$f(x; \theta) = g(\hat{\theta}; \theta)\, h(x)$$

Uno de los cuales es h(x), que debe ser independiente del parámetro en cuestión. Si es así, diremos que la estadística que forme parte de $g(\hat{\theta}; \theta)$ constituirá una estadística suficiente de θ.

EJEMPLO DE UN ESTIMADOR SUFICIENTE

Analytics AoZ

$$P(X_1,\cdots,X_n;\lambda) = \frac{\lambda^{X_1}e^{-\lambda}}{X_1!} * \frac{\lambda^{X_2}e^{-\lambda}}{X_2!} * \cdots * \frac{\lambda^{X_n}e^{-\lambda}}{X_n!} = \prod_{i=1}^{n}\frac{\lambda^{X_i}e^{-\lambda}}{X_i!}$$

$$= \frac{\lambda^{\sum_{i=1}^{n}X_i}e^{-n\lambda}}{\prod_{i=1}^{n}X_i!} = \lambda^{\sum_{i=1}^{n}X_i}e^{-n\lambda}\frac{1}{\prod_{i=1}^{n}X_i!}$$

$$f(X_i;\lambda) = g\left(\hat{\lambda};\lambda\right)h(X_1,\cdots,X_n)$$

$$g\left(\hat{\lambda};\lambda\right) = \lambda^{\sum_{i=1}^{n}X_i}e^{-n\lambda} \qquad h(X_1,\cdots,X_n) = \frac{1}{\prod_{i=1}^{n}X_i!}$$

PRINCIPALES ESTIMADORES PUNTUALES

Analytics AoZ

La media muestral es un estimador puntual de la media de la población.

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

La varianza muestral es un estimador puntual de la varianza de la población.

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

La proporción de la muestra es un estimador puntual de la proporción de la población.

$$\hat{\pi} = p = \frac{k}{n}$$

2. ESTIMACIÓN INTERVÁLICA

- Un método de estimación de Parámetros que se utiliza con frecuencia es la estimación mediante Intervalos de Confianza y cuya ventaja sobre la estimación puntual es que en este caso es posible determinar el error de estimación así como el nivel de confianza con el que se dan los resultados.

Definición:

Un intervalos de confianza es un rango de valores que se construye a partir de datos muestrales de modo que el parámetro, que se pretende estimar, está contenido dentro de dicho rango con una probabilidad especificada. A la probabilidad especificada se le conoce como Nivel de Confianza y se le denota con la letra griega γ

INTERVALOS DE CONFIANZA

Analytics AoZ

Es decir, dado un Parámetro de interés (que puede ser la media μ, la Proporción π , la varianza σ2 , etc) basándonos en la información de una muestra aleatoria y un nivel de confianza pre establecido , podremos decir que el parámetro está contenido en el intervalo:

$$\langle Li \ , \ Ls \ \rangle$$

donde:

Li: Límite inferior del intervalo

Ls: Límite superior del intervalo

Las fórmulas para estimar estos límites varían dependiendo del Parámetro que deseamos estimar.

Analytics AoZ

Intervalo de Confianza

Se dice que el intervalo (T1, T2) es un intervalo de confianza 1- α para estimar el parámetro θ, si y sólo sí P(T1< θ <T2) =1 - α, donde T1 es el límite inferior y T2 es el límite superior; además se tiene que $0 < \alpha < 1$.

En lo que sigue estableceremos los intervalos de confianza que nos permitan estimar los siguientes parámetros:

Media o Promedio	μ
Proporción	π
Varianza	$\sigma 2$

Estos intervalos se establecen a partir de datos de una muestra aleatoria y a un nivel de confianza establecida previamente.

Intervalo de Confianza a Estudiar

Intervalos para un parámetro		
μ	**σ^2 conocida**	$IC(\mu) = \bar{x} \pm z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}$
	σ^2 desconocida	$IC(\mu) = \bar{x} \pm t_{\left(n-1;\frac{1+\gamma}{2}\right)} \frac{S}{\sqrt{n}}$
π		$IC(\pi) = p \pm z_{\frac{1+\gamma}{2}} \sqrt{\frac{p(1-p)}{n}}$
σ^2		$IC(\sigma^2) = \left(\frac{(n-1)S^2}{\chi^2_{\left(n-1;\frac{1+\gamma}{2}\right)}}, \frac{(n-1)S^2}{\chi^2_{\left(n-1;\frac{1-\gamma}{2}\right)}} \right)$

INTERVALOS DE CONFIANZA PARA ESTIMAR μ

Analytics AoZ

a) Caso de varianza σ^2 conocida

Si se conoce el valor de la varianza de la población σ^2, se puede afirmar que el intervalo de confianza 1-α para estimar la media μ de la población esta dado por:

$$IC(\mu) = \left\{ \bar{x} \pm \left(z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \right\}$$

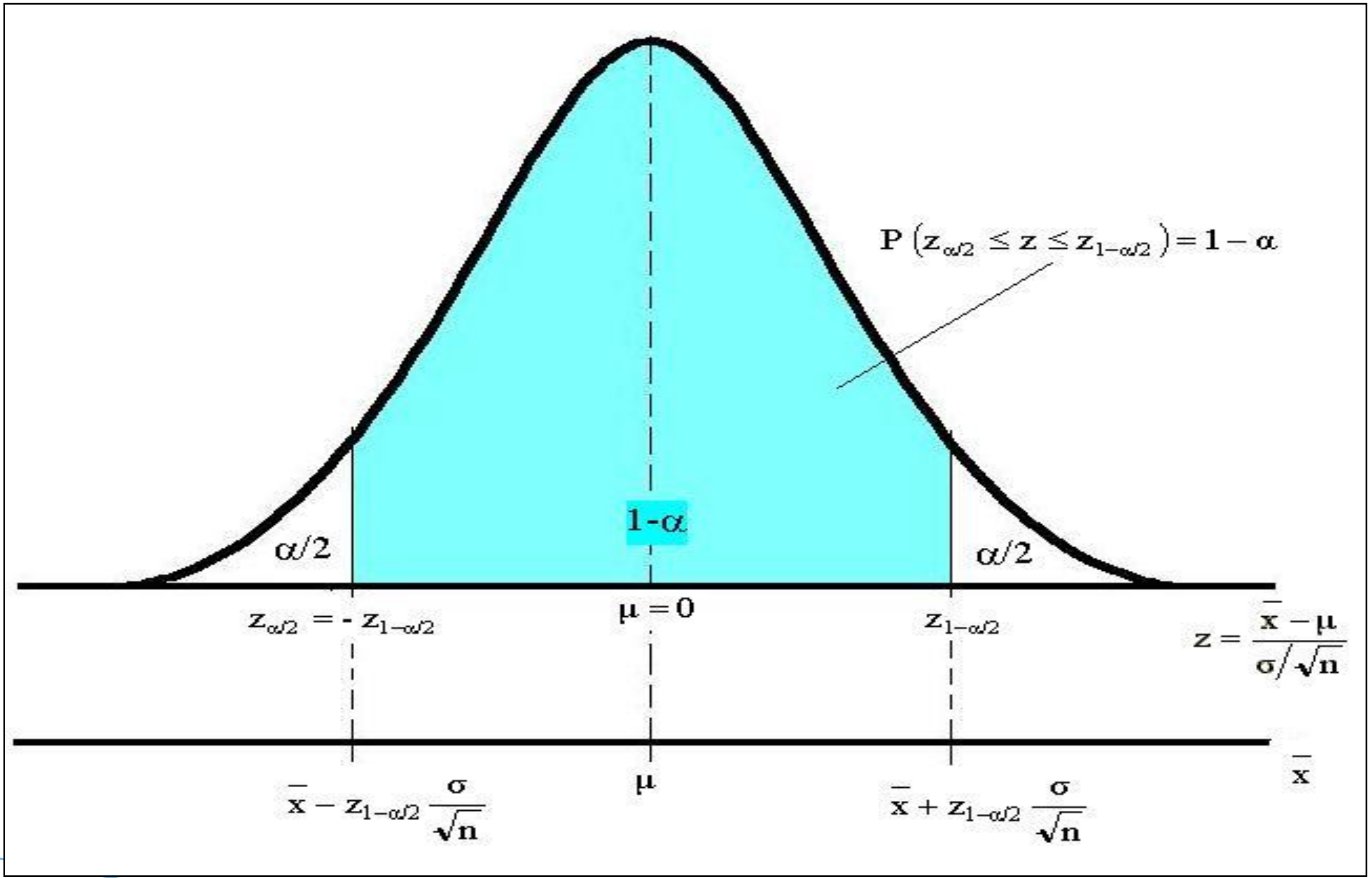
Error de estimación

donde $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ es el error estándar de la media muestral.

Si se conoce el tamaño de la población, el error estándar de la media muestral estará dado por:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$ siempre y cuando: $\dfrac{N-n}{N-1} \leq 0{,}95$ ó $n > \dfrac{N}{20}$

INTERVALOS DE CONFIANZA PARA ESTIMAR μ



INTERVALOS DE CONFIANZA PARA ESTIMAR μ

Analytics AoZ

b) Caso de varianza σ^2 desconocida

Si no se conoce el valor de la varianza de la población σ^2, se puede afirmar que el intervalo de confianza 1-α para estimar la media μ de la población esta dado por:

$$IC(\mu) = \left\{ \bar{x} \pm t_{(n-1;\, 1-\alpha/2)} \frac{s}{\sqrt{n}} \right\}$$

Error de estimación

donde $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$ es el error estándar estimado de la media muestral.

Si se conoce el tamaño de la población, el error estándar estimado de la media muestral estará dado por:

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)}$$ siempre y cuando: $$\frac{N-n}{N-1} \leq 0{,}95 \;\; ó \;\; n > \frac{N}{20}$$

EJEMPLO DE INTERVALOS DE CONFIANZA PARA ESTIMAR μ

Analytics AoZ

Un analista de investigación de mercados desea estimar el ingreso mensual promedio de los hogares de un determinado sector de Lima Metropolitana. Para tal efecto decide que su estimación debe tener una confianza del 95%. Además cuenta con los datos de una muestra de 100 hogares donde se comprobó que el ingreso promedio de los hogares entrevistados fue de US$ 1500, y según cifras oficiales la desviación estándar de los ingresos mensuales de los hogares de dicho sector es de US$ 300. Diga usted ¿entre qué valores se encuentra el ingreso promedio mensual de todos los hogares de dicho sector de Lima Metropolitana?

Solución.-

Se tiene que el nivel de confianza es de 100(1- α)% = 95%.

Esto es, 1-α =0,95 y 1-α/2 = 0,975. Luego se tiene que $z_{1-\alpha/2} = 1{,}96$.

Además se sabe que:

$$n = 100,\ \bar{x} = 1500,\ \sigma = 300\ \text{ y }\ \sigma_{\bar{x}} = \frac{300}{\sqrt{100}} = 30$$

EJEMPLO DE INTERVALOS DE CONFIANZA PARA ESTIMAR μ

Es un caso de varianza σ^2 conocida, luego se debe usar:

$$IC(\mu) = \left\{ \bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

Reemplazando datos se tiene que para el 95% confianza, el intervalo esta dado por:

$$IC(\mu) = \{ 1500 \pm (1,96)(30) \}$$

Es decir:

$$P(1441,2 \leq \mu \leq 1558,8) = 0,95$$

Haciendo uso de R tenemos:

EJEMPLO 2

Analytics AoZ

Los resultados de la revisión de una muestra de 100 cuentas de ahorros en US dólares de BANAMEX, mostraron que el saldo promedio de las cuentas fue de US$ 1000 con una desviación estándar de US$ 500.

a) ¿Cuál será el intervalo de confianza del 95% del saldo promedio de todas las cuentas de ahorros en US dólares de BANAMEX?

b) Si se sabe que BANAMEX tiene 1000 cuentas de ahorros en US dólares, ¿cuál será el intervalo de confianza del 95% del saldo promedio de todas las cuentas?

Solución a)

Se tiene que el nivel de confianza es de $100(1-\alpha)\% = 95\%$.

Esto es, $1-\alpha = 0{,}95$ y $1-\alpha/2 = 0{,}975$. Luego se tiene que $t_{(n-1;1-\alpha/2)} = 1{,}98422$.

Además se sabe que

$$n = 100,\ \bar{x} = 1000,\ s = 500\ \text{ y }\ \hat{\sigma}_{\bar{x}} = 50$$

EJEMPLO 2

Es un caso de varianza σ^2 desconocida, luego se debe usar:

$$IC(\mu) = \left\{ \bar{x} \pm t_{(n-1;1-\alpha/2)} \frac{s}{\sqrt{n}} \right\}$$

Reemplazando datos se tiene que para el 95% confianza el intervalo esta dado por:

$$IC(\mu) = \{ 1000 \pm (1,98422)(50) \}$$

Es decir:

$$P(900,8 \leq \mu \leq 1099,2) = 0,95$$

EJEMPLO 2

Solución b)

Como se me conoce el tamaño de la población, esto es, $N = 1000$. Luego se debe calcular el factor de corrección por finitud:

$$\frac{N-n}{N-1} = \frac{1000-100}{1000-1} = 0{,}9009 \leq 0{,}95$$

Luego se debe modificar el error estándar estimado de la media muestral, el mismo que estará dado por:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)} = \sqrt{\frac{500^2}{100}\left(\frac{1000-100}{1000-1}\right)} = 47{,}458$$

Reemplazando datos se tiene que para el 95% confianza el intervalo esta dado por:

$$IC(\mu) = \{1000 \pm (1{,}98422)(47{,}458)\}$$

Es decir:

$$P(905{,}833 \leq \mu \leq 1094{,}167) = 0{,}95$$

EJEMPLO 3

A una muestra de 35 cigarrillos de una marca conocida en el mercado se le midió el contenido promedio de nicotina obteniéndose un valor de 3,0 miligramos. Asumiendo que el contenido de nicotina por cigarrillo tiene una distribución normal con una desviación estándar de 1,0 miligramos.

a) ¿Cuál cree que sean los intervalos de confianza del 90%, 95%, y 99% que contengan el verdadero contenido promedio de nicotina por cigarrillo de dicha marca?

b) Si el tamaño de muestra fuese de 45 cigarrillos y se comprueba que el contenido promedio también es de 3,0 miligramos de nicotina por cigarrillo. ¿Cuál cree que sea el intervalo de confianza del 95% que contenga el verdadero contenido promedio de nicotina por cigarrillo de dicha marca?

Solución a)

Se tiene que los niveles de confianza son:

$1-\alpha = 0,90$ y $1-\alpha/2 = 0,95$, luego $z_{1-\alpha/2} = 1,645$

$1-\alpha = 0,95$ y $1-\alpha/2 = 0,975$, luego $z_{1-\alpha/2} = 1,96$

$1-\alpha = 0,99$ y $1-\alpha/2 = 0,995$, luego $z_{1-\alpha/2} = 2,58$

EJEMPLO 3

Analytics AoZ

Además se sabe que:

$$n = 35;\ \sigma = 1{,}0;\ \bar{x} = 3{,}0 \ \text{ y } \ \sigma_{\bar{x}} = 0{,}169$$

Es un caso de varianza σ^2 conocida, luego se debe usar:

$$IC(\mu) = \left\{ \bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

Reemplazando datos se tiene que el intervalo:

Para el intervalo de 90% de confianza $IC(\mu) = \{ 3{,}0 \pm (1{,}645)(0{,}169) \}$

➔ $P(2{,}72 \leq \mu \leq 3{,}28) = 0{,}90$

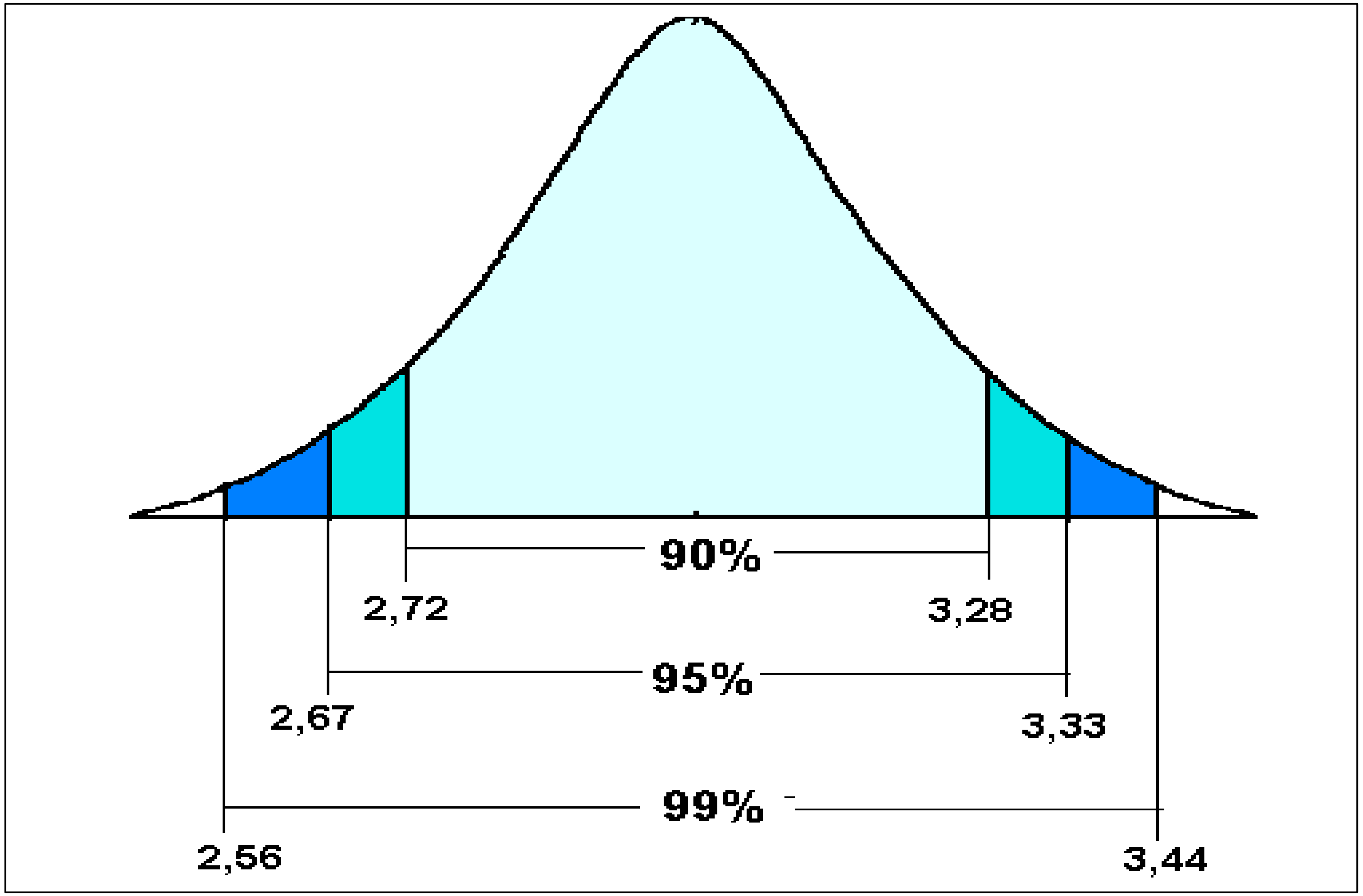
Para el intervalo de 95% de confianza $IC(\mu) = \{ 3{,}0 \pm (1{,}96)(0{,}169) \}$

➔ $P(2{,}67 \leq \mu \leq 3{,}33) = 0{,}95$

Para el intervalo de 99% de confianza $IC(\mu) = \{ 3{,}0 \pm (2{,}58)(0{,}169) \}$

➔ $P(2{,}56 \leq \mu \leq 3{,}44) = 0{,}99$

EJEMPLO 3.-



EJEMPLO 3.-

Analytics AoZ

Solución b)

Se tiene que el nivel de confianza es de 100(1- α)% = 95%

Esto es, 1-α =0,95 y 1-α/2 = 0,975.

Luego se tiene que $z_{1-\alpha/2}$ = 1,96.

Además se sabe que: $n = 45,\ \sigma = 1{,}0,\ \bar{x} = 3{,}0$ y $\sigma_{\bar{x}} = 0{,}149$

$$IC(\mu) = \left\{ \bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

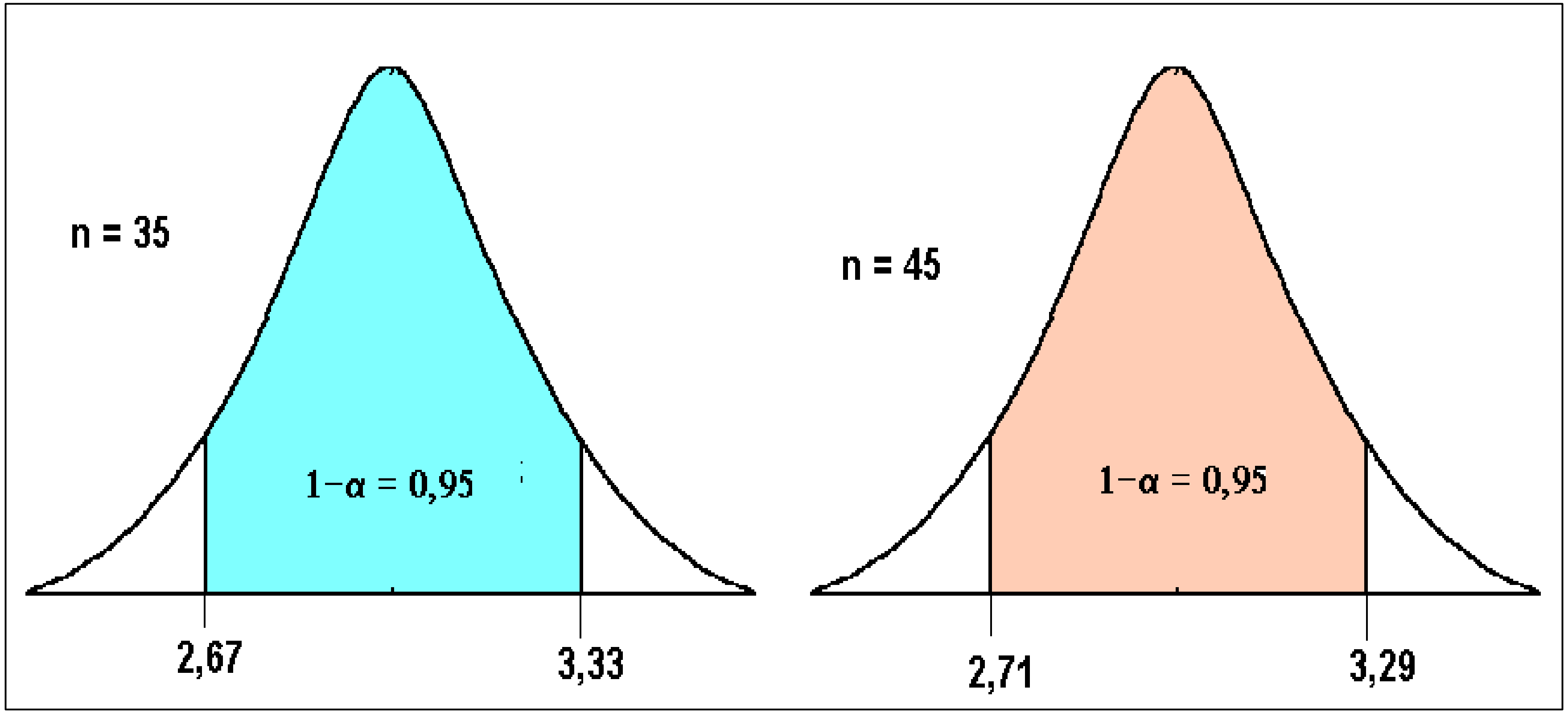
Reemplazando datos se tiene que para el 95% confianza el intervalo esta dado por:

$$IC(\mu) = \{ 3{,}0 \pm (1{,}96)(0{,}149) \}$$

Es decir:

$$P(2{,}71 \leq \mu \leq 3{,}29) = 0{,}95$$

EJEMPLO 3



Determinación del tamaño de la muestra para estima la media μ

- Uno de los primeros problemas a resolver cuando se pretende estimar algún Parámetro es el de determinar cual es el tamaño de muestra adecuado.
  - Fijado el nivel de confianza, se tienen dos factores principales que influyen en el tamaño de la muestra:
    1. La variabilidad de la población ($\sigma 2$)
    2. El grado de error que se está dispuesto a tolerar (E), y que es fijado por el investigador

TAMAÑO DE MUESTRA PARA ESTIMAR LA MEDIA

a) Caso de población infinita .-

El error de estimación de la media de la población μ, está dado por:

$$B = z_{1-\alpha/2}\sigma/\sqrt{n}$$

Luego se puede determinar el tamaño de la muestra mediante

$$n = \frac{z_{1-\alpha/2}^2\sigma^2}{B^2}$$

b) Caso de población finita .-

El error de estimación de la media de la población μ, está dado por:

$$B = z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$$

Luego se puede determinar el tamaño de la muestra mediante:

$$n = \frac{N z_{1-\alpha/2}^2 \sigma^2}{(N-1)B^2 + \sigma^2 z_{1-\alpha/2}^2}$$

Nota:

Si no se conoce la Varianza poblacional se toma una muestra piloto y se estima su valor.

Analytics AoZ

EJEMPLO 4

Analytics AoZ

Se desea tener una estimación de los montos por cobrar de los arbitrios municipales del presente trimestre. Se sabe que en el trimestre anterior la desviación estándar de dichos montos fue S/. 35 .

a) ¿Cuál será el tamaño de muestra necesario de contribuyentes, si se desea tener un límite de error de estimación de S/.3 y una certeza del 95%?

b) Si se sabe que el municipio tiene 5000 contribuyentes, ¿cual será el tamaño de muestra necesario de contribuyentes, si se desea tener un límite de error de estimación de S/.3 y una certeza del 95%?

Solución a) .-

Se asume que el nivel de confianza $1-\alpha = 0{,}95$, luego se tiene que $z_{1-\alpha/2} = 1{,}96$. Además se sabe que $B = 3$ y $\sigma = 35$. Luego se tiene:

$$n = \frac{z^2_{1-\alpha/2}\ \boldsymbol{\sigma}^2}{B^2} = \frac{(1{,}96)^2(35)^2}{3^2} = 522{,}88 \cong 523$$

EJEMPLO 4

Solución b) .-

Se asume que el nivel de confianza 1-α =0,95, luego se tiene que

$z_{1-\alpha/2}$ =1,96. Además B = 3, N = 5000 y σ = 35.

$$n = \frac{N \; z^2_{1-\alpha/2}\boldsymbol{\sigma}^2}{(N-1)B^2 + z^2_{1-\alpha/2}\boldsymbol{\sigma}^2} =$$

$$= \frac{5000 \quad (1{,}96^2)(35)^2}{(5000-1)(3^2) + (1{,}96^2)(35)^2} =$$

$$n = 473{,}46558 = 474$$

INTERVALOS DE CONFIANZA PARA ESTIMAR π

Analytics AoZ

Caso de n > 30

Se puede afirmar que el intervalo de confianza 1-α para estimar la proporción π de la población esta dado por:

$$IC(\pi) = \left\{ p \pm z_{1-\alpha/2}\sqrt{\frac{p(1\text{-}p)}{n}} \right\}$$

Error de estimación

donde $\hat{\sigma}_{\hat{p}} = \sqrt{\dfrac{p(1\text{-}p)}{n}}$ es el error estándar estimado de la proporción muestral.

Si se conoce el tamaño de la población, el error estándar de la proporción muestral estará dado por:

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{p(1\text{-}p)}{n}\left(\frac{N-n}{N-1}\right)}$$ siempre y cuando: $\dfrac{N-n}{N-1} \leq 0{,}95$ ó $n > \dfrac{N}{20}$

EN EL CÁLCULO DEL TAMAÑO DE MUESTRA PARA ESTIMAR LA PROPORCIÓN POBLACIONAL (P), SE DEBEN DE TENER LAS SIGUIENTES CONSIDERACIONES:

Analytics AoZ

En el cálculo del tamaño de muestra para estimar la proporción poblacíonal (π), se deben de tener las siguientes consideraciones:

- Si se tiene información acerca del valor de π **por experiencia u otros estudios realizados** se utiliza este como valor "p" de la fórmula.
- Si no se cuenta con dicha información, se toma una **muestra piloto** y se estima el valor de π, y este se utiliza en la formula.
- Si ninguna de las situaciones anteriores se da, entonces se utiliza el valor de p=0.5.

EJEMPLO DE INTERVALOS DE CONFIANZA PARA ESTIMAR π

El gerente de producción de artefactos eléctricos garantiza que el 95% de los artefactos que se producen están de acuerdo con las especificaciones estándares exigidas. Examinando una muestra de 200 unidades de dichos artefactos se encontró que 25 son defectuosos.
Si se pone en duda la afirmación del gerente de producción ¿cuál será el intervalo de confianza del 96% para la proporción de artefactos que están de acuerdo con las especificaciones estándares exigidas?

Solución.-

Se tiene que el nivel de confianza es de 100(1- α)% = 95%.

Esto es, $1-\alpha = 0,95$ y $1-\alpha/2 = 0,975$. Luego se tiene que $z_{1-\alpha/2} = 1,96$.

Además se sabe que: $n = 200 \quad k = 175 \quad p = 0,875 \quad \hat{\sigma}_p = \sqrt{\dfrac{0,875(0,125)}{n}} = 0,0234$

Luego, si reemplaza los datos anterior en la formula siguiente, se tiene:

$$IC(\pi) = \left\{ p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right\} = \{\ 0,875 \pm 1,96(0,0234)\ \} = P(0,829166 \le \pi \le\ 0,920834) = 0,95$$

TAMAÑO DE MUESTRA PARA ESTIMAR LA PROPORCIÓN

Analytics AoZ

a) Caso de población infinita .-

El error de estimación de la proporción de la población π, está dado por:

$$B = z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Luego se puede determinar el tamaño de la muestra mediante

$$n = \frac{z^2_{1-\alpha/2}\ p(1-p)}{B^2}$$

b) Caso de población finita .-

El error de estimación de la media de la población μ, está dado por:

$$B = z_{1-\alpha/2}\sqrt{\frac{p(1\text{-}p)}{n}\left(\frac{N-n}{N-1}\right)}$$

Luego se puede determinar el tamaño de la muestra mediante:

$$n = \frac{Nz^2_{1-\alpha/2}\,p(1-p)}{(N-1)B^2 + z^2_{1-\alpha/2}\,p(1-p)}$$

EJEMPLO 6

Se desea realizar una encuesta de mercado para estimar la proporción de amas de casa que prefieren un producto al que vende la competencia. Asimismo requiere que el error al estimar la proporción no sea mayor de 4 puntos porcentuales con un grado de confianza del 95%. El Dpto. de muestreo estima que el 20% de las amas de casa podrían preferir el producto. Si cuesta US$ 4 000 poner en marcha la encuesta y US$ 65 por entrevista, ¿cuál será el costo total de la encuesta?

Solución.- Se sabe que el nivel de confianza $1-\alpha = 0{,}95$, luego se tiene que $z_{1-\alpha/2} = 1{,}96$. Además se sabe que $B = 0{,}04$ y $p = 0{,}20$

$$n = \frac{z^2_{1-\alpha/2}\,p(1-p)}{B^2} = \frac{(1{,}96)^2(0{,}20)(0{,}80)}{(0{,}04)^2} = 384{,}16 \cong 384$$

Luego el Costo Total = 4 000+(65)(384) = 28 960

EJEMPLO PROPUESTO DE IC Y TAMAÑO DE MUESTRA PARA ESTIMAR π

Analytics AoZ

Suponga que usted ha realizado una encuesta para estimar la verdadera proporción de consumidores que adquirirían su producto en una población de 25000 habitantes.

a) Si se seleccionó una muestra aleatoria 500 personas y se encontró que el 25% adquirirían el producto. estime la proporción de consumidores que adquirirían su producto mediante un nivel de confianza de 95%

b) Si se exige que la estimación tenga una confiabilidad del 95% con una precisión de 4 puntos porcentuales. y si el trabajo muestral tiene un costo fijo de S/. 2400 y a cada encuestador se le paga S/ 4,0 por entrevista. Halle el costo total de la encuesta si no se tiene estimados previos sobre la proporción real de consumidores que adquiere el producto.

INTERVALO DE CONFIANZA DE LA VARIANZA

Analytics AoZ

El Intervalo de Confianza 1-α para estimar la varianza de la población σ^2 está dado por:

$$IC(\sigma^2) = \left\{ \frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}} \; ; \; \frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}} \right\}$$

El Intervalo de Confianza 1-α para estimar la desviación estándar de la población σ está dado por:

$$IC(\sigma) = \left\{ \sqrt{\frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}}} \; ; \; \sqrt{\frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}}} \right\}$$

EJEMPLO 7

En la fabricación de anillos para motores, se sabe que el diámetro promedio es de 5 cm. con una desviación estándar igual a 0,006 cm. El proceso es vigilado en forma periódica mediante la selección aleatoria de 24 anillos, midiendo sus diámetros. Así en la última muestra se obtuvo una desviación estándar de 0,0065 y se considero que la variabilidad de los diámetros estaba bajo control ¿Cuáles son los límites de la variabilidad esperados con un nivel del 95% de confianza?

Solución.-

Se sabe que $n = 24$, $s = 0,0065$ y que el nivel de confianza $1-\alpha = 0,95$, luego se tiene que:

$$\chi^2_{(n-1;\alpha/2)} = \chi^2_{(23;0,025)} = 11,689 \qquad \chi^2_{(n-1;1-\alpha/2)} = \chi^2_{(23;0,975)} = 38,076$$

$$IC(\sigma^2) = \left\{ \frac{(n-1)s^2}{\chi^2_{(n-1,1-\alpha/2)}} ; \frac{(n-1)s^2}{\chi^2_{(n-1,\alpha/2)}} \right\} = \left\{ \frac{(23)(0,0065)^2}{38,076} ; \frac{(23)(0,0065)^2}{11,689} \right\}$$

$$P\left(0,000026 \le \sigma^2 \le 0,000083\right) = 0,95$$

$$P\left(0,00505 \le \sigma \le 0,00912\right) = 0,95$$

EJEMPLO PROPUESTO DE INTERVALO DE CONFIANZA PARA ESTIMAR σ^2

Se espera tener cierta variación nominal en el espesor de las láminas de plástico que una máquina produce. Para determinar cuándo la variación en el espesor se encuentra dentro de ciertos límites. cada día se seleccionan en forma aleatoria 12 láminas de plástico y se mide en milímetros su espesor.

Los datos que se obtuvieron son los siguientes:

12,6; 11,9; 12,3; 12,8; 11,8; 11,7; 12,4; 12,1; 12,3; 12; 12,5; 12,9

Si se supone que el espesor es una variable aleatoria distribuida normalmente.

a) Obtenga los intervalos de confianza del 90%, 95% y 99% para la varianza del espesor.

b) ¿Qué ocurre cuando el nivel de confianza aumenta?

3. PRUEBA DE HIPÓTESIS Y TIPOS DE ERRÓRES

Analytics AoZ

Supongamos que se forma alguna conjetura o supuesto con respecto a una característica desconocida de la población, como por ejemplo:

"La concentración de Cadmio en el lago Titicaca es 0.0025 ppm"

En este caso será necesario contrastar la validez de dicha conjetura que se ha planteado con respecto al promedio de concentración.

Este supuesto o esta conjetura, la llamaremos ***HIPÓTESTIS ESTADÍSTICA*** y tiene que ser contrastada para determinar su validez. ¿Podemos aceptar como válida la afirmación de que la concentración de Cadmio en el lago Titicaca es 0.0025 ppm?

Este proceso de contrastación se realiza en base a la información obtenida a partir de una muestra aleatoria. Nuestro objetivo al tomar una muestra es extraer alguna conclusión o realizar alguna inferencia respecto a la población.

PRUEBAS DE HIPÓTESIS

Analytics AoZ

- Hipótesis Estadística

Es una aseveración que se hace acerca del valor de un parámetro (de las relaciones entre parámetros, o de la distribución de una variable). Esta afirmación se debe analizar con base a la evidencia de los datos de una muestra.

Otra forma de definición es : “Una hipótesis estadística es un supuesto o una conjetura que se plantea con respecto al valor de algún parámetro de una o más poblaciones”.

Ejemplos:

H1: La concentración de Cobre promedio en las vetas del Batolito de Pataz es igual a 0.84 ppm (μ=0.84ppm).

H2: La proporción de rocas ígneas y metamórficas en la Cuenca Hornillos Alto es a lo más 25% (π<=0.025).

H3: Ambos proyectos de inversión minera tienen el mismo nivel de riesgo ($\boldsymbol{\sigma_1}^2 = \boldsymbol{\sigma_2}^2$)

El objetivo de las prueba de hipótesis es evaluar las proposiciones o conjeturas que se plantean acerca de los parámetros de la población en estudio.

Este objetivo se logra utilizando información obtenida en una muestra aleatoria que es tomada de la población de estudio, con un nivel de significancia (α) fijado y el estadístico de contraste adecuado al parámetro que está sometido a prueba.

Tipos de Hipótesis

Hipótesis Nula.- Es una afirmación que se hace sobre un parámetro de la población, la que esperamos rechazar o desaprobar y está denotada por H_0

Hipótesis Alternativa.- *Es un enunciado que afirma lo opuesto a la hipótesis nula, y representa la conclusión a la que se llegaría si hubiera suficiente evidencia muestral para decidir que la hipótesis nula no será rechazada. Se* denotada por H_1

NOTA: En lo siguiente que desarrollaremos las técnicas que usaremos para resolver pruebas de hipótesis que involucren un promedio μ, una proporción π o una varianza $\sigma_1{}^2$. Así como que involucren dos promedios (μ1, μ2) , dos proporciones(π1, π2) y dos varianzas $(\sigma_1{}^2, \sigma_2{}^2)$ en el caso que se tenga dos muestras independientes. Existen muchos casos más.

PRUEBAS DE HIPÓTESIS

En una Prueba de Hipótesis se evalúa una hipótesis nula H_0 frente a una hipótesis alternativa H_1.

- Pruebas Unilaterales o de una cola
  - Prueba de Cola Inferior o Izquierda.-

    $H_0: \theta = \theta_0$ vs $H_1: \theta < \theta_0$

  - Prueba de Cola Superior o Derecha.-

    $H_0: \theta = \theta_0$ vs $H_1: \theta > \theta_0$

- Prueba Bilateral o de dos Colas

  $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$

DECISIONES POSIBLES DE UNA PRUEBA DE HIPÓTESIS

Analytics AoZ

Tabla Nº 1.1: Estados de la naturaleza y decisiones sobre la hipótesis nula

Decisiones sobre $\mathbf{H_0}$	Estados de la naturaleza	
	$\mathbf{H_0}$ es Verdadera	$\mathbf{H_0}$ es Falsa
No rechazar $\mathbf{H_0}$	**Decisión CORRECTA** Probabilidad $= 1 - \alpha$ $100(1-\alpha)\%$ se denomina **nivel de confianza**	**Error TIPO II** Probabilidad $= \beta$
Rechazar $\mathbf{H_0}$	**Error TIPO I** Probabilidad $= \alpha$ $100\alpha\%$ se denomina **nivel de significación**	**Decisión CORRECTA** Probabilidad $= 1 - \beta$ $100(1-\beta)\%$ se denomina **potencia dela prueba**

__Error Tipo I__.- Se comete cuando rechazamos H_0 siendo ésta verdadera (α)

__Error Tipo II__.- Se comete cuando no rechazamos H_0 siendo ésta falsa (β)

__Potencia de la Prueba__.- Es la probabilidad de rechazar H_0 siendo ésta falsa. Al comparar dos pruebas, se elije aquella que tiene la mayor potencia.

__Nivel de significación de la Prueba__.- Es la probabilidad de rechazar H_0 siendo ésta verdadera.

__Estadística de Prueba__.- Es una estadística muestral que a su vez es un estimador insesgado del parámetro a probar.

__Valor Crítico__.- Es aquel valor de la estadística de prueba que permite rechazar H_0.

EJEMPLO

Considere el siguiente caso como una prueba de hipótesis. Se acaba de recibir un paracaídas sobre el cual un inspector postula la siguiente hipótesis nula "este paracaídas funcionará".

a) Establezca cuidadosamente los 4 posibles resultados al tomar la decisión.

b) Decida sobre la gravedad de los dos errores posibles.

c) Si se pudiesen controlar estadísticamente α y β, ¿qué conjunto de probabilidades preferiría el usuario del paracaídas?.

α = 0,10 y β = 0,001;

α = 0,001 y β = 0,10;

α= 0,05 y β = 0,05.

EJEMPLO

<u>*Solución a)*</u>

Hipótesis planteadas:

H_0: Este paracaídas funcionará

H_1: Este paracaídas no funcionará

Decisiones posibles:

- Afirmar que el paracaídas funcione
- Afirmar erróneamente que el paracaídas no funcione
- Afirmar erróneamente que el paracaídas funcione.
- Rechazar que el paracaídas funcione

EJEMPLO

Solución b)

Errores posibles:

- Error Tipo I: Afirmar erróneamente que el paracaídas no funcionará.
- Error Tipo II: Afirmar erróneamente que el paracaídas funcionará

El error más grabe es el Error Tipo II.

Solución c)

De las tres opciones:

$\alpha = 0{,}10$ y $\beta = 0{,}001$; $\alpha = 0{,}001$ y $\beta = 0{,}10$; $\alpha = 0{,}05$ y $\beta = 0{,}05$.

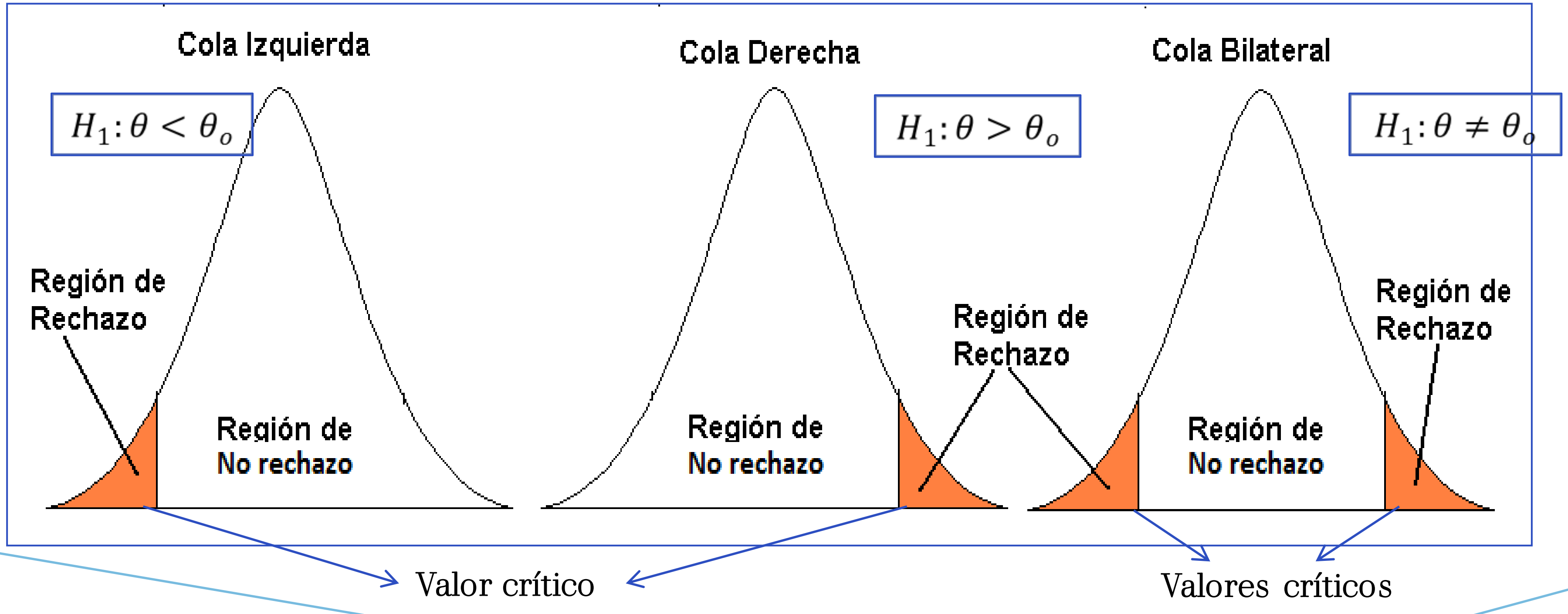
Dada la Gravedad del Error Tipo II, seria más conveniente elegir la opción: $\alpha = 0{,}10$ y $\beta = 0{,}001$.

REGLA DE DECISIÓN

Se establecen en base al valor crítico de la estadística de prueba:

- Región de Rechazo (RR) o Crítica. - Contiene los resultados de la estadística de prueba para rechazar H_0.
- Región de No Rechazo (RNR). - Contiene los resultados de la estadística de prueba para NO rechazar H_0.

$$\boxed{H_0: \theta = \theta_o}$$



P-VALOR

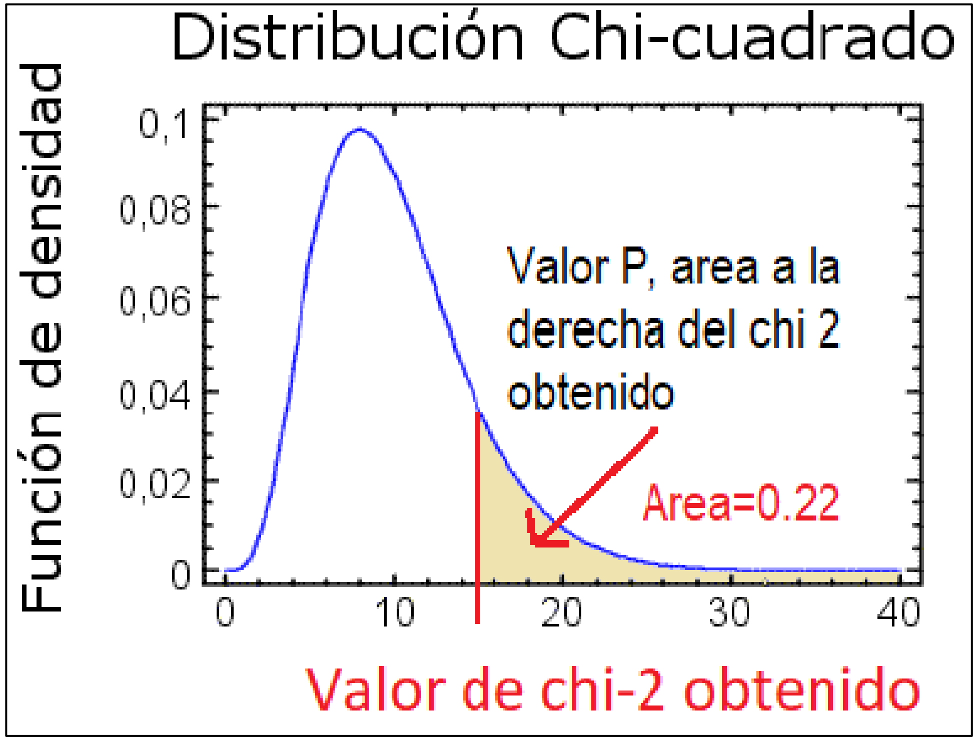
Analytics AoZ

El p-valor se define como la probabilidad correspondiente del estadístico, suponiendo que la hipótesis nula es correcta. Si se cumple con la condición que el p-valor es menos al nivel de significancia (alfa) impuesto arbitrariamente, entonces la hipótesis nula será, eventualmente, rechazada. Es fundamental tener en cuenta que el valor de p está basado en la asunción de la hipótesis nula, por lo tanto es una medida de significación estadística. Es la probabilidad que se presente el resultado obtenido en uno o más extremo. Si el p-valor es pequeño, se dice que H es inverosímil. El valor de P puede interpretarse como <<cuán inverosímil es el resultado observado si H fuera cierta>> o <<hasta que punto el resultado observado es probabilísticamente compatible con H>>. Lo que suele interpretarse , cuando P es pequeña, como que hay <<suficiente evidencia o pruebas en contra de H>> para creer que el resultado es <<estadísticamente significativo>>.

Ver el siguiente link:
https://www.youtube.com/watch?v=eyknGvncKLw&t=23s&ab_channel=DrNic%27sMathsandStats

El valor P de la figura, estaría dado por el área obtenida entre los alores de Chi 2 de 15 hasta el infinito (área de color mostaza), supongamos que esa área sea 0.22 (dato obtenido por cálculo de función densidad) entonces diremos que p-value=0.22.

Si bien es así como se concibe actualmente el valor de P, es necesario recordar que así no fue la concepción original de R.A.Fisher y en su concepción original es incompatible con las pruebas de hipótesis de Neymar Pearson a las que se ha incorporado. El valor P se ideó como medida inferencial flexible mientras que las pruebas de hipótesis fueron ideadas como reglas de conducta, no de inferencia. La combinación de los dos métodos ha llevada a reinterpretar el valor P simultáneamente como una "tasa observada de error" y como medida de carácter probatorio.



Ambas interpretaciones son problemáticas y su combinación no solo ha ocultado las importantes diferencias que tenía Neymar y Fisher respecto a la naturaleza del método científico sino que ha impendido entender las consecuencias filosóficas de estos métodos básicos hoy de uso corriente. El análisis mediante otro método propuesto por Fisher el de la verosimilitud matemática muestra que el valor P exagera mucho el carácter probatorio de los datos contra la hipótesis nula, la verosimilitud aclara la distinción entre las tasas de error y el carácter probatorio inferencial y es un instrumento cuantitativo para expresar la fuerza probatoria.

Fuente: Valores P, pruebas de hipótesis y verosimilitud: las consecuencias para la epidemiología de un debate histórico ignorado Shen N. Goodmar Bol Oficina Sanit Panam 1995;118(2):141-155

Ambas interpretaciones son problemáticas y su combinación no solo ha ocultado las importantes diferencias que tenía Neymar y Fisher respecto a la naturaleza del método científico sino que ha impendido entender las consecuencias filosóficas de estos métodos básicos hoy de uso corriente. El análisis mediante otro método propuesto por Fisher el de la verosimilitud matemática muestra que el valor P exagera mucho el carácter probatorio de los datos contra la hipótesis nula, la verosimilitud aclara la distinción entre las tasas de error y el carácter probatorio inferencial y es un instrumento cuantitativo para expresar la fuerza probatoria.

Fuente: Valores P, pruebas de hipótesis y verosimilitud: las consecuencias para la epidemiología de un debate histórico ignorado Shen N. Goodmar Bol Oficina Sanit Panam 1995;118(2):141-155

CONTRASTE DE HIPÓTESIS

Analytics AoZ

El contraste de hipótesis es un instrumento para tomar una decisión manteniendo controlados los riesgos de error.

Para probar la hipótesis se debe comparar los que arrojan los datos, frente a lo que se espera.

Lo que arrojan los datos, o lo que ha resultado usualmente se traduce en una Prueba estadística. Paramétricas (Chi-cuadrado, t student, ANOVA, entre otras), No Paramétrica (Wilcoxon, Mann-Whitney, entre otros) y Semi Paramétricas. Y esta arroja un valor que se compara con los valores que se esperan (son aquellos que se espera que ocurran o aparezcan si la hipótesis nula es cierta). Esta valor que se ha obtenido en la prueba estadística se le asocia con el Valor P. Por lo que una vez calculado el valor de la prueba estadística, se tiene a lado un valor P.

Lo que se espera que ocurra es nuestra confianza (usualmente mayor 95%). Lo que no se espera que ocurra, será el complemento del la confianza , y le llaman el error alfa (usualmente menor que 0.05 o 5%).

Entonces una vez obtenido el valor P se compara con el alfa, y si el P es menor que el alfa , se rechaza la hipótesis nula. Si el Valor P es mayor que el alfa, no rechazaremos la hipótesis nula, algunos libros de estadística, hablan de PRUEBA DE SIGNIFICACION y CONTRASTE DE HIPOTESIS, y proceden de la siguiente manera:

PRUEBA DE SIGNIFICACIÓN		
Si el valor de *P* es...	**Grande (p. ej. 0,634)**	**Pequeño (p. ej. 0,0001)**
H es...	verosímil	inverosímil
La diferencia...	es explicable por...	no es explicable por...
La diferencia...	no es estadísticamente significativa	si es estadisticamente significativa
A nivel práctico...	no hemos logrado demostrar que la moneda está cargada	creemos que la moneda está cargada
CONTRASTE DE HIPÓTESIS		
Si el estadístico se sitúa en...	**Región de aceptación**	**Región crítica**
	Se acepta H_0	Se rechaza H_0
	Se toma la acción A_0	Se toma la acción H_1

ETAPAS DE UNA PRUEBA (CONTRASTE) DE HIPÓTESIS

1. Plantear las hipótesis nula H_0 y alternativa H_1
2. Especificar el nivel de significación (α) que se va utilizar.
3. Elegir la estadística de prueba (o estadístico de contraste) que debe ser especificada en términos de un estimador insesgado del parámetro que se está contrastando.
4. azar H_0 (Establecer la o las regiones Determinar el valor o valores críticos para rechazar o no regiones críticas)
5. Establecer las reglas de decisión de la prueba.
6. Tomar la decisión correspondiente (rechazar o no rechazar H_0).
7. Formular las conclusiones (Decisión e Interpretación)

Paso 1: Planteamiento de las hipótesis

Analytics AoZ

Hipótesis Nula Ho:

Las características de la hipótesis nula son:

a) Se va a considerar como cierta hasta que se tenga suficiente evidencia de lo contrario.

b) Siempre se incluye el signo de *igualdad*.

c) Es la base para el análisis estadístico de la de la prueba de decisión tomar.

Es la que se contrasta y esta es rechazada o no, dependiendo de la evidencia que presenta la información muestral.

Es la Hipótesis se supone que no hay diferencia, o no hay asociación, o no efecto, se plantea con la intención de ser rechazada.

Hipótesis Alternativa H1:

En resumen las características de la H1:

a) Es lo contrario a la Ho.

b) En general esta hipótesis se establece en términos de la evidencia que se está buscando ($\neq, \leq, \geq$)

c) Es la que define la dirección de la zona de rechazo.

En términos generales esta hipótesis se plantea lo que se está tratando de probar, y está sujeta a la evidencia observada de la muestra.

EJEMPLO

La gerencia general de una cadena de supermercados, contempla la posibilidad de abrir una tienda en cierta zona de la ciudad, si encuentra evidencias de que el gasto promedio mensual en consumo por familia es superior a S/. 2000. La decisión la tomará con base a una encuesta aplicada a 500 familias del sector.

- Identifique la variable en estudio y el parámetro poblacional correspondiente.
- Formule convenientemente las hipótesis nula y alternativa.
- Indique, en términos del enunciado, en que consisten los errores de tipo I y tipo II.

Paso 2: Establecer el nivel de significancia (α)

El nivel de significancia representa la probabilidad de cometer un error de decisión al concluir la prueba (será definido con mayor detalle más adelante cuando hablemos de teoría de errores).
Como es una probabilidad de error, entonces es una probabilidad pequeña (5%, 2%, 1%, 0.1%, etc).
Este nivel de significancia es fijado por el investigador.

Paso 3: Estadístico de Prueba

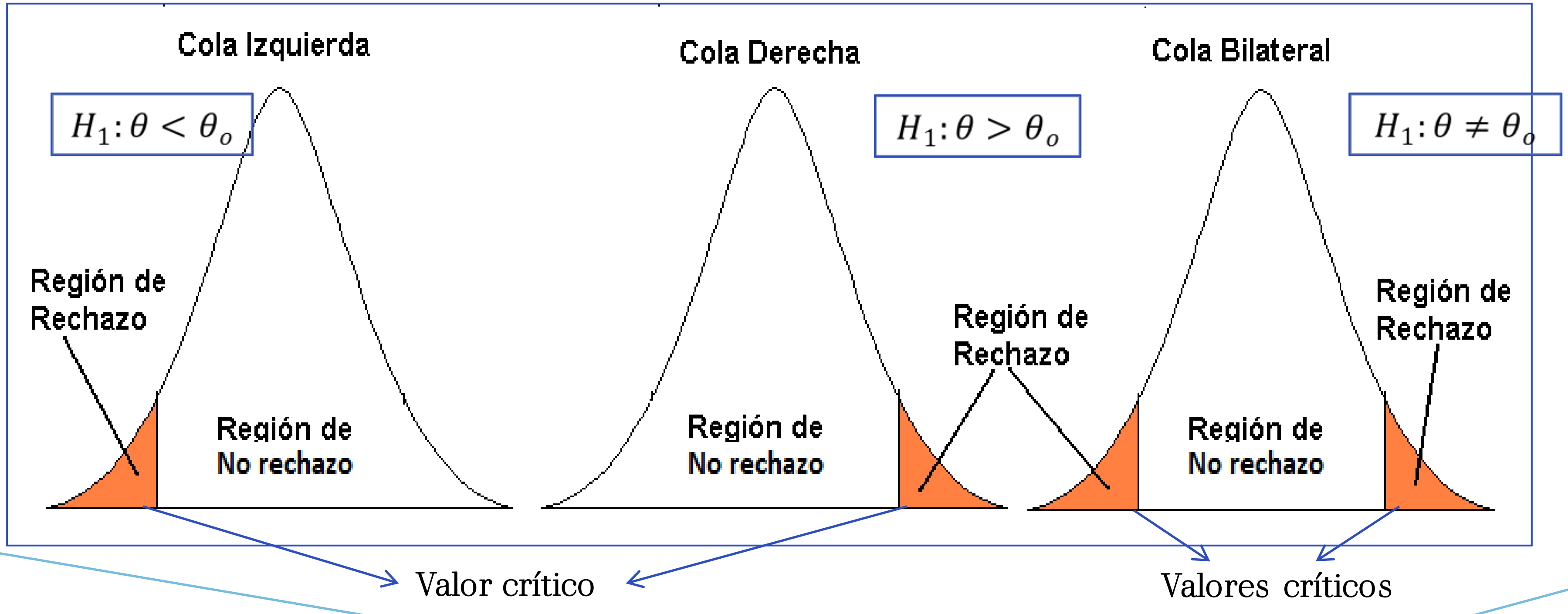
En este paso se selecciona el respectivo estadístico de prueba, es decir la "fórmula" que se utilizará para realizar el contraste.
Esta depende del parámetro sometido a prueba y de la información disponible.
Está basada en la información muestral y se calcula bajo el *supuesto que la Hipótesis Nula es verdadera*.

Paso 4: Determinar el valor crítico o valores críticos

Se establecen en base al valor crítico de la estadística de prueba:

- Región de Rechazo (RR) o Crítica. - Contiene los resultados de la estadística de prueba para rechazar H_0.
- Región de No Rechazo (RNR). - Contiene los resultados de la estadística de prueba para NO rechazar H_0.

$$\boxed{H_0: \theta = \theta_o}$$



Paso 5: Reglas de Decisión de la Prueba

Observar si el estadístico de prueba pertenece o no a la zona de rechazo o de no rechazo (aceptación) de la hipótesis. Además debemos también fijarnos en el p-valor.

Paso 6: Tomar la Decisión

Acorde a la comparación obtenida de la regla de decisión mediante el estadístico de prueba, valor crítico y el p-valor se toma la decisión de rechazar o aceptar la hipótesis nula.

Paso 7: Formular las conclusiones

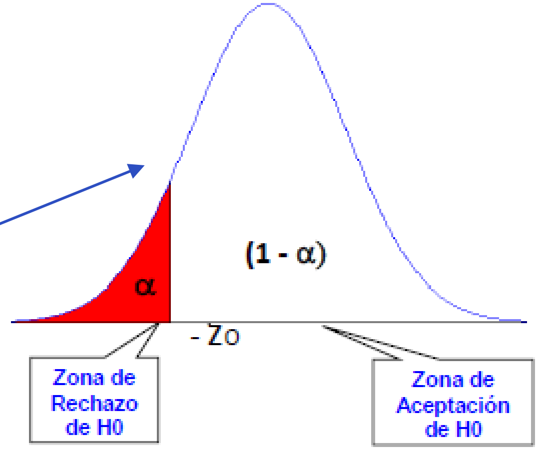
Redactar la conclusión llegada en base al rechazo o aceptación de la hipótesis. Se debe redactar con claridad y siendo lo más explicito posible.

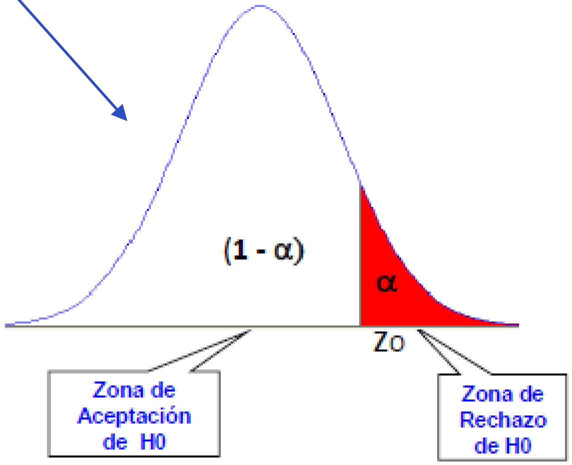
PRUEBA DE HIPÓTESIS PARA LA MEDIA μ

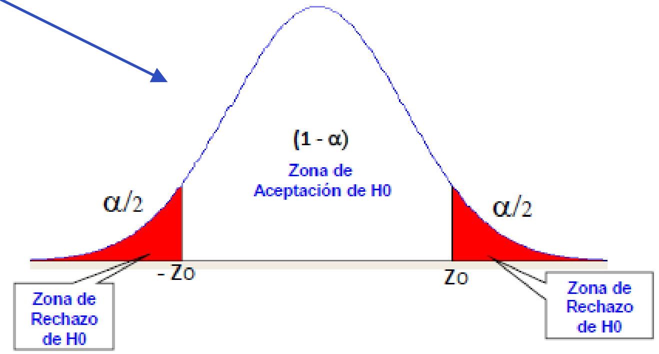
a) *Caso de varianza σ^2 conocida*

Criterio del punto crítico

Hipótesis	Estadística de Prueba	Valores Críticos	Regla para rechazar H_0
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$z_c = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	z_{α}	$z_c < z_{\alpha}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$z_{1-\alpha}$	$z_c > z_{1-\alpha}$
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$z_{1-\alpha/2}$	$\lvert z_c \rvert > z_{1-\alpha/2}$







Analytics AoZ

a) Caso de varianza σ^2 conocida

Criterio del p-valor

Hipótesis	Estadística de Prueba	Càlculo del p-valor	Regla para rechazar H_0
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$z_c = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$P(z < z_C) = P\left(z < \dfrac{\bar{X} - \mu_O}{\sigma/\sqrt{n}}\right)$	$p\text{-valor} < \alpha$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$P(z > z_c) = P\left(z > \dfrac{\bar{X} - \mu_o}{\sigma/\sqrt{n}}\right)$	
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$2P(Z < Z_c)$ o, $2P(Z > Z_c)$	

b) *Caso de varianza σ^2 desconocida*

Criterio del punto crítico

Hipótesis	Estadística de Prueba	Valores críticos	Reglas para rechazar H_0
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_c = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	t_α	$t_c < t_\alpha$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_{1-\alpha}$	$t_c > t_{1-\alpha}$
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$t_{1-\alpha/2}$	$\mid t_c \mid > t_{1-a/2}$

b) Caso de varianza σ^2 desconocida

Criterio del p-valor

Hipótesis	Estadística de Prueba	cálculo del p-valor	Reglas para rechazar H_0
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_c = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$P(t < t_c) = P\left(t < \dfrac{\bar{X} - \mu_o}{s/\sqrt{n}}\right)$	p-valor < α
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$P(t > t_c) = P\left(t > \dfrac{\bar{X} - \mu_o}{s/\sqrt{n}}\right)$	
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$2P(t < t_c)$ o, $2P(t > t_c)$	

Analytics AoZ

EJEMPLO 1.

Un fabricante afirma que mediante el uso de un aditivo especial en la gasolina, los automóviles podrían recorrer por término medio, 3 kilómetros más por litro. Para evaluar este producto se usa una muestra aleatoria de 100 automóviles, alcanzando un incremento medio de 3,4 kilómetros por litro, con una desviación estándar de 1,8 kilómetros. ¿Con α = 0,05, se puede afirmar que con el uso del aditivo, los automóviles incrementarán su recorrido?

Solución .-

De los datos tenemos:

$$\mu_0 = 3{,}00 \quad n = 100 \quad \bar{x} = 3{,}40 \quad s = 1{,}80$$

1. Hipótesis a plantear

H_0: El uso del aditivo no permite el incremento del recorrido.

$H_0: \mu = 3$

H_1: El uso del aditivo (SI) permite el incremento del recorrido.

$H_1: \mu > 3$

Analytics AoZ

EJEMPLO 1.-

2. **Nivel de significación:** $\alpha = 0,05$
3. **Estadística de Prueba (En este caso: σ^2 desconocido)**

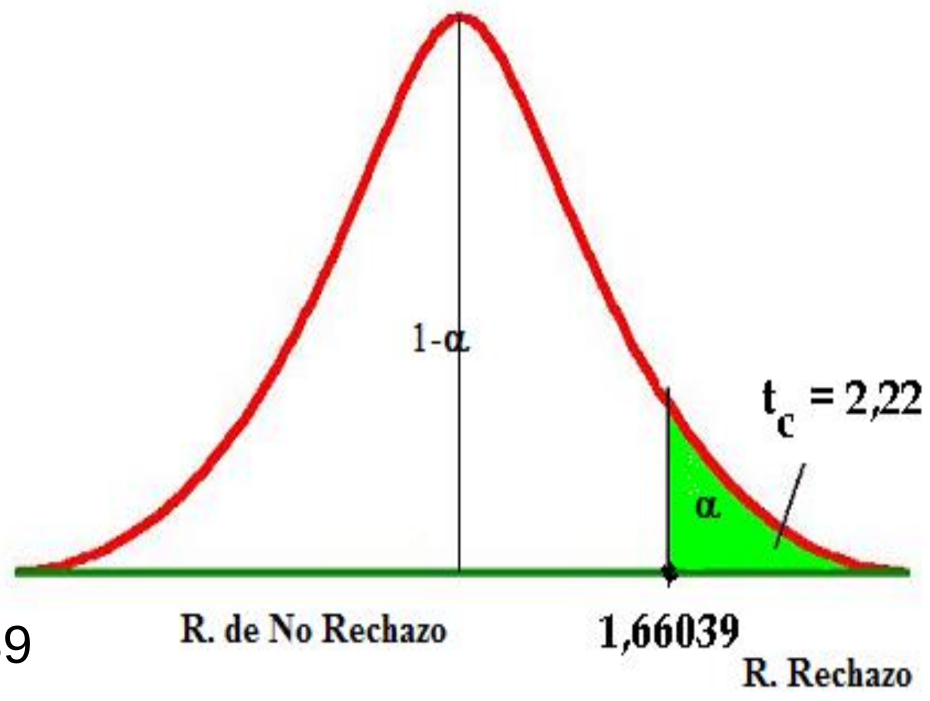
$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3,4 - 3,0}{1,8/\sqrt{100}} = 2,22$$

4. **Valor crítico (Regiones críticas):**

$t_{(n-1;\ 1-\alpha)} = t_{(99;\ 0,95)} = 1,66039$

5. **Regla de Decisión:**

Si $t_c > t_{(n-1;\ 1-\alpha)}$ se rechaza H_0



6. **Decisión de la Prueba:**

Como $t_c=2,22 > t_{(99;0,95)}=1,66039$

la decisión es: Rechazar H_0

7. **Conclusión:**

Con un nivel de significación del 5%, SI SE PUEDE AFIRMAR que el uso del aditivo permite el incremento del recorrido del automóvil.

EJEMPLO 2.-

Un proceso funciona correctamente cuando produce frascos de champú con un contenido promedio de 200 gramos. Una muestra aleatoria de 8 frascos de una remesa presentó los siguientes pesos: 197, 206, 197, 208, 201, 197, 203, 209. Asumiendo que la distribución de los pesos es normal; al nivel del 5%, ¿hay razones para creer de que el proceso no está funcionando correctamente?

Solución.-

De los datos tenemos: $\mu_0 = 200 \quad n = 8 \quad \bar{x} = 202{,}25 \quad s = 5{,}04$

1. Hipótesis a plantear

   H_0: El proceso No está funcionando incorrectamente

   H_0: $\mu = 200$

   H_1: El proceso (SI) está funcionando incorrectamente.

   H_1: $\mu \neq 200$

EJEMPLO 2.-

2. **Nivel de significación**: $\alpha = 0{,}05$
3. **Estadística de Prueba (En este caso: σ^2 desconocido)**

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{202{,}25 - 200}{5{,}04/\sqrt{8}} = 1{,}26$$

4. **Valor crítico:**

$t_{(n-1;\ 1-\alpha/2)} = t_{(7;\ 0{,}975)} = 2{,}365$

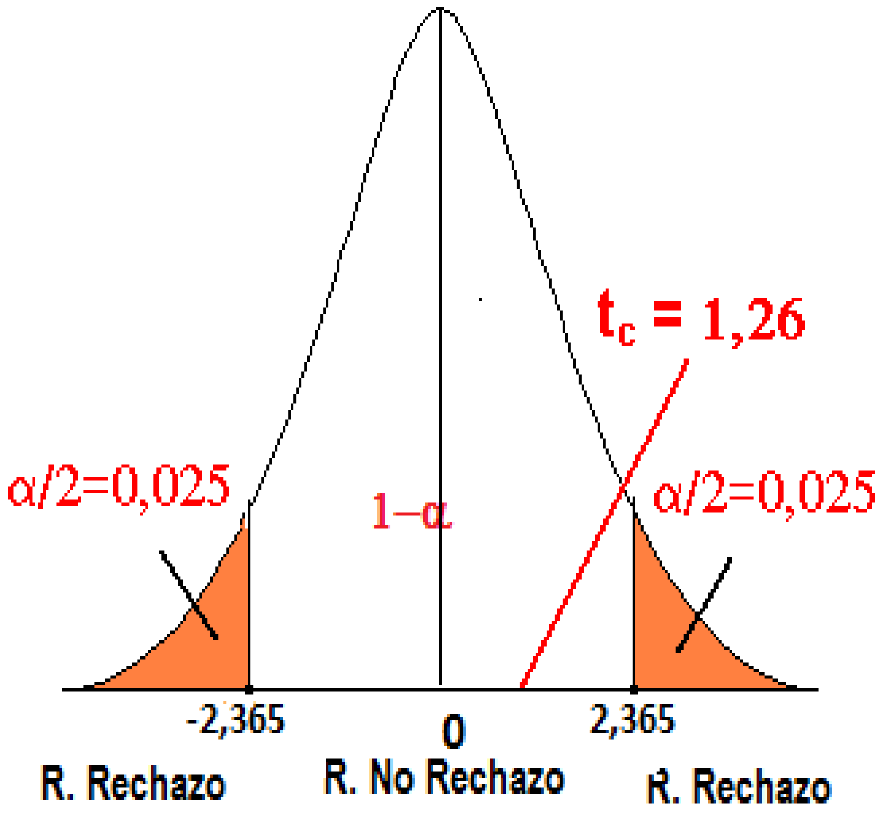
5. **Regla de Decisión:**

Si $|\, t_c \,| > t_{(n-1;\ 1-\alpha/2)}$ se rechaza H_0

6. **Decisión de la Prueba**:

Como $|\, t_c \,| = 1{,}26 < t_{(7;\ 0{,}975)} = 2{,}365$

la decisión es: No Rechazar H_0



7. **Conclusión:**

Con un nivel de significación del 5%, NO SE PUEDE AFIRMAR que el proceso está funcionando incorrectamente.

Analytics AoZ

PROBABILIDAD DE ERROR TIPO I: COLA IZQUIERDA

- P-value Para un valor de $\mu = \mu_0$ se tiene que:

$$P_{value} = P(H_0 \text{ verdadera}) = P\left(z < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

✱ **El Valor Crítico de la Media Muestral:**

Si α es conocido, el valor crítico de la media se puede calcular a partir de la siguiente manera:

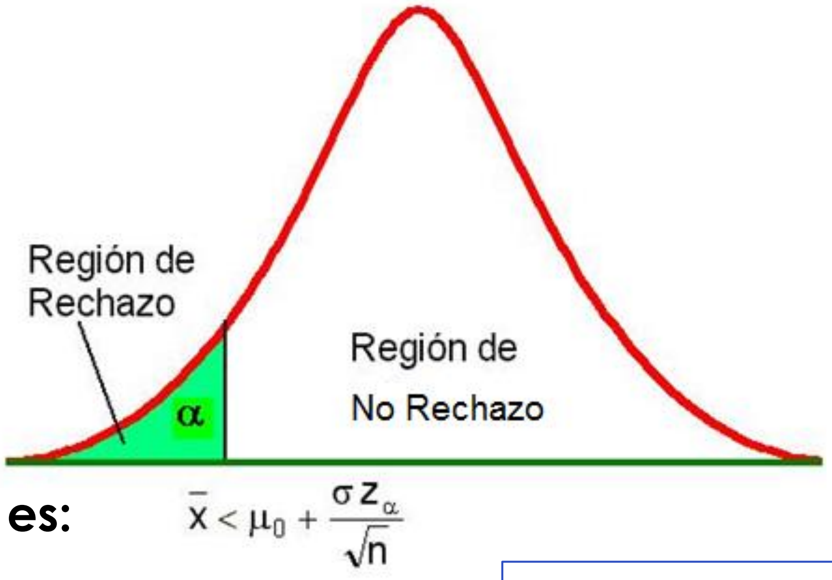
$$\alpha = \mathbf{P(Error\ Tipo\ I) = P(Rechazar\ H_0 / H_0\ es\ verdadera)}$$

Pero cuando Ho es verdadera, Ho se rechaza si: $Z_c < Z_\alpha$, luego:

$$\boldsymbol{\alpha = P(Z_c < Z_\alpha)}$$

$$\text{Es decir}: \alpha = P\left(\frac{\bar{x}_c - \mu_0}{\sigma/\sqrt{n}} < z_\alpha\right)$$

$$\text{Luego}: \alpha = P\left(\bar{x}_c < \mu_0 + \frac{\sigma\, z_\alpha}{\sqrt{n}}\right)$$



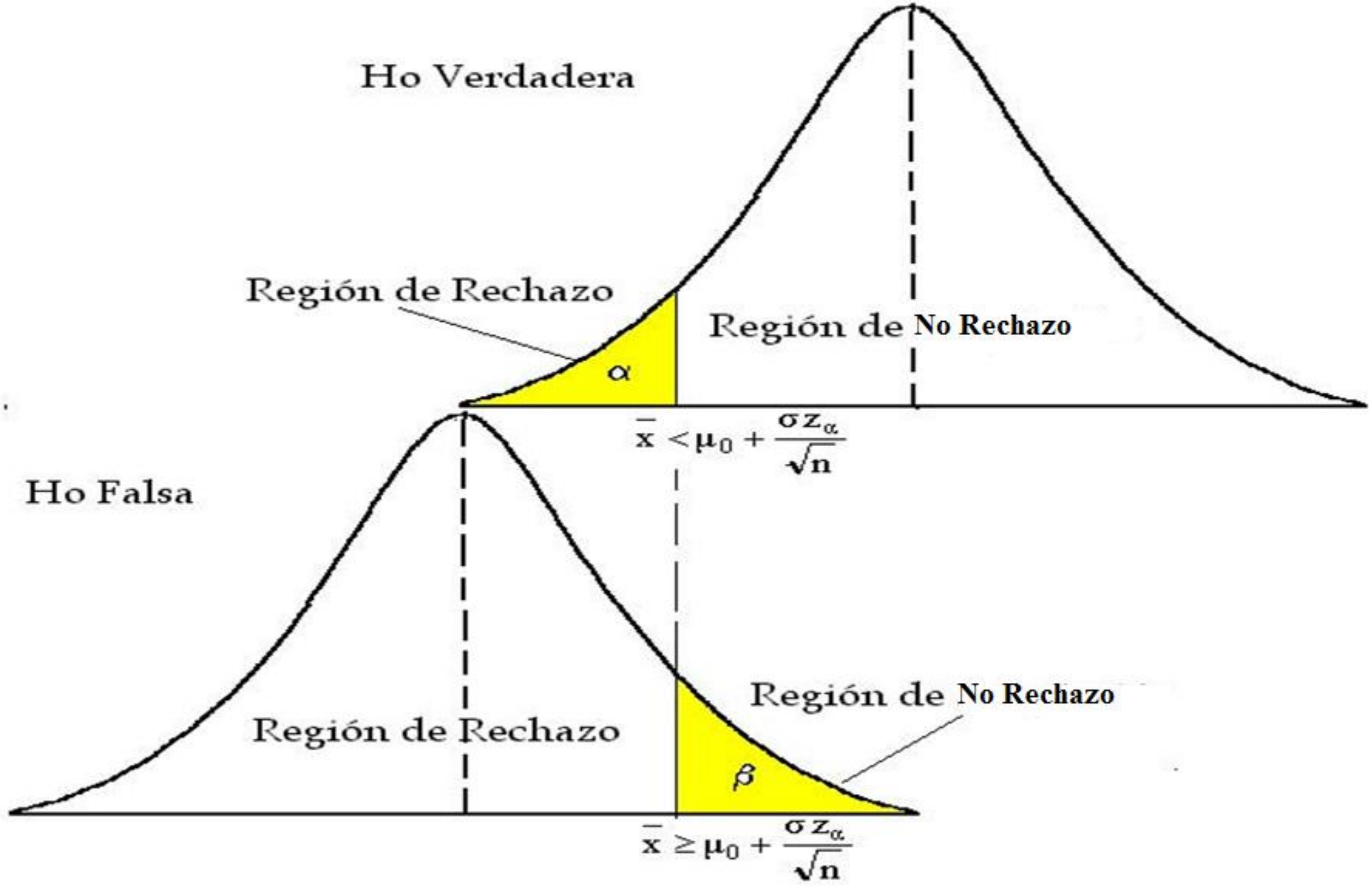
De modo que el valor crítico de la media es:

$$\therefore \quad \bar{x}_c \cong \mu_0 + \frac{\sigma\, z_\alpha}{\sqrt{n}}$$

Luego, Ho se rechaza si $\bar{x}$ $\bar{x}_c$

<

ERROR TIPO I Y TIPO II: COLA IZQUIERDA



PROBABILIDADES DE ERROR TIPO II: COLA IZQUIERDA

✱ Error Tipo II: Para una Hipótesis Alternativa H_1 ($\mu=\mu_1$) y para un α conocido, se tiene que:

$$\beta = P(Error\ \ Tipo\ \ II) = P(No\ \ rechazar\ \ Ho/Ho\ \ es\ \ Falsa)$$

Para el cálculo de β, tenemos 3 alternativas:

i) Estandarizando $\bar{X}_C$

$$\beta(\mu = \mu_1) = P\left(\bar{X}_c \geq \mu_0 + \frac{\sigma\, z_\alpha}{\sqrt{n}}\right)$$

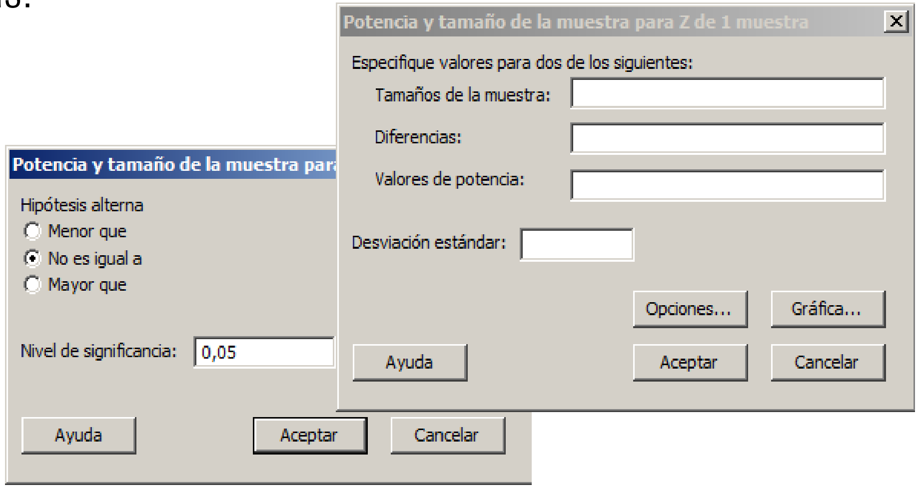
$$= P\left(z \geq \frac{\left(\mu_0 + \frac{\sigma\, z_\alpha}{\sqrt{n}}\right) - \mu_1}{\frac{\sigma}{\sqrt{n}}} \right)$$

ii) Utilizando la distribución de $\bar{X}_C$

$$\bar{X}_C \sim N(\mu_1, \frac{\sigma^2}{n})$$

iii) Utilizando Minitab

Ruta: Estadísticas/Potencia y tamaño de muestra/Elegir: z, t, etc, según corresponda/



Para obtener el valor de la *Potencia* de la prueba, se ingresa el tamaño *n* de la muestra y la diferencia: $\mu_1 - \mu_0$.

Si selecciona "<" en H_1, ingrese diferencia **(-)** Si selecciona ">" en H_1, ingrese diferencia **(+)**

Luego se obtiene:

β = 1 – Potencia.

EJEMPLO 3.-

La cadena de restaurantes BOCOTAS S.A. afirma que el tiempo de espera de los clientes tiene una media de 5 minutos con una desviación estándar de 1 minuto. El departamento de aseguramiento de la calidad de BOCOTAS encontró una muestra de 50 clientes en su local del Suburvia Plaza, donde el tiempo medio de espera fue de 4.25 minutos.

a) En el nivel de significación de 0.05, ¿es posible concluir que el tiempo medio de espera se ha reducido?. Utilice los tres criterios para el desarrollo de una prueba de hipótesis.

b) Hallar la probabilidad de cometer error de tipo II cuando el verdadero tiempo promedio de espera es de 4 minutos. $\alpha = 0.05$

Solución a) *Utilizando el criterio del punto o valor crítico*

De los datos tenemos:

$$\mu_o = 5 \quad n = 50 \quad \bar{x} = 4.25 \quad \sigma = 1.0 \quad \alpha = 0,05$$

1. Planteamiento de las hipótesis

H_0: El tiempo medio de espera NO se ha reducido. H_0: $\mu = 5$

H_1: El tiempo medio de espera (SI) se ha reducido. H_1: $\mu < 5$

EJEMPLO 3.-

2. **Nivel de significación**: $\alpha = 0,05$
3. **Estadística de Prueba (En este caso: $\sigma^2 = 1$ es conocido)**

$$z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{4,25 - 5}{1/\sqrt{50}} = -5,30$$

4. **Valor crítico:**

   $z_\alpha = z_{0,05} = -1,645$

5. **Regla de Decisión:**

   Si $z_c < z_\alpha$, se rechaza H_0

6. **Decisión de la Prueba**:

   Como $z_c = -5,30 < z_{0,05} = -1.645$

   la decisión es: Rechazar H_0

7. **Conclusión:**

   Con un nivel de significación del 5%, SE PUEDE AFIRMAR que el tiempo de espera (SI) se ha reducido.

EJEMPLO 3.-

Solución a): Utilizando el ***criterio del valor crítico de la media muestral*** que en este caso $\bar{X}_C$ es el máximo valor de la media de una muestra de tamaño 50 que permite rechazar H_0.

$$\bar{x}_c \cong \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\bar{x}_c \cong 5 - 1{,}645 \frac{1}{\sqrt{50}}$$

$$\bar{x}_c \cong 4{,}7674$$

Como $\bar{X} = 4{,}25 < \bar{X}c$ entonces Ho se rechaza

Solución a): ***Utilizando el criterio del p-valor***

$$p - valor = P(Z < Z_C) =$$
$$= P(Z < -5{,}30) = 0$$

Luego, como:

$p - valor = 0 < \alpha = 0{,}05$, **Ho se Rechaza**

Como puede verse, luego de aplicar los tres criterios:

- La decisión es: Rechazar H_0.
- La conclusión es: Con un nivel de significación del 5%, SE PUEDE AFIRMAR que el tiempo de espera (SI) se ha reducido.

.

EJEMPLO 3.-

Solución b). Se quiere hallar β, es decir, la probabilidad de cometer error tipo II, cuando el verdadero tiempo promedio de espera $\mu = 4$ minutos. El procedimiento utiliza el valor crítico de la media muestral, veamos:

β = P(No rechazar Ho/Ho es falso)

Esto es, hallar el valor de β cuando $\mu = 4$ (ya que Ho : $\mu = 5$, es **Falso**).

$\beta(\mu = 4) = ?$

El valor crítico de la media muestral es:

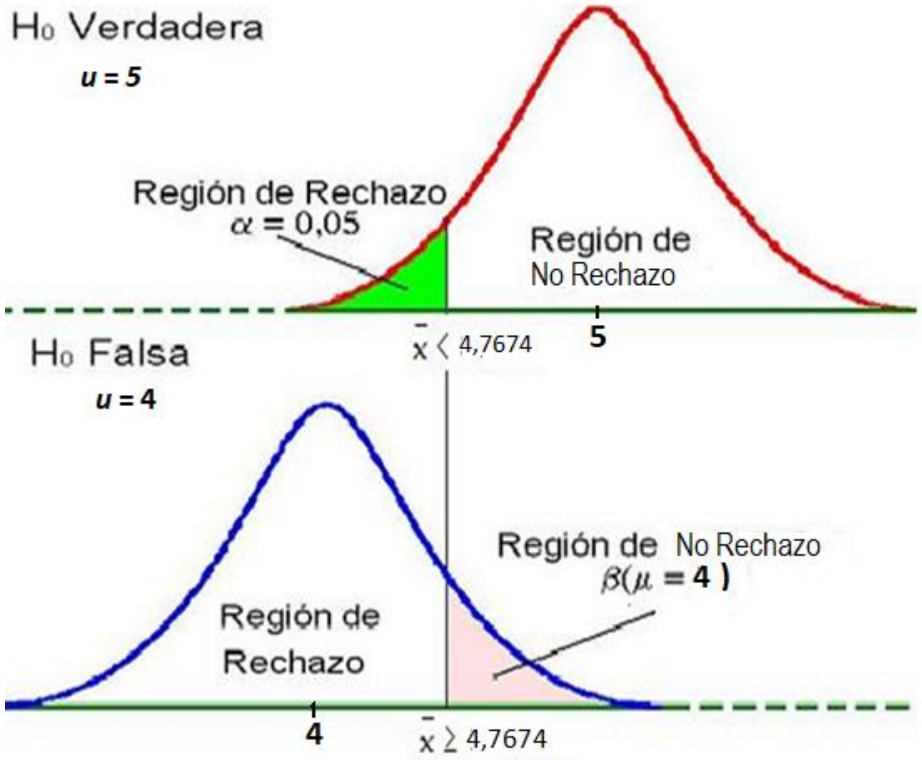
$\bar{X}_C = 4.7674$, luego:

$$\beta = \beta(\mu = 4) = P(\bar{X} \geq \bar{X}_C)$$
$$= P(\bar{X} \geq 4.7674)$$

Pero $\bar{X} \sim N(4;\ {}^{1}/_{50})$ 0.141421^2

Luego:

$$\beta = \beta(\mu = 4) = P(\bar{X} \geq 4.7674) = 0$$



EJEMPLO 4

Un inspector de calidad investiga las acusaciones contra una embotelladora por el deficiente llenado de botellas que debe ser, en promedio 32,5 onzas. Para ello toma una muestra de 60 botellas, encontrando que el contenido medio es de 31,9 onzas de líquido. Se sabe que la máquina embotelladora debe producir un llenado con una desviación estándar de 3,6 onzas.

a) Con α = 0,05, ¿puede el inspector llegar a la conclusión que se están llenando las botellas por debajo de su especificación de contenido?. Use los tres criterios conocidos.

b) ¿Cuál es la probabilidad de que el inspector decida no rechazar H_0 cuando en realidad μ = 31,0

Solución a) De los datos tenemos:

$$\mu_0 = 32,5 \quad n = 60 \quad \bar{x} = 31,9 \quad \sigma = 3,6 \quad \alpha = 0,05 \quad z_\alpha = -1,645$$

Las hipótesis a plantear son las siguientes:

H_0: El llenando de botellas NO es deficiente (promedio NO está por debajo de su capacidad).

H_0: $\mu = 32,5$

H_1: El llenando de botellas es deficiente (promedio está por debajo de su capacidad).

H_1: $\mu < 32,5$

- Criterio del *valor o punto crítico*

2. Nivel de significación: $\alpha = 0{,}05$

3. Estadística de Prueba (En este caso: $\sigma^2 = 3{,}6^2$ es conocido)

$$Z_o = \frac{\bar{X} - \mu_o}{3{,}6/\sqrt{60}} = \frac{31{,}9 - 32{,}5}{3{,}6/\sqrt{60}} = -1{,}29$$

4. Valor crítico:

$z_\alpha = z_{0,05} = -1{,}645$

5. Regla de Decisión:

Si $z_c < z_\alpha$, se rechaza H_0

6. Decisión de la Prueba:

Como $z_c = -1{,}29 > z_{0,05} = -1.645$

la decisión es: No Rechazar H_0

7. Conclusión:

Con un nivel de significación del 5%, NO SE PUEDE AFIRMAR que el el llenado de botellas es deficiente.

EJEMPLO 4

- Criterio del ***valor crítico de la media muestral*** $\bar{x}_c$, que en este caso es el máximo valor de la media de una muestra de tamaño 60 que permite rechazar H_0.

$$\bar{x}_c \cong \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\bar{x}_c \cong 32,5 - 1,64485 \frac{3,6}{\sqrt{60}}$$

$$\bar{x}_c \cong 31,73554$$

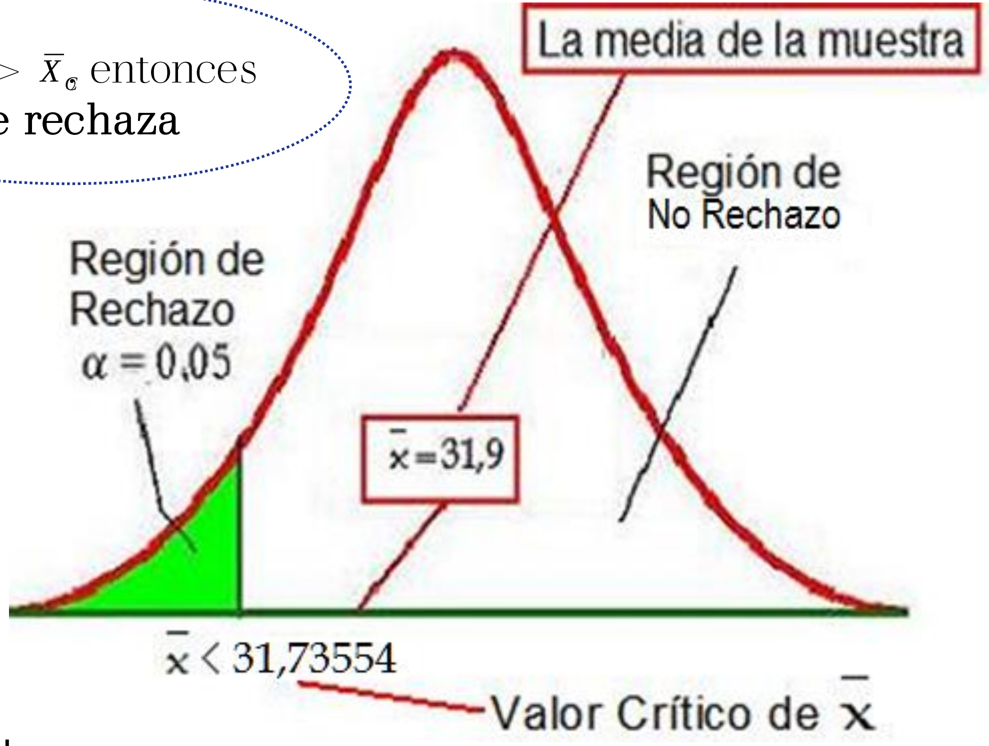
Como $\bar{x} > \bar{x}_c$ entonces Ho NO se rechaza

- ***Criterio del p-valor***

$$p-\text{valor} = P\left(z < z_C = \frac{31,9 - 32,5}{3,6/\sqrt{60}} \right)$$

$$p-\text{valor} = P(z < -1,29)$$

$$p-\text{valor} = 0,098 > \alpha = 0.05$$



Luego, la **decisión** será: No Rechazar H_0.
La **conclusión**: con un NS del 5% NO es posible afirmar que el llenado de botellas es deficiente, ya que su promedio NO está por debajo de su capacidad.

EJEMPLO 4

Solución b): El objeto es calcular la probabilidad de cometer Error Tipo II, es decir:

β = P(No rechazar Ho/Ho es falso)

Esto es, hallar el valor de β cuando $\mu = 31$ (no cuando $\mu = 32{,}5$).

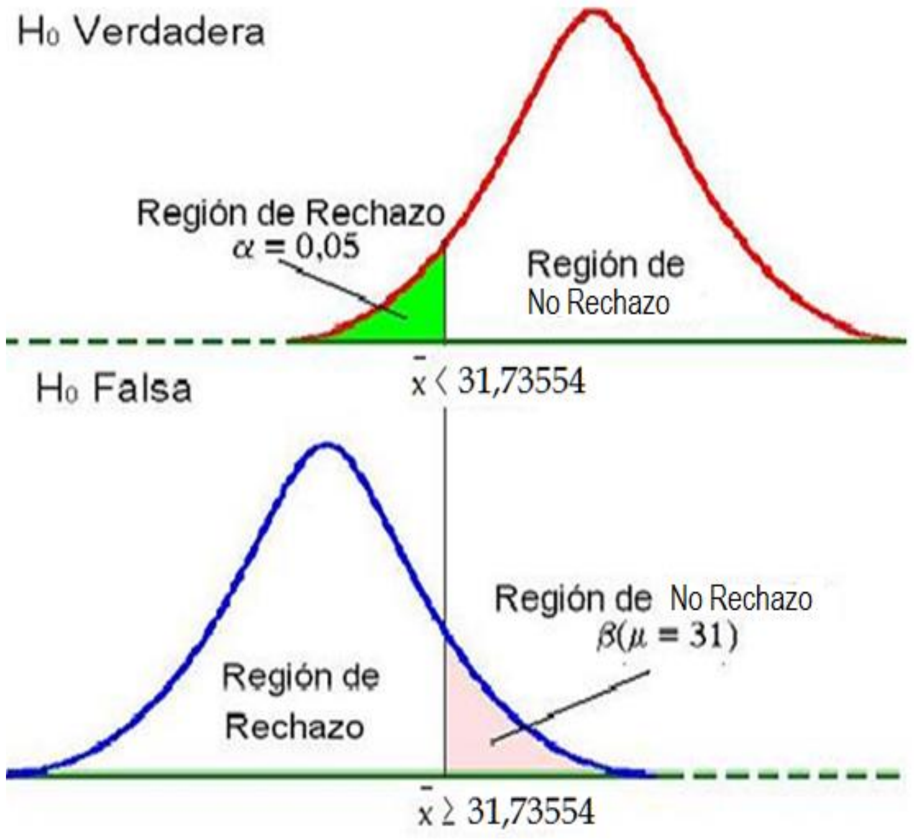
$\beta = P(\mu = 31{,}0)$

$$\beta(\mu = 31) = P(\bar{x}_c \geq 31{,}73554)$$

$$= P\left(z \geq \frac{31{,}73554 - 31}{3{,}6/\sqrt{60}} \right)$$

$$= P(z \geq 1{,}58263)$$

$$\beta(\mu = 31) = 0{,}056753$$



ERROR TIPO I: COLA DERECHA

✱ **P-Value:** Para Ho Verdadera ($\mu = \mu_0$) se tiene que

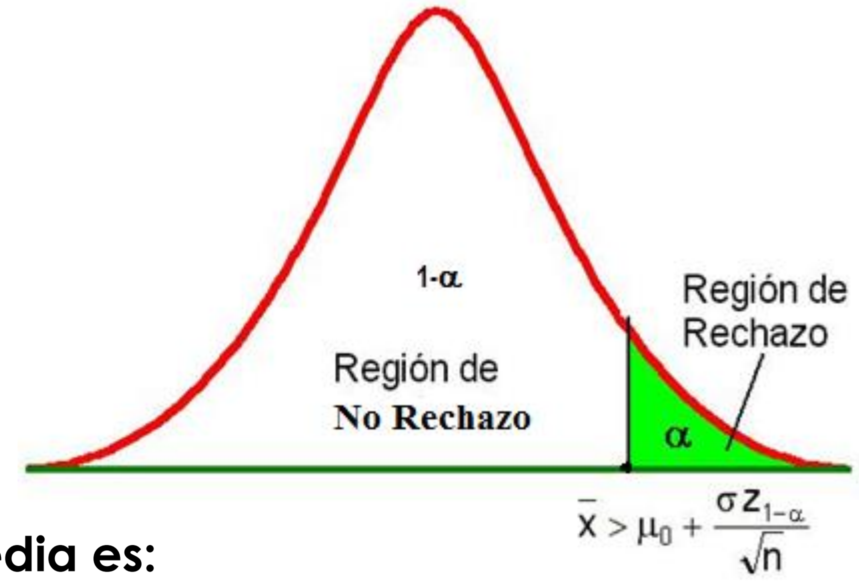
$$p - value = P\left(Z > Z_C = \frac{\bar{X} - \mu_o}{\sigma/\sqrt{n}} \right)$$

✱ El Valor Crítico de la Media Muestral

$$\alpha = P(Error\,Tipo\,I) = P(\text{Re}\,chazar\;Ho / Ho\,es\,V)$$
$$= P(Z_C > Z_{1-\alpha})$$

$$\text{Es decir}: \alpha = P\left(\frac{\bar{x}_c - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha} \right)$$

$$\text{Luego}: \alpha = P\left(\bar{x}_c > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right)$$

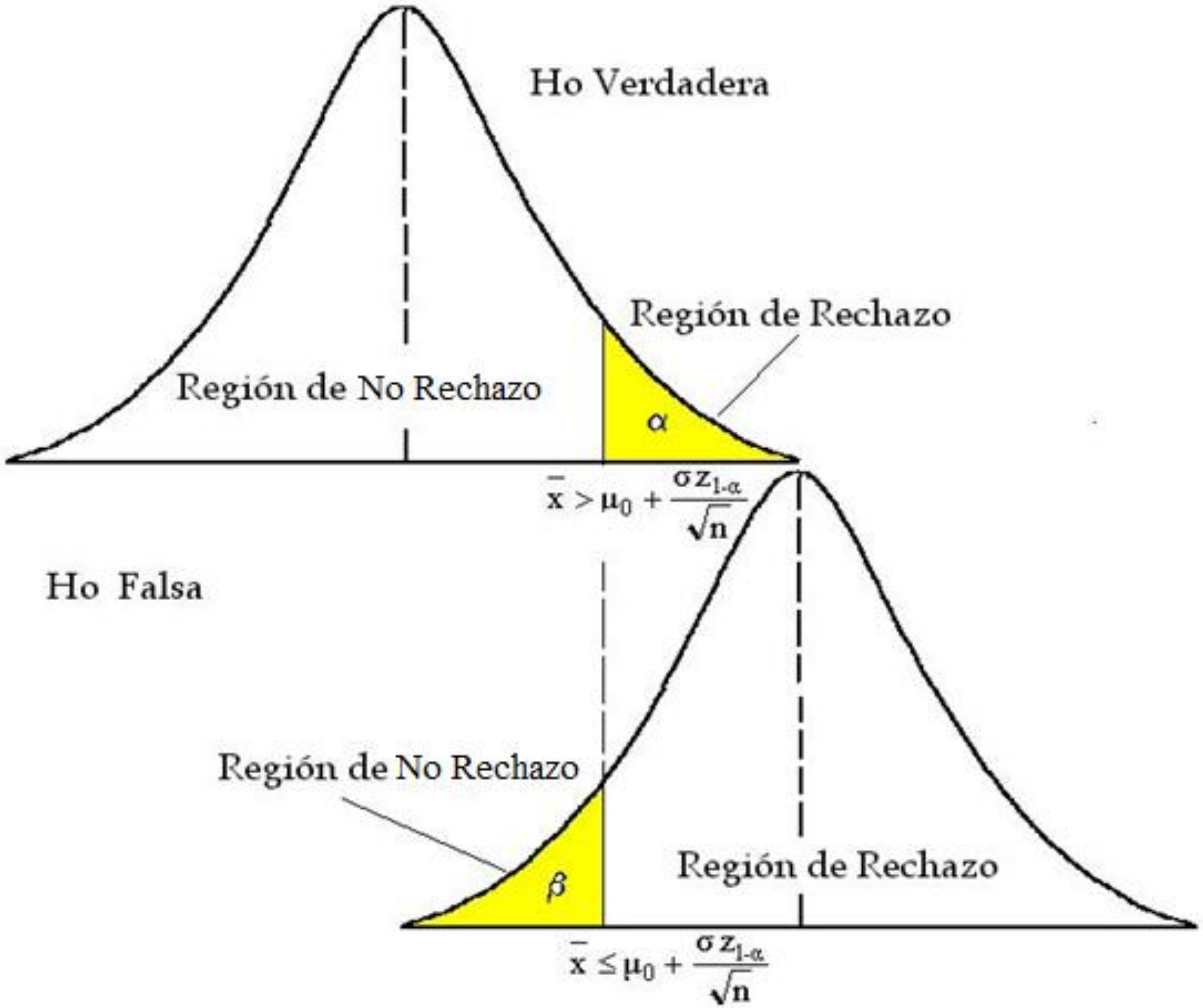


De modo que el valor crítico de la media es:

$$\therefore \bar{x}_c \cong \mu_0 + z_{1-\alpha}\,\sigma/\sqrt{n}$$

Luego, Ho se rechaza si $\bar{X}_C$
>

ERROR TIPO I Y TIPO II: COLA DERECHA



PROBABILIDADES DE ERROR TIPO II: COLA DERECHA

- Error Tipo II. Para un valor de la Hipótesis Alternativa H_1 ($\mu=\mu_1$), y para un α conocido, se tiene que:

$$\beta(\mu=\mu_1) = P(\text{Error Tipo II}) = P(\text{No Rechazar Ho/Ho es Falsa})$$

$$= P\left(\bar{x}_c \leq \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(z \leq \frac{(\mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}) - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right)$$

Nota: Recuerde que existen 3 alternativas para encontrar el valor de $\boldsymbol{\beta}$, tal como se indica en el mismo procedimiento para hipótesis de cola izquierda.