

Unidad III

Parte I

Estimación de parámetros:

Estimación puntual

Estimación por intervalos

Inferencia

Existen dos formas de hacer **inferencia** acerca de un parámetro de una población:

- Estimar su valor desconocido .
- Tomar una decisión acerca de la validez de la afirmación respecto a un valor hipotético de un parámetro de una o más poblaciones.

Ejemplo:

1. Estimar el tiempo medio μ de conexión a internet de un usuario que disfrute de tarifa plana en cierta región.
2. Decidir si el tiempo medio de conexión a internet μ excede o no los 2.5 horas.

Los métodos que consisten en la toma de decisiones acerca de uno ó más parámetros de una población, se denominan **Pruebas de Hipótesis**.

Parámetro

Una v.a. o población **X** está caracterizada por su función de densidad ó de distribución de probabilidad y ésta permite calcular cualquier parámetro poblacional **θ** .

Un **parámetro** es aquella constante que informa de manera sintética de una propiedad relevante o característica de la población o de la variable aleatoria.

Los parámetros clásicos son:

μ = $E(X)$ el valor medio

σ^2 = $V(X)$ la varianza

σ = $V(X)^{1/2}$ el desvío estándar

Estimadores

Si el valor de un **parámetro** θ es desconocido, los **estimadores** son los que nos permitirán la estimación del mismo.

Definición:

Un **estimador** es una v. a. definida a partir de una muestra aleatoria de la v. a. X , ie, X_1, X_2, \dots, X_n . Se lo denota:

$$\hat{\theta} = f(X_1, X_2, \dots, X_n) \text{ o simplemente } \hat{\theta}.$$

Un estimador $\hat{\theta}$ de un parámetro θ es una **fórmula** que muestra como utilizar las observaciones de una muestra para calcular:

- **un sólo número** (un punto) ó,
- **un intervalo.**

Estimadores

Tipos de estimadores:

- **Estimador puntual:** es una **fórmula** que permite calcular un número en base a las observaciones contenidas en una muestra. El número que resulta del cálculo es una **estimación puntual**.

$$\hat{\theta} = f(X_1, X_2, \dots, X_n)$$

Estimador puntual

- **Estimador de intervalo** es una **fórmula** que permite calcular **dos puntos** a partir de los datos de una muestra. El objetivo es formar un intervalo que contenga al verdadero valor del parámetro con un grado de confianza adecuado.

$$\left(\hat{\Theta}_1(X_1, X_2, \dots, X_n), \hat{\Theta}_2(X_1, X_2, \dots, X_n) \right)$$

Estimador por intervalo

IMPORTANTE!!!



Si $\hat{\theta}$ es *estimador puntual* del parámetro θ
→ es un *estadístico*.

Por ser una v. a. definida a partir de una muestra aleatoria de la v. a. X , ie, X_1, X_2, \dots, X_n . Se lo denota:

$$\hat{\theta} = f(X_1, X_2, \dots, X_n) \text{ o simplemente } \hat{\theta}$$

OJO!! La recíproca no es cierta!!!!

IMPORTANTE!!!!



Una **estimación puntual**, por ser un sólo **número**, no proporciona por sí mismo información alguna sobre la **precisión** y **confiabilidad** de la estimación.

Una **estimación por intervalo**, si proporciona información sobre la **magnitud de error máximo de estimación** con que se construyó este intervalo.

Estimadores puntuales - Estimación puntual

Ejemplos

Se desea estimar el **tiempo medio**, μ , de conexión a internet de un usuario que disfrute de tarifa plana en cierta región.

1. Estimación puntual de $E(X) = \mu$

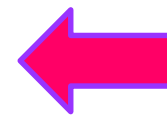
De todos los usuarios de esa región que gozan de tarifa plana se seleccionaron al azar **10** usuarios y se obtuvieron los siguientes tiempos.

1	2	3	4	5	6	7	8	9	10
2.5	3.1	4.3	2.7	3.6	2.7	4.2	2.1	2.8	3.4

El tiempo de conexión medio de **3.14** horas

- El valor $\bar{X} = 3.14$ es la **estimación puntual** de μ , que corresponde a un valor del **estimador puntual**:

$$\bar{X} = \frac{X_1 + \cdots + X_{10}}{10}$$



Estimador puntual

Interpretación de la estimación: El tiempo de conexión promedio de un usuario que dispone de tarifa plana es de aproximadamente 3.14 hs.

Estimadores puntuales - Estimación puntual

Se desea estimar la **variabilidad** del **tiempo**, σ^2 , de conexión a internet de los usuarios que disfrutan de tarifa plana en cierta región.

1. Estimación puntual de $V(X) = \sigma^2$

De los **10** usuarios seleccionados al azar de esa región que gozan de tarifa plana y los tiempos fueron:

1	2	3	4	5	6	7	8	9	10
2.5	3.1	4.3	2.7	3.6	2.7	4.2	2.1	2.8	3.4

Se obtuvo también una **varianza muestral** del tiempo de conexión de **0.53 hs²**.

- **$s^2 = 0.53 \text{ hs.}^2$** , es la *estimación puntual de la variabilidad* en los tiempos de conexión a Internet en usuarios que gozan de tarifa plana, ie, es una estimación de σ^2 .

La *estimación puntual* corresponde a un **valor** particular del *estimador puntual*



$$S^2 = \frac{\sum_{i=1}^{10} (X_i - \bar{X})^2}{10 - 1}$$

Interpretación de la estimación: La fluctuación promedio del tiempo de conexión respecto del tiempo promedio de conexión de un usuario que dispone de tarifa plana es de aproximadamente 0.53 hs.²

Estimación puntual

Resumiendo

- $\hat{\theta}$ **estimador puntual** del parámetro θ
- **Estimación puntual** de θ , corresponde al valor del estimador $\hat{\theta}$ puntual evaluado en los valores observados x_1, x_2, \dots, x_n .

Estimadores puntuales

Parámetros

$\hat{\mu} = \bar{X}$ media muestral,



μ media poblacional

$\hat{\sigma}^2 = S^2$ varianza muestral,



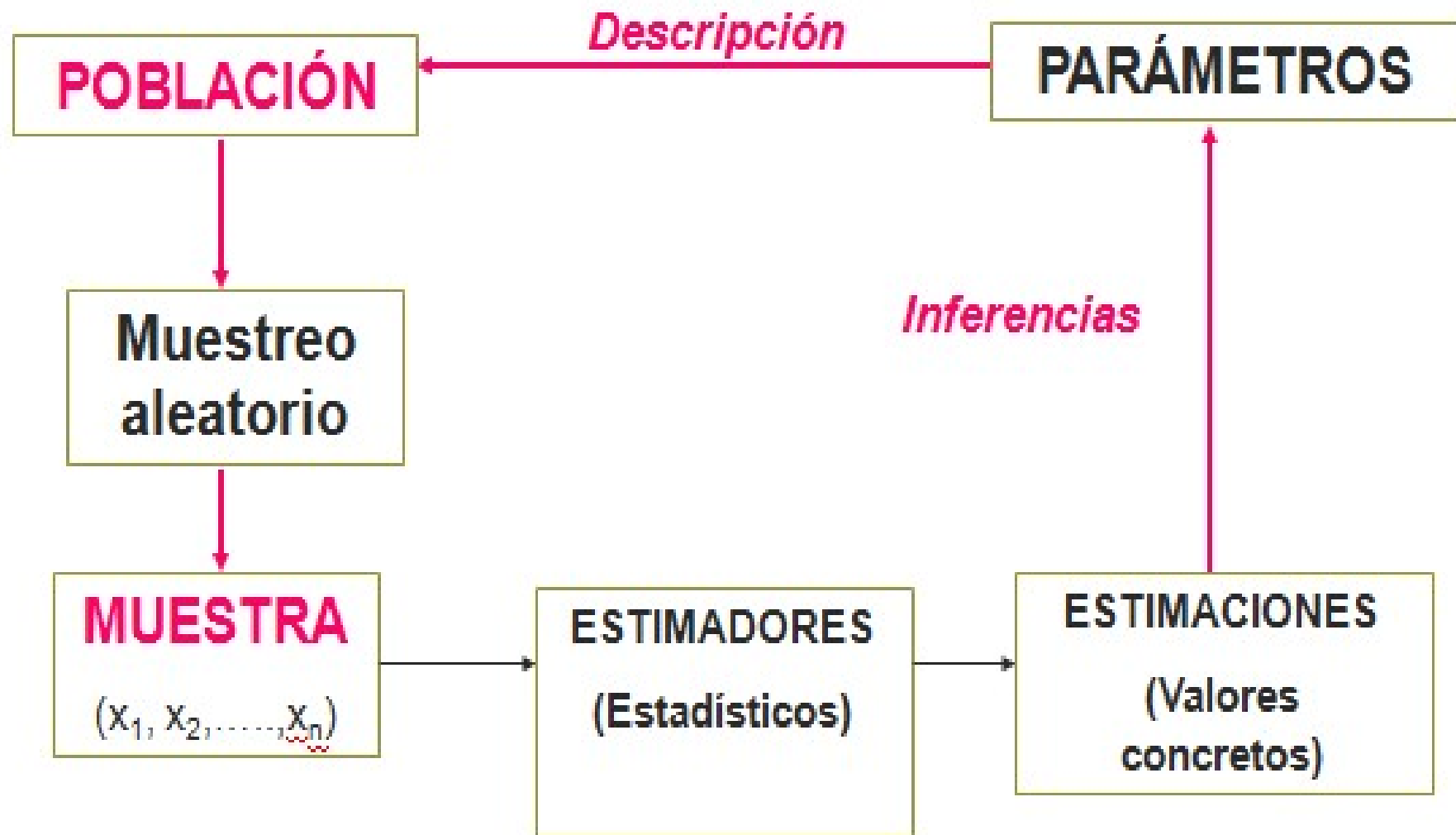
σ^2 varianza poblacional

$\hat{\sigma} = S$ desvío estándar muestral



σ el desvío estándar poblacional

Estimación de Parámetros



Estimación puntual

A una muestra de 25 baterías para teléfono celular de cierta marca se les midió la duración en horas, con el aparato sin operar, obteniéndose un valor promedio de 120 horas y un desvío de estándar de 15 horas.



- a) Estimar puntualmente la duración promedio de todas las baterías de esta marca.
- b) Estimar puntualmente la variabilidad de la duración de todas las baterías de esta marca.

Estimadores de intervalos - Estimación por intervalo

2. Estimación por Intervalo de $E(X) = \mu$ y de $V(X) = \sigma^2$

Se denota: **IC μ** _{95%} (1.2 , 4.27).

Interpretación: El tiempo medio de conexión a Internet, μ , de un usuario con tarifa plana oscila entre 1.2 y 4.27 hs con una confianza del 95%.

Se denota: **IC σ^2** _{95%} (0.88 , 1.47).

Interpretación: La variabilidad del tiempo de conexión a Internet, σ^2 , de un usuario con tarifa plana oscila entre 0.88 y 1.47 hs² con una confianza del 95%.

Estimación por Intervalo

La **estimación por intervalos** de un parámetro θ consiste en determinar de un intervalo (a, b) , que contendrá al verdadero valor del parámetro con un nivel de confianza $1 - \alpha$, $0 < \alpha < 1$, fijado por el consultor estadístico.

$$P\left(\theta \in (\hat{\Theta}_1, \hat{\Theta}_2)\right) = 1 - \alpha$$

Para la construcción de un Intervalo se necesita:

- Una muestra aleatoria X_1, X_2, \dots, X_n de tamaño n extraída de la población X .
- Un estimador puntual $\hat{\theta}$ del parámetro poblacional θ , con distribución de probabilidad conocida.
- El nivel de confianza $1 - \alpha$, establecido a priori por el experimentador (los valores más utilizados son 0.95, 0.90 y 0.99). Se expresa también en porcentaje.

Estimación por Intervalo

Dada una muestra aleatoria de la v.a. \mathbf{X} , X_1, X_2, \dots, X_n de tamaño n , se denomina **intervalo de confianza a un nivel $1 - \alpha$** al intervalo:

$$\left(\hat{\Theta}_1(X_1, X_2, \dots, X_n), \hat{\Theta}_2(X_1, X_2, \dots, X_n) \right)$$

$\hat{\Theta}_1$ y $\hat{\Theta}_2$ son v. a. que dependen del **estimador puntual** del parámetro θ .

Estimación por Intervalo

Si se extrae una muestra de tamaño n y con las observaciones, x_1, x_2, \dots, x_n , se calculan los extremos del intervalo, se dispondrá del **intervalo de confianza** (a, b) para el parámetro θ , es decir,

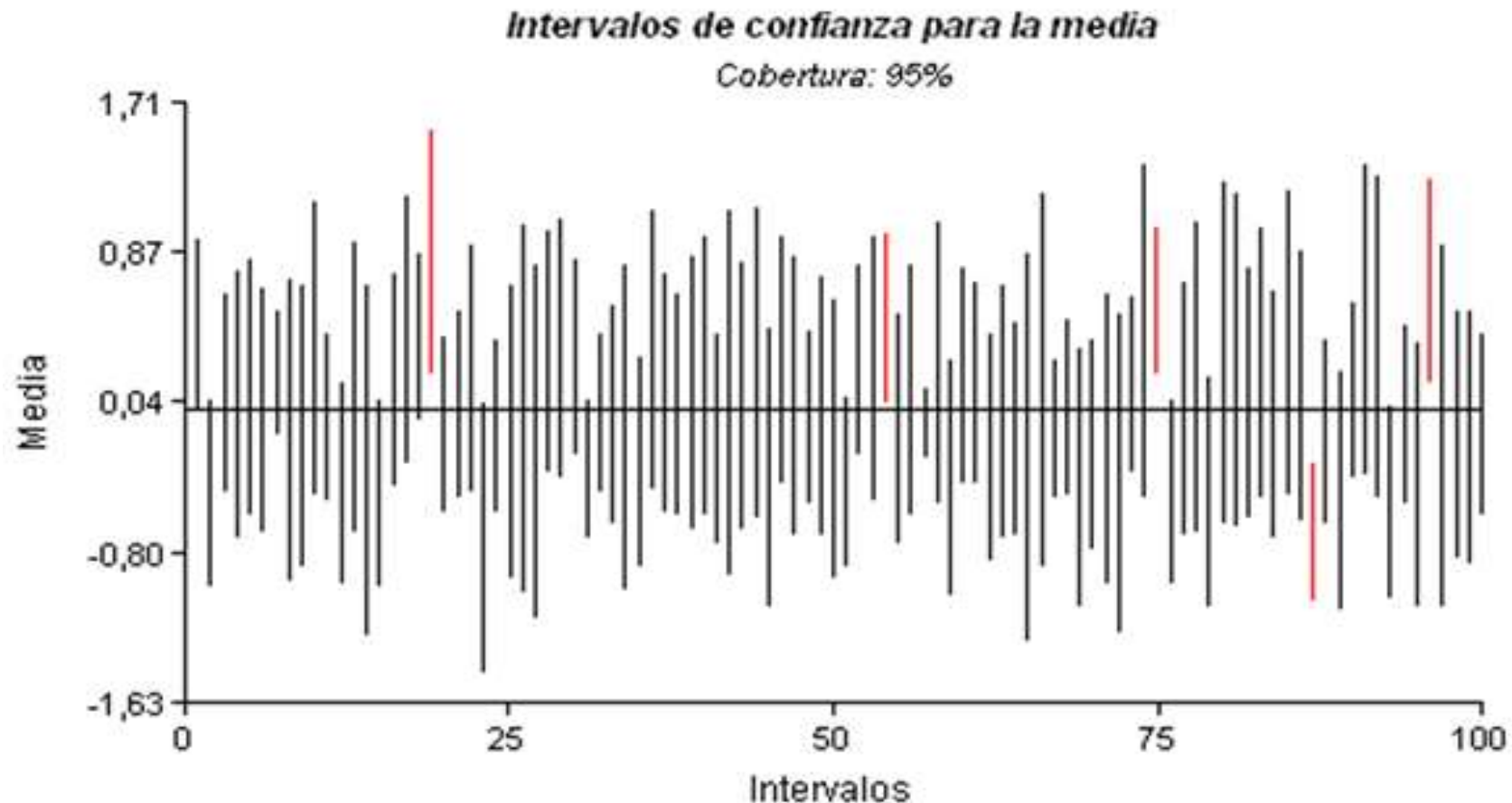
$$\text{extremo inferior} = \mathbf{a} = \hat{\Theta}_1(x_1, x_2, \dots, x_n)$$

$$\text{extremo superior} = \mathbf{b} = \hat{\Theta}_2(x_1, x_2, \dots, x_n)$$

Interpretación del intervalo de confianza:

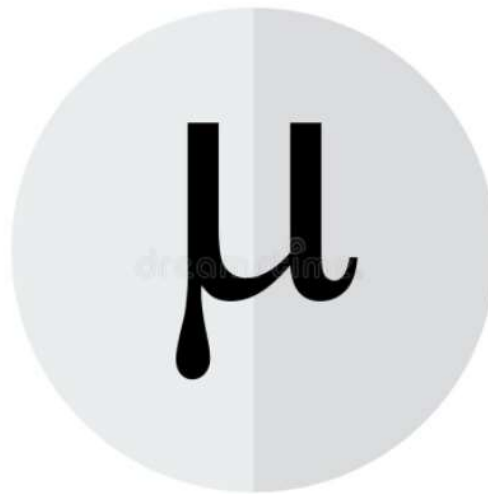
1. El parámetro θ oscila entre **a y b** con un nivel de confianza de $1 - \alpha$ o un porcentaje de confianza de $(1 - \alpha) 100 \%$.
2. El intervalo **(a, b)** contendrá al verdadero valor del parámetro con un nivel de confianza $1 - \alpha$ o con un porcentaje de confianza de $(1 - \alpha) 100 \%$.

Estimación por Intervalo: nivel de confianza



De 100 muestra de tamaño n extraídas de la población, un **nivel de confianza de 95%** implica que el **95%** de las muestras daría lugar a un intervalo que incluye a $\mu = 0.04$, y sólo **5%** de las muestras producirá un intervalo erróneo. Cuanto mayor sea el nivel de confianza podremos creer que el valor del parámetro que se estima está dentro del intervalo.

Intervalo de confianza para estimar la media de una población



Intervalo de confianza para la media de una población

Sea $\mathbf{X} \sim N(\mu, \sigma^2)$, donde μ es el parámetro desconocido. Sea X_1, X_2, \dots, X_n una muestra aleatoria de la v.a. \mathbf{X} y sea \bar{X} el estimador puntual media muestral.

Caso I: varianza σ^2 conocida

Como $X \sim N(\mu, \sigma^2)$, con σ^2 conocida entonces $\bar{X} \sim N(\mu, \sigma^2/n)$ de donde resulta que

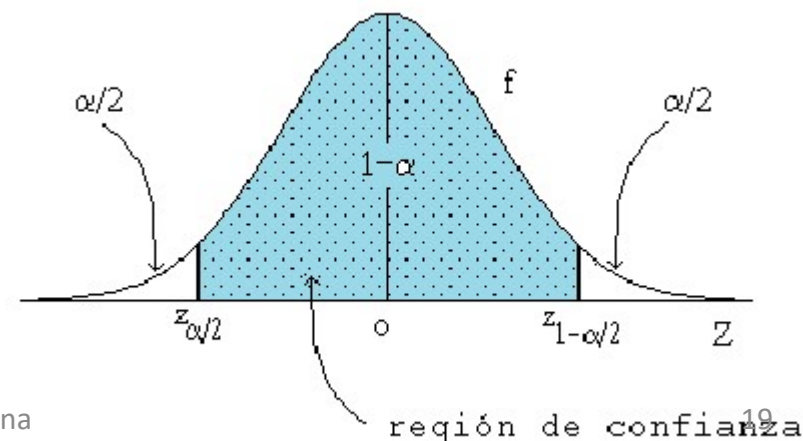
$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = Z \sim N(0, 1)$$

Notar que aunque \mathbf{Z} depende de μ , su distribución de probabilidad no.

Dado $1 - \alpha$, el nivel de confianza elegido para construir el intervalo, es posible encontrar z_0 tal que:

$$P(-z_0 < Z < z_0) = 1 - \alpha$$

donde $-z_0 = z_{\alpha/2}$ $z_0 = z_{1-\alpha/2}$



Intervalo de confianza para la media de una población

$$P(-z_0 < Z < z_0) = 1 - \alpha$$

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

Si se reemplaza la v.a. **Z** por su expresión, esta igualdad resulta:

$$P\left(-z_0 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_0\right) = 1 - \alpha$$

$$P\left(-z_0 * \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < +z_0 * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(-\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z_0 * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{X} + z_0 * \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Si se despeja el parámetro μ en la expresión dentro del paréntesis se obtiene:

$$P\left(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Intervalo de confianza para la media de una población

$$P\left(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Obsérvese que:

$$\hat{\Theta}_1(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \bar{X} - z_0 * \frac{\sigma}{\sqrt{n}} \quad \text{donde} \quad -z_0 = z_{\alpha/2}$$

y

$$\hat{\Theta}_2(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n) = \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}} \quad \text{donde} \quad z_0 = z_{1-\alpha/2}.$$

Por lo tanto, con una probabilidad $1 - \alpha$, el **intervalo aleatorio**

$$\left(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}}\right) \text{ contendrá el parámetro } \mu.$$

Intervalo de confianza para la media de una población

Si se extrae una muestra de tamaño n y con observaciones, x_1, x_2, \dots, x_n , se calculan los extremos del intervalo, se dispondrá del **intervalo de confianza**

$$(\bar{x} - z_0 * \sigma / \sqrt{n}, \bar{x} + z_0 * \sigma / \sqrt{n}) = (\mathbf{a}, \mathbf{b})$$

para el parámetro μ para un nivel de confianza de $(1-\alpha)$ 100 %.

IMPORTANTE!!!!

Este intervalo debe interpretarse ***muy cuidadosamente***. Es incorrecto, decir que la probabilidad de que el parámetro μ se encuentre contenido en el intervalo $(\bar{x} - z_0 * \sigma / \sqrt{n}, \bar{x} + z_0 * \sigma / \sqrt{n})$ es igual a $1 - \alpha$.

No puede asociarse ningún valor de **probabilidad** al intervalo obtenido

$$\bar{x} - z_0 * \sigma / \sqrt{n} < \mu < \bar{x} + z_0 * \sigma / \sqrt{n}$$

debido a que sólo contiene **constantes**.

Si se permite asignar a dicho intervalo un **grado de confianza** o **porcentaje de confianza** de que el parámetro μ se encuentre contenido en él .

Intervalo de confianza para la media de una población

Ejemplo 1 (caso I)

Un directivo de cierta empresa de software ha comprobado que los resultados obtenidos en los tests de aptitud por los solicitantes de un determinado puesto de trabajo sigue una distribución normal con una desviación estándar de 32,4 puntos. La media de las calificaciones de una muestra aleatoria de nueve tests es de 187,9 puntos.

- a) Estimar **puntualmente** el verdadero valor de la calificación media poblacional del grupo de solicitantes actual.
- b) Construir un **intervalo de confianza del 95%** para estimar la calificación media poblacional del grupo de solicitantes actual.
- c) A partir de estos resultados muestrales, un estadístico construyó otro **intervalo de confianza**, para estimar la calificación media poblacional, con extremos 165,8 a 210,0 puntos. Determine el nivel de confianza que utilizó en para calcular dicho intervalo.
- d) Sin realizar los cálculos, determinar si un intervalo de confianza del 90% para estimar la calificación media poblacional tendría mayor, menor o la misma longitud que el calculado en el inciso **b)**. ¿Cuál de los dos intervalos es más preciso?

X = puntaje obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo(en puntos) ”

$$X \sim N(\mu, \sigma^2 = 32,4^2 \text{ puntos}^2)$$

Datos: $n = 9$ tests, $\bar{X} = 187,9$ puntos

a) Estimar puntualmente el verdadero valor de la calificación media poblacional del grupo de solicitantes actual.

Se desea estimar puntualmente el parámetro μ :

μ = puntaje promedio obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo.

El **estimador puntual** de μ es la v.a. media muestral \bar{X} . Una **estimación puntual** es $\bar{X} = 187,9$ puntos .

Luego, afirmamos que:

El puntaje promedio obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo es de aprox. 187,9 puntos.

b) Construir un intervalo de confianza del 95% para estimar la calificación media poblacional del grupo de solicitantes actual.

X = puntaje obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo(en puntos) ”

X ~ N (μ , $\sigma^2 = 32,4^2$ puntos²)

Se desea estimar mediante un intervalo de confianza el parámetro μ :

μ = puntaje promedio obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo.

Datos:

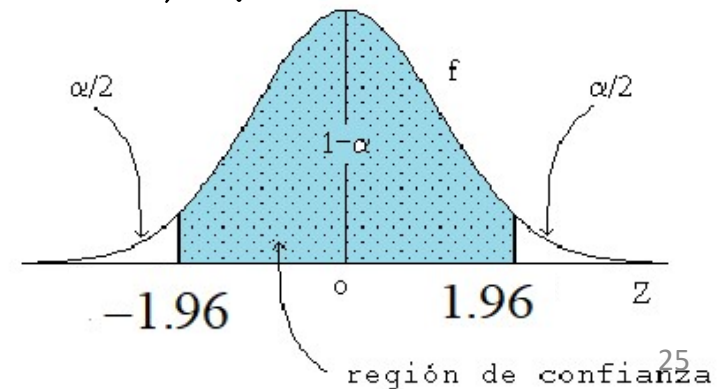
$n = 9$ tests, $\bar{X} = 187,9$ puntos , $\sigma = 32,4$ puntos (conocido)

Nivel de confianza = $1 - \alpha = 0.95$ entonces $\alpha = 0.05$ y $\alpha/2 = 0.025$

Intervalo a usar es el del caso I: $(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}})$

$$-z_0 = z_{\alpha/2} = z_{0.025} = -1.96$$

$$z_0 = z_{1-\alpha/2} = z_{0.975} = 1.96$$



$$(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}})$$

Reemplazando los datos en:

$$(187.9 - 1.96 * \frac{32.4}{\sqrt{9}}, 187.9 + 1.96 * \frac{32.4}{\sqrt{9}})$$

$$(187.9 - 21.16, 187.9 + 21.16)$$

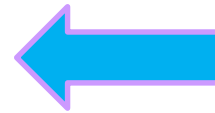
$$(166.732, 209.068) = \mathbf{(a, b)}$$

El puntaje promedio obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo oscila entre **166.732 puntos** y **209.068 puntos**, con un porcentaje de confianza del 95%.

c) A partir de estos resultados muestrales, un consultor estadístico construyó otro intervalo de confianza, para estimar la calificación media poblacional, con extremos 165,8 a 210,0 puntos. Determine el nivel de confianza que utilizó para calcular dicho intervalo.



$$(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}})$$



Se desconoce el nivel de confianza empleado.

Con lo cual la incognita es el valor de z_0 que utilizó para construir el intervalo

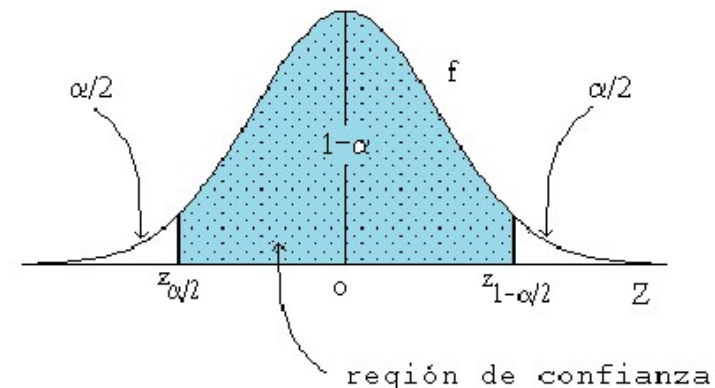
Amplitud de un intervalo

$$\begin{aligned} A &= \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}} - (\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}) = \\ &= 2 z_0 * \frac{\sigma}{\sqrt{n}} = 2 z_0 * \frac{32.4}{\sqrt{9}} \end{aligned}$$

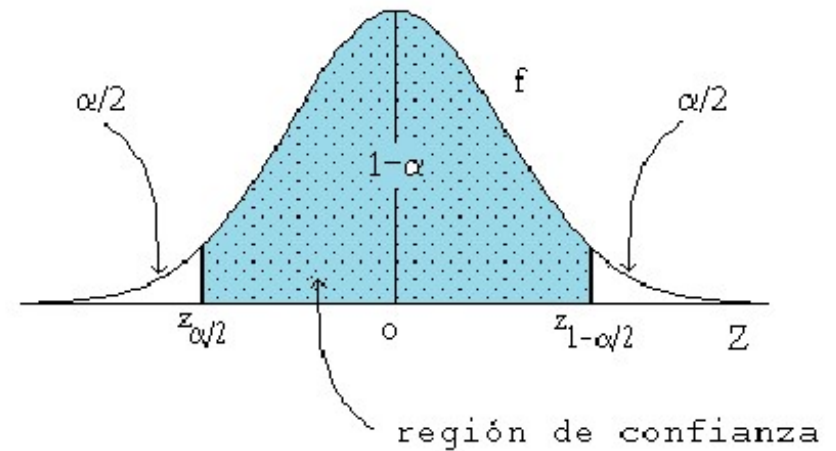
Luego

$$210 - 165.8 = 44.2 = A = 2 z_0 * \frac{32.4}{\sqrt{9}}$$

$$2.05 = \frac{44.2}{2 * 32.4} \sqrt{9} = z_0$$



$$2.05 = \frac{44.2}{2} \frac{\sqrt{9}}{32.4} = z_0$$



$$P(-2.05 < Z < 2.05) = 1 - \alpha$$

$$P(-2.05 < Z < 2.05) = P(Z < 2.05) - P(Z < -2.05) =$$

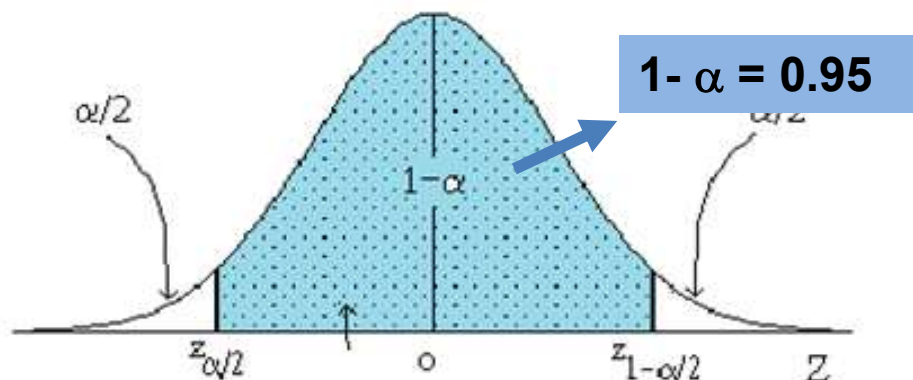
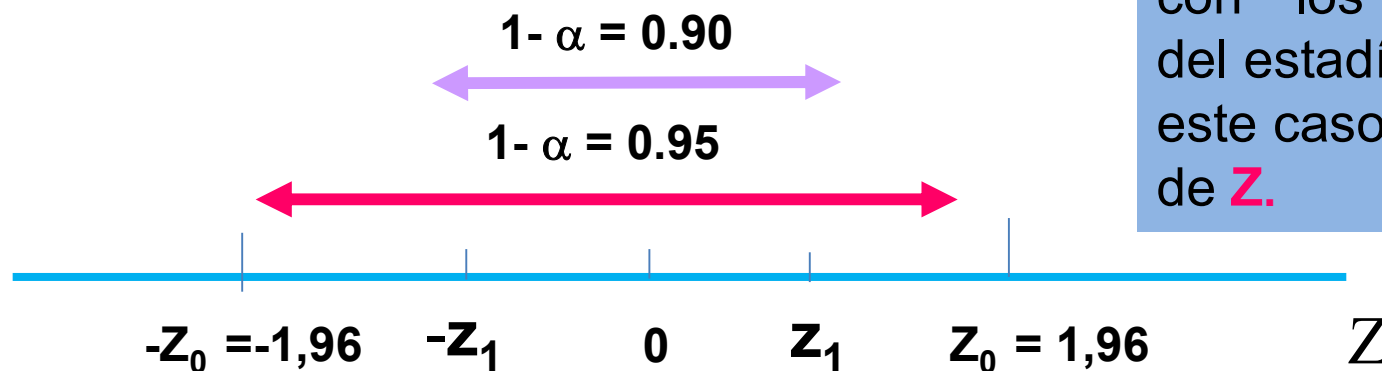
$$= 0.9798 - 0.0202 = 0.9596$$

El nivel de confianza utilizado por el consultor estadístico para construir el intervalo para estimar el puntaje promedio obtenido por un solicitante en el test de aptitud para acceder a un determinado puesto de trabajo es de $1 - \alpha = 0.96$ aprox.

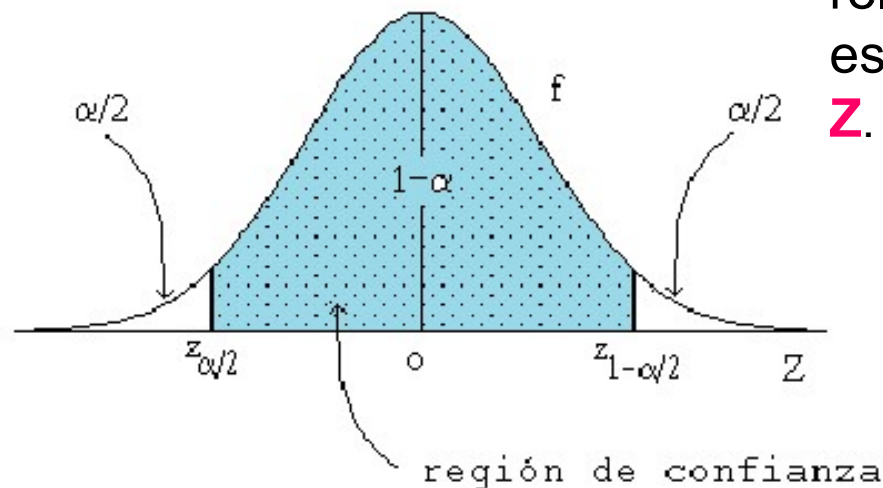
d) Sin realizar los cálculos, determinar si un intervalo de confianza del 90% para estimar la calificación media poblacional tendría mayor, menor o la misma longitud que el calculado en el inciso **b)**. ¿Cuál de los dos intervalos es más preciso?

$$(\bar{X} - z_0 * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_0 * \frac{\sigma}{\sqrt{n}})$$

El nivel de confianza **1 - α** está relacionado con los valores del estadístico, en este caso, valores de **Z**.



d) Sin realizar los cálculos, determinar si un intervalo de confianza del 90% para estimar la calificación media poblacional tendría mayor, menor o la misma longitud que el calculado en el inciso **b)**. ¿Cuál de los dos intervalos es más preciso?



El nivel de confianza $1 - \alpha$ está relacionado con los valores del estadístico, en este caso, valores de Z .

A mayor confianza, mayor amplitud del intervalo, con lo cual el intervalo de Confianza del inciso **b)** es más amplio que uno en el que se usa un nivel de **0.9**. Este último tiene menor longitud, y en consecuencia, es más preciso.