ESTIMADORES. INTERVALOS DE CONFIANZA

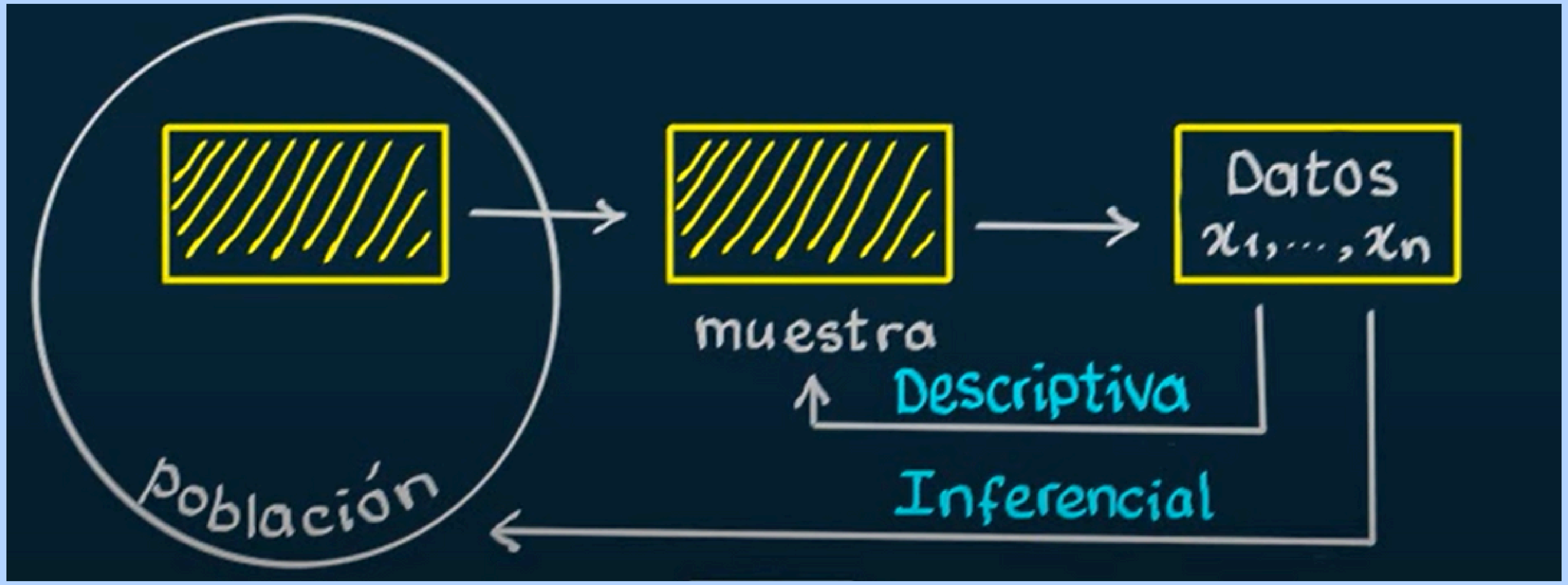
Florencia Gonzalez

INFERENCIA ESTADÍSTICA

¿Qué significa INFERIR?

Es sacar conclusiones a partir de un MUESTRA.

VAMOS A TRABAJAR EN ASEGURARNOS QUE ESTAMOS TOMANDO UNA MUESTRA ADECUADA Y REPRESENTATIVA PARA SACAR CONCLUSIONES



DIFERENCIA ENTRE PARÁMETROS Y ESTADÍSTICOS

PARÁMETROS

ESTUDIO SOBRE LA TOTALIDAD DE LOS ELEMENTOS DE UN **POBLACIÓN.**

ESTADÍSTICOS

ESTUDIO SOBRE UNA **MUESTRA** CONFIABLE Y REPRESENTATIVA**.**

Los estadísticos muestrales son cuantificadores de los datos de una muestra. Estos se calcula con el objetivo de inferir conclusiones soble la población en la que ha sido extrída la muestra.

El estadístico calculado de la muestra sirve para estimar el parámetro poblacional

Estadístico		Parámetro
Un **estadístico** es un valor que describe una característica de una **muestra**. El valor de un estadístico varía de una muestra a otra: **NO TIENE UN VALOR ÚNICO.**	Estimación →	Un **parámetro** es un valor que describe una característica de una **población**. Bajo el enfoque de la estadística clásica el valor de un parámetro poblacional es: **ÚNICO.**

$\bar{X}$ (media muestral) → μ (media poblacional)

S^2 (varianza muestral) → σ^2 (varianza)

LA MEDIA ARITMÉTICA COMO ESTIMADOR

Cuando la aplicamos a una muestra, se denomina **media muestral** y es uno de los **estadísticos muestrales**. Se trata de un estimador fiable, fácil de calcular y conocido por todos.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

En donde $\bar{X}$ es la media muestral que se desea obtener, n es la cantidad de elementos contenidos en la muestra y los x_i son cada uno de los valores de la muestra.

PROPIEDADES DE UN BUEN ESTIMADOR

PROPIEDADES DE UN BUEN ESTIMADOR

Denotaremos con $\hat{\theta}$ al estimador genérico de un determinado parámetro θ. Cuando se repiten las muestras para calcular el estimador, a cada uno de ellos se lo suele denominar $\hat{\theta}_n$, en donde n a menudo puede indicar la cantidad de elementos considerados en la muestra.

1. Un buen estimador es insesgado

El sesgo de un estimador es la diferencia que hay entre la esperanza matemática del estimador y el valor hallado para el mismo. Se considera que un buen estimador tiene sesgo nulo o casi nulo. Recordemos que la esperanza es el valor esperado para una determinada prueba. Simbólicamente resulta:

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

2. Un buen estimador es consistente

Se dice que un estimador es consistente cuando, a medida que crece el tamaño de la muestra, su valor se aproxima cada vez más al valor del parámetro considerado. Simbólicamente:

$$\lim_{n \to \infty} \hat{\theta} = \theta$$

3. Un buen estimador es eficiente

Un estimador es más eficiente, o más preciso, cuanto menor sea su varianza. Usualmente se utiliza esta característica en la comparación de dos o más estimadores, o sea:

$$\sigma^2(\theta_1) < \sigma^2(\theta_2) \Rightarrow \theta_1 \; es \; más \; eficiente \; que \; \theta_2$$

4. ***Un buen estimador es suficiente***

Un estimador es suficiente si para calcularlo se utiliza toda la información relevante suministrada por la muestra.

DISTRIBUCIONES EN MUESTREO

Una estadística es una variable aleatoria que depende básicamente de la muestra seleccionada, con lo cual debe tener una distribución de probabilidad, a la que llamaremos **distribución muestral**, la cual, a su vez, dependerá de:

- *El tamaño de la población*
- *El tamaño de la muestra*
- *El método utilizado para la selección de las muestras*

EL TEOREMA DEL LÍMITE CENTRAL

Sea $(x_1, x_2, \dots, x_n)$ una muestra de n variables aleatorias, independientes e idénticamente distribuidas de una distribución con media μ y varianza finita σ^2 finita y distinta de cero, y donde $\bar{X}$ es la media muestral, entonces la forma límite de la distribución de

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

cuando $n \to \infty$, es la distribución normal estándar.

El teorema central del límite (TCL) es una teoría estadística que establece que, dada una muestra aleatoria suficientemente grande de la población, la distribución de las medias muestrales seguirá una distribución normal.

- Si el tamaño de la muestra es suficientemente grande, la distribución de las medias muestrales seguirá aproximadamente una distribución normal. El TCL considera una muestra como grande cuando el tamaño de la misma es superior a 30. Por tanto, si la muestra es superior a 30, la media muestral tendrá una función de distribución próxima a una normal. Y esto se cumple independientemente de la forma de la distribución con la que estamos trabajando.
- La media poblacional y la media muestral serán iguales. Es decir, la media de la distribución de todas las medias muestrales será igual a la media del total de la población.
- La varianza de la distribución de las medias muestrales será σ^2/n. Que es la varianza de la población dividido entre el tamaño de la muestra.

ESTIMACIÓN PUNTUAL

- Una estimación es **puntual** cuando se usa un solo valor extraído de la muestra para estimar el parámetro desconocido de la población. Al valor usado se le llama estimador.

- La media de la población se puede estimar puntualmente mediante la media de la muestra:

$$\bar{x} = \mu$$

- La proporción de la población se puede estimar puntualmente mediante la proporción de la muestra:

$$\hat{p} = p$$

- La desviación típica de la población se puede estimar puntualmente mediante la desviación típica de la muestra, aunque hay mejores estimadores:

$$s = \sigma$$

ESTIMACIÓN POR INTERVALOS DE CONFIANZA

En este caso, en lugar de indicar simplemente un único valor como estimación del parámetro poblacional θ, lo que haremos es ofrecer un intervalo de valores en el que se tiene cierta probabilidad (confianza) de que se encuentre el verdadero valor de θ.

$$\hat{\theta} - \varepsilon < \theta < \hat{\theta} + \varepsilon$$

Intervalo de confianza: Es el intervalo de las estimaciones (probables) sobre el parámetro. **Límites de los intervalos de confianza** : Son los dos valores extremos del intervalo de confianza. **Amplitud del intervalo o margen de error...**

CÁLCULO DE INTERVALO DE CONFIANZA PARA LA MEDIA, CONOCIDA LA DESVIACIÓN TÍPICA DE LA POBLACIÓN EN UNA VARIABLE ALEATORIA NORMAL

- **Intervalo de confianza** es el intervalo que contiene al parámetro que se está estimando con un cierto nivel de confianza.
- **Nivel de confianza (1 - α)** , significa que el (1 - α) · 100% de los intervalos de confianza contienen el parámetro poblacional que se está estimando.

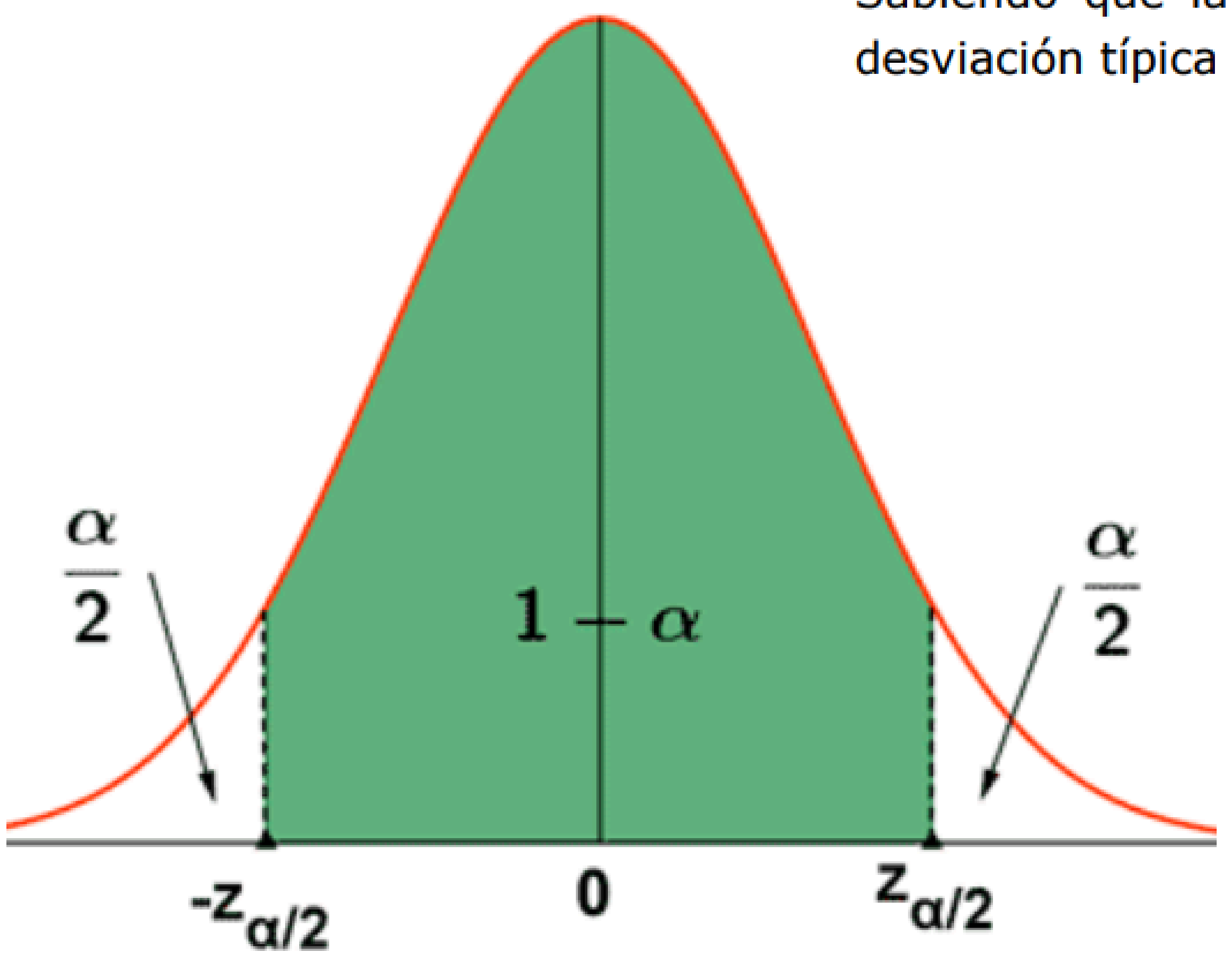
A cada nivel de confianza (N_c) le corresponde un **valor crítico $z_{\alpha/2}$** correspondiente a la distribución normal N(0, 1) y que cumple:

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha$$

A los extremos del intervalo de confianza se les llama **límites de confianza**.

En una distribución N(μ , σ) el intervalo correspondiente a una probabilidad p = 1 -α es:

$$\left(\mu - z_{\alpha/2} \cdot \sigma \ , \quad \mu + z_{\alpha/2} \cdot \sigma \right)$$



Sabiendo que la media muestral $\bar{x}$, sigue una distribución normal de media μ y desviación típica $\sigma/\sqrt{n}$, se calculan los extremos del intervalo como:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \quad \text{y} \quad \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

El **valor crítico $z_{\alpha/2}$** correspondiente a un nivel de confianza N_c en tanto por ciento, se calcula mediante la expresión:

$$P(Z \le z_{\alpha/2}) = \frac{1 + \frac{N_c}{100}}{2}$$

Y después buscando el resultado **dentro** de las tablas de la distribución normal.

Ejemplo 1:

Si fijamos el nivel de confianza N_c = 95 % hallar el valor crítico $z_{\alpha/2}$.

$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 0{,}95$$

$$P(Z \le z_{\alpha/2}) = \frac{1 + \frac{N_c}{100}}{2} = \frac{1 + \frac{95}{100}}{2} = \frac{1{,}95}{2} = 0{,}975$$

En las tablas de la distribución normal, el número 0,975 corresponde a : 1,96

Situación Problemática 1

Se realizan 10 tomas de los tiempos utilizados por un corredor para realizar una prueba de velocidad, con el propósito de estimar la media del tiempo empleado por el atleta.

El resultado de las medidas tomadas arroja una media de 21,5 segundos.

Sabiendo por datos anteriores que la desviación típica de esta variable para este corredor es de 0,4 segundos, obtener un intervalo de confianza con un 95% de confianza. ¿Cuántas pruebas habría que cronometrar para que el margen de error en la estimación de la media fuese inferior a tres décimas de segundo?

NOTA: Se establece que la variable que mide el tiempo del corredor sigue una distribución normal.

Solución:

Los datos con los que contamos son:

$\bar{x} = 21.5$ seg.

$n = 10$

$\sigma = 0.4$ seg.

$1 - α = 0.95$

Buscaremos entonces el valor de $z_{\alpha/2}$ para una distribución normal que tenga un área de probabilidad igual a $\alpha/2$ por su derecha, que en nuestro caso equivale a

(1-0.95) /2, es decir 0.025.

La función de distribución de probabilidad nos da el área de probabilidad acumulada, es decir a la izquierda, con lo cual tenemos que ver qué valor de z nos deja a la izquierda 0,975, y este valor se corresponde con un z=1.96.

Con estos datos podremos calculas los extremos del intervalo:

$$\left(21.5 - \frac{0.4}{\sqrt{10}}1.96\,, 21.5 + \frac{0.4}{\sqrt{10}}1.96\right)$$

$$(21.5 - 0.248\,, 21.5 + 0.248)$$

$$(21.252\,, 21.748)$$

LINK

OTRO EJEMPLO

Interpretación de los resultados:

Se estima que la media es de 21.5 seg., más o menos un error de 0.248 seg.