Tema 2. Conceptos básicos de inferencia estadística

Integrantes:

Amaro Cantoral Edgar
Durán López Teresita de Jesús
Cornejo Aguilar Clara Luz
Serralde Flores Andrea
Zarazúa Ramírez Johan Axel

Objetivo: El alumno describirá los conceptos más usuales de la inferencia estadística.

Contenido:

Conceptos importantes de la inferencia estadística: definiciones de parámetro, muestra aleatoria, estadístico y estimador de un parámetro.

Parámetro

Def. Un parámetro es una característica numérica de la distribución de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés. Regularmente se denota genéricamente por θ

Estadístico

Def. Un estadístico T es cualquier función de las variables aleatorias que se observaron en la muestra de manera que esta función no contiene cantidades desconocidas. Esto es: $T = h(X_1, X_2, \dots, X_n)$

Muestra aleatoria.

Def. Si las variables aleatorias X_1, X_2, …, X_n tienen la misma función (densidad) de probabilidad que la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces X_1, X_2, …, X_n forman un conjunto de n variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una muestra aleatoria de la población.

Obtención de una muestra aleatoria

1.- Se diseñó un experimento y se lleva a cabo para proporcionar la observación X_1 de la característica medible X. El experimento se repite bajo las mismas condiciones proporcionando el valor de X_2. El proceso continúa hasta tener n observación $X_1, X_2, \ldots, X_n$ de la característica X.

2.- Después de llevar a cabo una mezcla adecuada de los objetos de la población, se extrae uno y se observa la característica medible. Esta observación será X_1. El objeto se regresa a la población y está vuelve a mezclarse; después se extrae el segundo objeto. X_2 se constituye la segunda observación. El proceso se continúa de esta forma hasta que sean han extraído n objetos para tener una muestra de observaciones $X_1, X_2, \ldots, X_n$ de la característica X.

3.- Después de una mezcla adecuada de los objetos que constituyen la población, n de estos se seleccionan uno después de otro sin reemplazo. Este proceso proporciona una muestra de observaciones $X_1, X_2, \ldots, X_n$ de la característica X.

Distribución de muestreo

Def. La distribución de muestreo de un estadístico T es la distribución de probabilidad de T que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño n, provenientes de la población de interés.

Estimador.

Un estimados es una estadística T que identifica al mecanismo funcional por medio del cual, una vez que las observaciones en la muestra se realizan, se obtiene una estimación. Se utiliza para aproximar el valor del parámetro desconocido θ de la población. Se denota por $\hat{\theta}$

2.2 Teorema del límite central:

No importa el tipo de modelo de probabilidad a partir del cual se obtenga la muestra; mientras la media y la varianza existan, la distribución de muestreo de X se encontrará aproximada por N(m , σ/√n) para valores grandes de n. Lo anterior constituye uno de los más importantes teoremas en inferencia estadística, y se conoce como teorema central del límite.

probabilidad más importante.

TEOREMA

Teorema del límite central (TLC)

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces si n es suficientemente grande, $\bar{X}$ tiene aproximadamente una distribución normal con $\mu_{\bar{X}} = \mu$ y $\sigma^2_{\bar{X}} = \sigma^2/n$, y T_o también tiene aproximadamente una distribución normal con $\mu_{T_o} = n\mu$, $\sigma^2_{T_o} = n\sigma^2$. Mientras más grande es el valor de n, mejor es la aproximación.

2.3 Los conceptos y las definiciones de la distribución de la población o poblacional, distribución de la media y la varianza muestral y sus parámetros.

Sea X_1, X_2,….,X_n una muestra aleatoria que consiste de n variables aleatoria independientes normalmente distribuidas con medias $E(X_i)=\mu$ y varianzas$Var(Xi) = \sigma^2$, i = 1, 2,…., n. Entonces la distribución de la media muestral $\bar{\boldsymbol{X}}$ es normal com media μ y varianza σ^2/n.

2.4 La ley de los grandes números.

Es un teorema fundamental de la probabilidad que señala que si se realiza varias veces un mismo experimento, la frecuencia con la que se repita un determina un determinado proceso se acercara a una constante, dicha constante será la probabilidad de que el evento ocurra. Es decir, el promedio de los resultados obtenidos de un gran número de ensayos debe estar cerca del valor esperado.

La Ley (que en realidad son varias) describe el comportamiento de la media ($\bar{X}$) de una muestra aleatoria con $E[X_i] = \mu$, conforme aumenta el n, el tamaño de la muestra . $\bar{X}$ $converge\ a\ \mu$ casi con certeza (con probabilidad 1) esto es:

$$\boldsymbol{P(\lim_{n\to\infty} \bar{X} = \mu) = 1}$$

Esto es, cuando la muestra es grande la media muestral es la media poblacional

Ejercicios.

I. Una marca particular de jabón para lavadora de platos se vende en tres tamaños: 25 oz, 40 oz y 65 oz. El 20% de todos los compradores seleccionan la caja de 25 oz, 50% seleccionan una caja de 40 oz y el 30% restante seleccionan la caja de 65 oz. Sean $X1$ y $X2$ los tamaños de paquete seleccionados por dos compradores independientemente seleccionados.

a. Determine la distribución de muestreo de $\bar{X}$, calcule $E(\bar{X})$, y compare con μ

b. Determine la distribución de muestreo de la varianza muestral S^2, calcule $E(S^2,)$ y compare con σ^2.

II. Una compañía naviera menja contenedores en tres diferentes tamaños: (1) 27 pies^3 (3 x 3 x 3), (2) 125 pies^3 y (3) 512 pies^3. Sea X_i (i = 1, 2, 3) el número de contenedores de tipo i embarcados durante una semana dada. Con $\mu_i = E(X_i)$

y $\sigma^2_i = V(Xi)$, suponga que los valores medios y las desviaciones estándar son como sigue:

$$\mu_1 = 200 \qquad \mu_2 = 250 \qquad \mu_3 = 100$$
$$\sigma_1 = 10 \qquad \sigma_2 = 12 \qquad \sigma_3 = 8$$

a. Suponiendo que X_1, X_2, X_3 son independientes, calcule el valor esperado y la varianza del volumen total embarcado [*Sugerencia*: Volumen = $27X_1 + 125X_2 + 512X_3$.]

b. ¿Serían sus cálculos necesariamente correctos si las X_i no fueran independientes? Explique.

III. Suponga que la densidad de un sedimento de un espécimen seleccionado al azar de cierta región está normalmente distribuida con media de 2.65 y desviación estándar de 0. 85.

a) Si se selecciona una muestra aleatoria de 25 especímenes ¿cuál será la probabilidad de que la densidad del sedimento promedio muestral sea cuando mucho de 3? ¿Cuál sería la probabilidad de que se encuentre entre 2.65 y 3?

b) ¿Qué tan grande debe ser un tamaño de muestra para garantizar que la 1ª probabilidad en el inciso a sea por lo menos 0.99?

IV. La cantidad de una impureza particular en un lote de cierto producto químico es una variable aleatoria con valor medio de 4 g y desviación estándar de 1.5 g, si se preparan 50 lotes en forma independiente ¿Cuál es la probabilidad (aproximada) de que la cantidad promedio muestral de la impureza sea de 3.5 g a 3.8 g? Utilice TCL.

V. La resistencia a la ruptura de un remache tiene un valor medio de 10000 lb⁄pulg^2 y una desviación estándar de 500 lb⁄pulg^2.

¿Cuál es la probabilidad de que la resistencia a la ruptura media de una muestra aleatoria de 40 remaches sea de entre 9900 y 10200?

VI. Se sabe que la dureza Rockwell de "pernos" de un tipo tiene un valor medio de 50 y una desviación estándar de 1.2.

a) Si la distribución es normal, ¿cuál es la probabilidad de que la dureza media de una muestra aleatoria de 9 pernos sea por lo menos de 51?

b) ¿Cuál es la probabilidad (aproximada) de que la dureza media de una muestra aleatoria de 40 pernos sea por lo menos de 51?

VII. En la producción de cierto material para solar se sabe que la desviación estándar de la tensión de ruptura de este material es de 25 libras. ¿Cuál debe ser la tensión de ruptura promedio del proceso si, con base en una muestra aleatoria de 50 especímenes, la probabilidad de que la media muestral tenga un valor mayor de 250 libras es de 0.95?

VIII. La resistencia es una característica importante de los materiales utilizados en casas prefabricadas. Suponga que un fabricante de casas afirma la resistencia de los materiales que utiliza es tal que cuando se someten a esfuerzo severo, los anchos máximos de las grietas resultantes siguen una distribución normal con media μ=1.5 y varianza σ^2 =0.8

Cada uno de n = 11 elementos de placa prefabricados se sometieron a prueba de esfuerzo severo y se registró el ancho máximo (mm) de las grietas resultantes. Los datos proporcionados en la siguiente tabla aparecieron en el artículo ("Prefabricated Ferrocement Ribbed Elements for Low-Cost Housing", J. Ferrocement, 1984: 347-364).

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0.684	−0.9841	0.9685
2.540	0.8719	0.7602
0.924	−0.7441	0.5537
3.130	1.4619	2.1372
1.038	−0.6301	0.3970
0.598	−1.0701	1.1451
0.483	−1.1851	1.4045
3.520	1.8519	3.4295
1.285	−0.3831	0.1468
2.650	0.9819	0.9641
1.497	−0.1711	0.0293
$\sum x_i = 18.349$	$\sum(x_i - \bar{x}) = -0.0001$	$S_{xx} = \sum(x_i - \bar{x})^2 = 11.9359$
$\bar{x} = 18.349/11 = 1.6681$		

a) Calcular la varianza s^2 y la desviación estándar s de la muestra.

b) Calcular la media y la varianza de $\bar{X}$

c) Calcule la probabilidad de que $\bar{X}$ se mayor que el resultado observado en la muestra

d) Calcule la probabilidad de que *s*2 sea mayor que el valor que obtuvo en el inciso a)

IX. Encuentra la probabilidad de que una muestra aleatoria de 25 observaciones, de una población normal con varianza $\sigma^2 = 6$, tenga una varianza muestral S^2.

a. Mayor a 9.1

b. Entre 3.462 y 10.745

X. Las calificaciones de un examen de colocación que se aplicó a estudiantes de primer año de una universidad durante los últimos cinco años están distribuidas aproximadamente de forma normal con una media $\mu = 74$ y una varianza $\sigma^2 = 8$. ¿Consideraría aún que $\sigma^2 = 8$ es un valor válido de la varianza si una muestra aleatoria de 20 estudiantes, quienes realizan tal examen de colocación este año, obtienen un valor de $s^2 = 20$?

Punto adicional en la serie

En la simulación del ejercicio 10 de la guía tema 1 calcule $\frac{x}{n}$ $para\ n = 50, n = 100\ y\ n = 500$. Observe que en la medida que $x \to \infty$ $\frac{x}{n} \to 0.75$

Notas Técnicas

Nota Técnica 1

Inferencia estadística

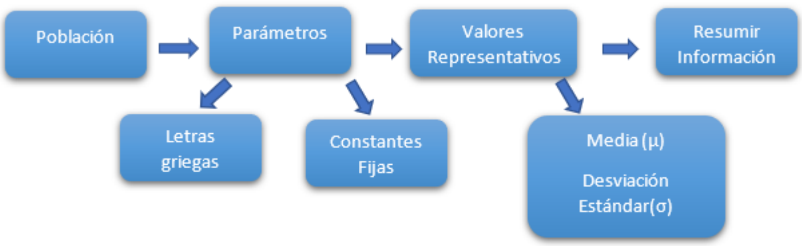
Conjunto de métodos que permiten conocer el comportamiento de una determinada población a través de una muestra. La inferencia estadística estudia entonces, como sacar conclusiones sobre los parámetros de una población de datos.

La inferencia estadística resuelve dos tipos de problemas:

1. Estimación de una o varias características de la población
2. Prueba de hipótesis o suposiciones acerca las características de la población

Parámetro y estadístico.

Parámetro: Medida descriptiva de la población.

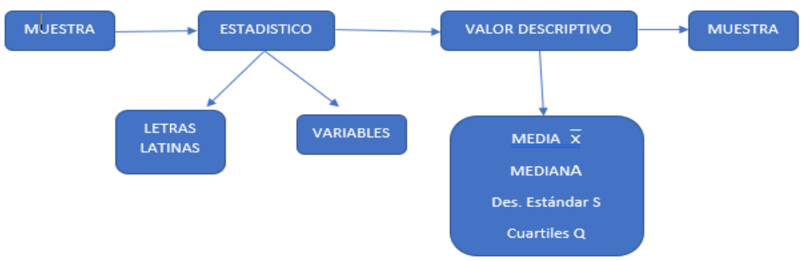


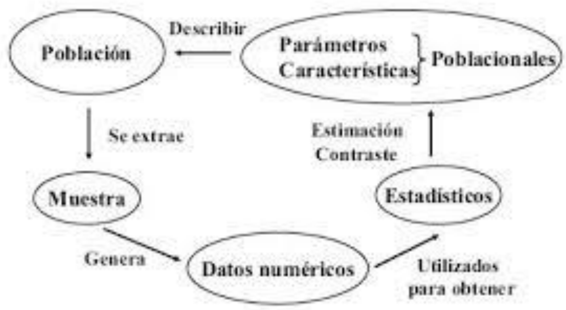
Ejemplos de parámetros son el promedio, la mediana, la desviación estándar o los cuartiles.

Los parámetros estadísticos se clasifican según la información que resumen. Los dos tipos más comunes de parámetros estadísticos son:

- De tendencia central: Son las medidas que informan acerca de la mayor o menor agrupación o concentración de los datos entorno a la media: **media**, **moda** y **mediana**.
- De dispersión: Son los parámetros que indican la mayor o menor concentración de los datos alrededor de los parámetros de centralización: **desviación respecto de la media**, **desviación media**, **varianza** y **desviación estándar**.

Estadístico: Medida descriptiva de la muestra.





consultar material de apoyo:

- Diferencia entre parámetro y estadístico.
  https://www.youtube.com/watch?v=D6d8MD6RDyU
- Ejercicios de parámetros estadísticos
  https://www.youtube.com/watch?v=rRc-7KE7VUQ

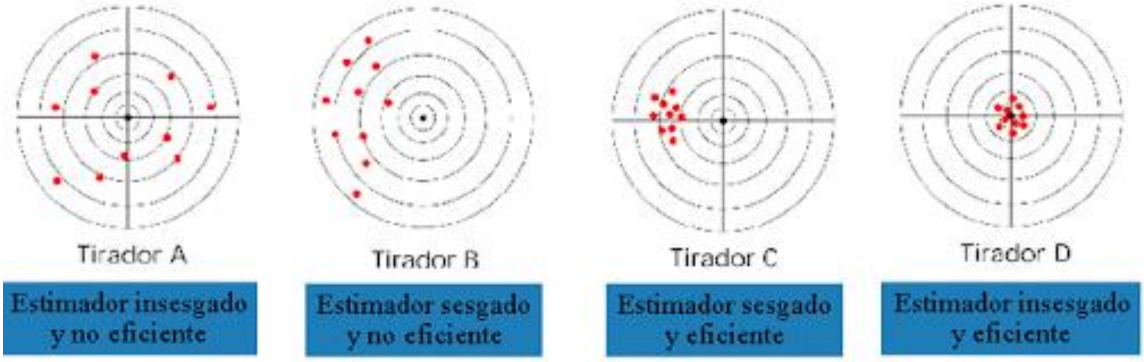
Estimador

Estadístico usado para estimar un parámetro desconocido de la población.Es obtienen mediante una muestra aleatoria esto significa que el valor del estimador cambia dependiendo de la muestra. Es decir, son aleatorios y en consecuencia tienen una distribución.

Ejemplo:

Propiedades de Eficiencia y sesgo de un estimador

Cuatro tiradores han efectuado 10 disparos sobre una diana. Si traducimos cada disparo en una estimación, efectuada por un determinado estimador,sobre una muestra, podemos interpretar las propiedades de los estimadores de la siguiente forma:



consultar material de apoyo:

- Propiedades de los estimadores : https://www.youtube.com/watch?v=Hd6rmhof_Iw

Muestra

Tipos de muestra:

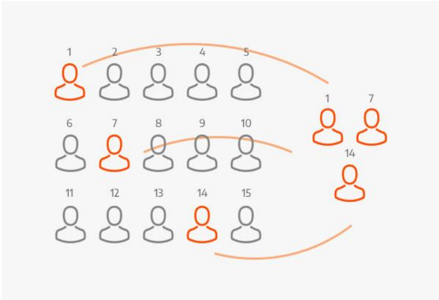
Muestra aleatoria

Conjunto de tamaño "n" formado por variables aleatorias (independientes e idénticamente distribuidas).

Muestra aleatoria simple

Conjunto de tamaño "n" formado por variables aleatorias donde cada variable tiene la misma posibilidad de ser elegida y por ende cada conjunto de tamaño “n” tiene la misma posibilidad de ser seleccionado como muestra.

La muestra debe ser probabilística, esto quiere decir que debe ser seleccionada al azar; pero una buena muestra además de ser aleatoria también debe ser representativa de la población de donde proviene.



El muestreo estratificado
implica seleccionar muestras independientes de un número de subpoblaciones, grupo o estratos dentro de la población. Por ejemplo, si queremos analizar los datos de unas elecciones por género o por grupo de edad, deberemos cerciorarnos de obtener muestras representativas de todas las subpoblaciones.

Muestreo estadístico
Es un método de investigación por el cual observamos solamente una parte de la población llamada muestra para obtener información y conocimiento de la población total.
Como parte de este método de investigación tenemos la inferencia estadística.

consultar material de apoyo:

- Técnicas para generar una muestra aleatoria simple ***https://es.khanacademy.org/math/ap-statistics/gathering-data-ap/sampling-methods/v/techniques-for-generating-a-simple-random-sample***
- ***https://es.khanacademy.org/math/ap-statistics/gathering-data-ap#sampling-methods***

Nota Técnica 2

TLC

El Teorema de Límite Central es un teorema fundamental de probabilidad y estadística que nos permite describir la media de una muestra aleatoria sin importar que tipo de modelo de probabilidad se trabaje para obtener aquella muestra.

Ejemplo:

Una empresa de mensajería que opera en la ciudad tarda una media de 35 minutos en llevar un paquete, con una desviación típica de 8 minutos. Supongamos que durante el día de hoy han repartido 200 paquetes. a) ¿Cuál es la probabilidad de que la media de los tiempos de entrega de hoy esté entre 30 y 35 minutos?. b) ¿Cuál es la probabilidad de que, en total, para los doscientos paquetes hayan estado más de 115 horas?.

Datos:

Variable X = “Tiempo de entrega del paquete”
Media: 35 minutos
Desviación típica: 8
n = 200 paquetes

a) Por el teorema del límite central sabemos que la media muestral se comporta como una normal de esperanza 35 y desviación típica:

$8 /\sqrt{200} = 0,566$

Si utilizamos esta aproximación, ya podemos calcular:

$P(30 \leq X \leq 35) = P((30-35)/0.566 \leq (X-35)/0.566 \leq (35-35)/0.566)$

$P((30-35)/0.566 \leq (X-35)/0.566 \leq (35-35)/0.566)= P(-8.83 \leq Z \leq 0)$

$P(-8.83 \leq Z \leq 0) = P(Z \leq 8.83)$

$P(Z \leq 8.83)= 0.5-0 = 0.5$

donde Z es una normal (0,1). Es decir, tenemos una probabilidad aproximada del ***0,4616*** de que la media del tiempo de entrega de hoy haya estado entre 30 y 35 minutos.

b) Para contestar ésta pregunta debemos observar que 115 horas por 60 minutos nos dan 6.900 minutos. Se nos pide que calculemos la probabilidad siguiente:

$P (Z>6900/20)=P(X>34.5)$

y como que sabemos que la media se distribuye aproximadamente como una normal de media 35 y desviación típica 0,566 , esta probabilidad se puede aproximar por la probabilidad de una distribución normal estándar Z:

P (Z>(34.5-35)/2000,566)=P(Z>0.88)=1-P(Z<0.88)=1-0.1894=***0.8194***

Tenemos una probabilidad de ***0.8194,*** de que, en total, para los doscientos paquetes hayan estado más de 115 horas.

consultar material de apoyo:

- Teorema del Límite central
  https://es.khanacademy.org/math/statistics-probability/sampling-distributions-library/sample-means/v/central-limit-theorem

- Ejemplos del Teorema de Límite Central

  https://www.youtube.com/watch?v=LY4v1AsTQGw

  https://www.youtube.com/watch?v=-IgvcerAu0s

Nota Técnica 3

Distribución poblacional.

Distribución de probabilidad de una población (Conocida o Desconocida).
Se puede caracterizar por:

- Medidas de tendencia central
- Medidas de dispersión

Distribución muestral.

Distribución de probabilidad de todas las muestras posibles que pueden ser tomadas de la población.
Se caracteriza por:

- *Medidas de tendencia central*
- *Medidas de dispersión (variabilidad)*
- *Función de probabilidad*

Propiedades útiles para el trabajo estadístico:

- *Aproximación a una distribución normal*
- *La media de la distribución es igual (o casi igual) a la media de la población.*
- *La dispersión (desviación estándar) es menor a la de la población.*

Distribución de la media muestral.

Distribución de probabilidad de todos los valores posibles de la media muestral que pueden ocurrir cuando se toma una muestra (de tamaño n) de la población, con media μ y varianza σ^2.

Casos según la selección de la muestra:

A. Población finita con reposición o infinita sin reposición.

$\overline{X} \longrightarrow$ Media muestral

$E[\overline{X}] = \mu$ $\qquad$ $Var[\overline{X}] = \sigma^2/n$

Para n lo suficientemente grande $(n \geq 30)$

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

★ $\overline{X} \longrightarrow N(\mu, \sigma^2/n)$ para $n \geq 30$

★ Si $X \sim N(\mu, \sigma^2)$ entonces $\overline{X} \sim N(\mu, \sigma^2/n)$ para $n > 2$

B. Población finita sin reemplazo.

$\overline{X} \longrightarrow$ Media muestral

$\mu_{\overline{X}} = \mu$ $\qquad$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

$\left(\frac{N-n}{N-1}\right)$ — factor de correccion (fcpf)

Cuando $N >> n$, $n/N \leq 0.05$ el fcpf ~ 1

$$Z = \frac{\overline{X} - \mu}{(\sigma/\sqrt{n})\left(\sqrt{\frac{N-n}{N-1}}\right)}$$

1. Distribución muestral de con población normal y σ^2 conocida

.

$$E[\bar{x}] = \mu_{\bar{x}} = \mu \qquad Var[\bar{x}] = \sigma^2_{\bar{x}} = \sigma^2/n$$

$$\bar{x} \longrightarrow N(\mu, \sigma/\sqrt{n})$$

por lo tanto:

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

2. Distribución muestral de con población desconocida y σ^2 conocida.

$$E[\bar{x}] = \mu_{\bar{x}} = \mu \qquad Var[\bar{x}] = \sigma^2_{\bar{x}} = \sigma^2/n \left(\frac{N-n}{N-1}\right)$$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1) \quad \text{para } n > 30$$

error estandar: $\sigma_{\bar{x}} = \left(\sigma/\sqrt{n}\right)\left(\sqrt{\frac{N-n}{N-1}}\right)$

Si $N/n \leq 0.05 \longrightarrow \sqrt{\frac{N-n}{N-1}} = 1$

3. Distribución muestral de con σ^2 desconocida.

1) Si la poblacion no es normal y n es pequeña, no se puede resolver

2) Si la poblacion no es normal y n es grande $(n \geq 30)$, sigue una distribucion t-student con n-1 grados de libertad

$$T = \frac{\bar{x} - \mu_{\bar{x}}}{s/\sqrt{n}} \longrightarrow t_{n-1}$$

3) Si la poblacion es normal y $n < 30$, $\bar{x}$ sigue t-student

$$T = \frac{\bar{x} - \mu_{\bar{x}}}{s/\sqrt{n}} \longrightarrow t_{n-1}$$

Si $n \geq 30$ entonces

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{s/\sqrt{n}} \longrightarrow N(0.1)$$

Distribución de la varianza muestral.

Cuando la población tenga una distribución normal y μ sea desconocida.

Estadístico $\chi^2 = \frac{n\sigma^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \longrightarrow$ Chi Cuadrado con n-1 grados de libertad χ^2_{n-1}

n = tamaño de la muestra

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

σ^2 = Varianza poblacional

Ejemplo (distribución de la media muestral):

Supóngase que el número de barriles de petróleo crudo que produce un pozo diariamente es una variable aleatoria con una distribución no especificada. Si se observa la producción en 64 días, seleccionados en forma aleatoria, y si se sabe que la desviación del número de barriles por día es $\sigma = 16$, determine la probabilidad de que la media muestral se encuentre a no más de cuatro barriles del verdadero valor de la producción por día.

$$P(\bar{X} < 4) = ?$$

Puesto que n es lo suficientemente grande, la distribución $\bar{X}$ es en forma aproximada, normal con media μ y desviación estándar $\sigma/\sqrt{n} = 16/\sqrt{64} = 2$. En forma equivalente, la distribución de **Z =** $(\boldsymbol{\bar{X}} - \boldsymbol{\mu})/\mathbf{2}$es, aproximadamente, N(0,1).

De acuerdo con lo anterior, la probabilidad deseada es:

$$P(|\bar{X} - \mu| < 4) = P(\mu - 4 < \bar{X} < \mu + 4)$$

$$= P[(\mu - 4 - \mu)/2 < Z < (\mu + 4 - \mu)/2]$$

$$= P(-2 < Z < 2)$$

De acuerdo al teorema del límite central:

$$P(a < x \leq b) = F(b) - F(a)$$

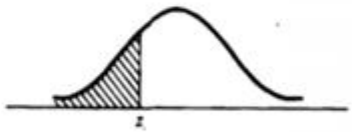
Sustituyendo:

$$P(-2 < Z < 2) = F(2) - F(-2)$$

Con ayuda de las tablas que nos muestran los valores de la función de distribución acumulativa normal estándar, buscaremos los valores:

TABLA D Valores de la función de distribución acumulativa normal estándar

$$P(Z \leq z) = F(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)dt$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183

Para F(2):

Ubicamos el valor de z = 2 en la primer columna, ya que no tiene más decimales. Podemos que observar que el valor de F(2) es 0.9772. De igual manera buscamos el valor de z = -2 y obtendremos que F(-2) es 0.0228.

z	.00
1.0	0.8413
1.1	0.8643
1.2	0.8849
1.3	0.9032
1.4	0.9192
1.5	0.9332
1.6	0.9452
1.7	0.9554
1.8	0.9641
1.9	0.9713
2.0	0.9772
2.1	0.9821
2.2	0.9861
2.3	0.9893
2.4	0.9918
2.5	0.9938
2.6	0.9953
2.7	0.9965
2.8	0.9974
2.9	0.9981

z	.00
−3.5	0.0002
−3.4	0.0003
−3.3	0.0005
−3.2	0.0007
−3.1	0.0010
−3.0	0.0013
−2.9	0.0019
−2.8	0.0026
−2.7	0.0035
−2.6	0.0047
−2.5	0.0062
−2.4	0.0082
−2.3	0.0107
−2.2	0.0139
−2.1	0.0179
−2.0	0.0228

Realizando la operación obtendremos que:

$$P(-2 < Z < 2) = 0.9772 - 0.0228 = 0.9544$$

Por lo tanto, la probabilidad de que la media muestral se encuentre a no más de cuatro barriles del verdadero valor de la producción por día es de 0.9544. Expresando el resultado de la siguiente manera:

$$P(\bar{X} < 4) = 0.9544$$

Material de apoyo:

- Las tablas de los valores de la función de distribución acumulativa normal estándar se encuentran en el libro Canavos, páginas de la 616 a 618. ***https://gsosa61.files.wordpress.com/2008/03/10-canavos-g-probabilidad-y-estadistica-aplicaciones-y-metodos.pdf***
- Distribución muestral de la media muestral
  Parte 1:
  https://es.khanacademy.org/math/statistics-probability/sampling-distributions-library/sample-means/v/sampling-distribution-of-the-sample-mean
  Parte 2:
  https://es.khanacademy.org/math/statistics-probability/sampling-distributions-library/sample-means/v/sampling-distribution-of-the-sample-mean-2
- Simulación que muestra el sesgo en la varianza muestral

https://es.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/more-on-standard-deviation/v/simulation-showing-bias-in-sample-variance

Nota Técnica 4

Las leyes de los grandes números explican por qué el promedio de una muestra al azar de una población de gran tamaño tenderá a estar cerca de la media de la población completa.

> “grandes números de individuos, actuando independientemente en un sistema, producen regularidades que no dependen de su coordinación mutua, de manera que es posible razonar sobre la colectividad sin ningún conocimiento detallado de los individuos”
>
> ‘Simeon Denis Poison’

Supongamos que lanzamos al aire una moneda de cinco francos, y observamos que, en 2.000 tiradas, la moneda sale cara 1.100 veces. Entendemos que hay una frecuencia o probabilidad constante de que la moneda salga cara, esto es, 11/20. Esta constante es la consecuencia de una causa común, de la manera en que está hecha la moneda y de la manera de arrojarla. Pero supongamos ahora que tiramos 2.000 monedas diferentes y obtenemos 1.100 caras. No podemos imaginar que las monedas tengan constituciones idénticas. Las causas y, por lo tanto, las probabilidades de salir cara, variarán de un caso a otro.

En resumen: “No se puede predecir el comportamiento individual, pero si el comportamiento promedio”.

Bibliografía:

- PROPIEDADES DE ESTIMADORES - Estadística Descriptiva en Edu.. (s.f.). Recuperado 1 abril, 2020, de https://sites.google.com/site/estadisticadescriptivaenedu/home/estimacion-porintervalos-1/propiedades-de-estimadores
- Estadística inferencial. (2015, 29 junio). Recuperado 1 abril, 2020, de https://www.slideshare.net/rbarriosm/4-estadistica-inferencial
- 403 Forbidden. (s.f.). Recuperado 1 abril, 2020, de http://www.estadistica.net/Aeronautica2016/estimadores.pdf
- Alejandro Quintela del Río (2019, 4 septiembre). 6.6 Las leyes de los grandes números | Estadística Básica Edulcorada. Recuperado 1 abril, 2020, de https://bookdown.org/aquintela/EBE/lasleyes-de-los-grandes-numeros.html
- DISTRIBUCIONES MUESTRALES PARA UNA POBLACIÓN NORMAL. (s.f.). Recuperado 1 abril, 2020, de https://www.uv.es/ceaces/tex1t/3%20infemues/dnormal.htm
- DistribuciÃ³n de la varianza muestral. (s.f.). Recuperado 1 abril, 2020, de https://prezi.com/unsorljfmjar/distribucion-de-la-varianza-muestral/
- El muestreo: qué es y por qué funciona. (s.f.). Recuperado 1 abril, 2020, de https://www.netquest.com/blog/es/blog/es/muestreo-que-es-porque-funciona
- El teorema del límite central: las medias de muestras grandes y aleatorias son aproximadamente normales - Minitab. (s.f.). Recuperado 1 abril, 2020, de https://support.minitab.com/esmx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/dataconcepts/about-the-central-limit-theorem/
- Estadística inferencial: estimación de parámetros y contrastes de hipótesis. (s.f.). Recuperado 1 abril, 2020, de http://www.hrc.es/bioest/Introducion.html
- Parámetros estadísticos. (2019, 29 noviembre). Recuperado 1 abril, 2020, de https://www.superprof.es/apuntes/escolar/matematicas/estadistica/descriptiva/parametro s-estadisticos.html
- POBLACIÓN, PARÁMETRO, MUESTRA, ESTADÍSTICO. (s.f.). Recuperado 1 abril, 2020, de http://desirestadisticasbasicas.blogspot.com/2010/07/poblacion-parametro-muestraestadistico.html
- Teorema central del limite. (2019a, 17 septiembre). Recuperado 1 abril, 2020, de https://www.superprof.es/apuntes/escolar/matematicas/estadistica/inferencia/teoremacentral-del-limite.html

- Universidad Nacional Mayor de San Marcos (s.f.). Distribuciones muestrales. Recuperado 1 abril, 2020, de https://es.slideshare.net/williamleon20/distribuciones-muestrales-49583318
- ¿Qué son parámetros, estimaciones de parámetros y distribuciones de muestreo? - Minitab. (s.f.). Recuperado 1 abril, 2020, de https://support.minitab.com/es-mx/minitab/18/help-and-howto/statistics/basic-statistics/supporting-topics/data-concepts/what-are-parametersparameter-estimates-and-sampling-distributions/