Agenda
Viernes 04 Jul 2025

- Estimación parámetros en MAS y M Sistemático

Sábado 05 Jul 2025

- Muestreo Estratificado (ME)
- Estimación parámetros en ME

Muestreo Probabilístico

Muestreo Aleatorio Simple (MAS)

- **Simple Random Sampling (SRS)**

Andy Domínguez

adominguez@utb.edu.co

Generalidades de un procedimento de muestreo

Un diseño muestral es un procedimiento mediante el cual se seleccionan las muestras, de modo que cada una tenga una determinada probabilidad de ser elegida.

Sea M un Diseño muestral aleatorio empleado y S el espacio muestral definido sobre M., entonces una función de probilidad P tal que :

- $P(S_i) \geq 0, \forall_i$
- $\sum_i P(S_i) = 1$

Generalidades de un procedimento de muestreo

el MAS es la forma mas básica de muestreo probabilistico y es el que provee las bases teóricas para las formas mas complejas.

Formas de MAS

Con reemplazamiento: cada unidad muestreada tiene probabilidad de $1/N$ cada vez, pero en general si se selecciona por ejemplo la misma persona para tomar un dato (Ej: Estatura) no será adecuado o útil.

Sin reemplazamiento: una muestra de tamaño n es selecionada de todos los posibles conjuntos de tamaño n de la población, habiendo $\binom{N}{n}$ posibles muestras con igual probabilidad de ser seleccionadas.

Simple Random Sampling (Muestreo Aleatorio Simple)

Definición

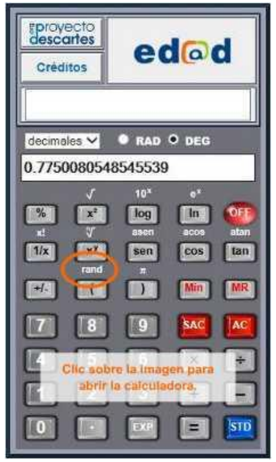
Al seleccionar una muestra de **n** mediciones de una población finita de **N** mediciones, si el muestreo se lleva a cabo de forma que todas las muestras posibles de tamaño n tengan la misma probabilidad de ser seleccionadas, el muestreo se llama aleatorio y el resultado es una muestra aleatoria simple

¿Cómo seleccionar una verdadera muestra aleatoria?

- Tabla de números aleatorios
- Calculadora - tecla RND.
- **Software Especializado**
    1. Python: random.sample(poblacion, n)
    2. R: sample(poblacion, n)
    3. Excel: =ALEATORIO.ENTRE(inf, sup)
    4. SPSS, STATA, Minitab, etc.

TABLA DE NÚMEROS ALEATORIOS

Fila					
1	7 9 9 3 0	3 7 3 4 1	8 6 8 2 2	1 4 4 6 3	3 6 0 1 2
2	1 4 1 7 6	6 6 0 4 3	1 6 2 5 2	3 9 8 4 5	1 7 3 2 5
3	7 1 8 7 4	2 0 5 9 9	2 8 6 8 9	0 1 9 7 9	0 9 5 0 9
4	4 9 1 1 3	0 9 0 6 4	3 7 6 1 1	7 3 8 3 5	2 8 2 6 8
5	8 7 9 2 4	9 6 1 3 6	3 0 7 5 8	1 6 1 5 0	5 3 4 9 0
6	4 3 4 6 2	9 3 3 0 5	0 9 2 5 1	6 0 3 1 0	6 0 3 5 9
7	8 5 2 5 1	5 2 0 3 7	7 6 7 8 3	1 0 0 5 6	2 9 8 2 5
8	1 6 8 9 0	6 1 5 4 7	9 4 8 3 0	1 8 1 8 4	2 6 6 6 1
9	7 6 8 5 6	2 4 8 3 0	8 1 7 8 3	4 7 4 4 7	8 1 7 8 4
10	6 3 0 3 2	7 2 4 8 3	4 7 4 5 3	8 4 6 7 2	3 5 1 3 5
11	7 0 8 0 9	5 1 8 4 3	1 6 1 5 0	5 6 9 7 9	4 7 0 5 5
12	6 1 1 4 7	3 3 5 4 3	0 2 0 4 9	6 1 1 2 1	5 2 1 9 8
13	5 7 7 5 4	6 0 7 8 1	9 1 8 8 8	7 3 9 2 6	7 9 6 3 4
14	9 7 8 8 3	0 6 7 3 3	0 2 7 3 4	6 7 3 3 6	1 4 4 9 8
15	8 7 1 9 9	2 7 4 9 9	7 9 1 5 4	2 8 1 3 8	0 5 2 5 1
16	1 6 7 9 6	9 7 1 1 2	6 3 7 4 1	7 1 8 1 8	4 1 5 4 3
17	8 1 3 1 9	6 3 2 8 6	2 3 3 5 2	8 0 4 0 8	9 6 2 3 7
18	9 3 8 6 9	7 6 8 6 3	6 6 3 4 3	7 6 2 1 8	9 3 8 8 6
19	9 0 9 5 2	6 4 1 2 8	0 2 6 2 2	8 7 6 5 1	3 7 6 8 0
20	4 6 3 5 6	7 1 2 2 8	4 7 2 4 0	5 1 7 1 5	7 5 1 8 2
21	8 2 7 0 7	3 1 6 3 8	7 3 2 7 2	5 1 9 2 5	8 1 5 3 9
22	9 4 9 9 6	0 0 3 7 6	7 8 8 7 6	6 7 7 5 2	6 7 0 6 8
23	1 2 3 0 7	1 0 7 8 3	2 2 8 4 6	4 8 7 5 2	1 0 8 6 7
24	2 6 6 4 9	5 6 0 1 6	1 6 7 6 8	2 4 8 4 3	9 2 9 3 0
25	3 1 8 4 2	3 6 2 3 7	3 5 0 7 1	0 5 0 9 6	2 0 7 7 8



Generalidades de un procedimiento de muestreo

Un diseño muestral es un procedimiento mediante el cual se seleccionan las muestras, de modo que cada una tenga una determinada probabilidad de ser elegida.

Sea M un Diseño muestral aleatorio empleado y S el espacio muestral definido sobre M., entonces una función de probilidad P tal que :

- $P(S_i) \geq 0, \forall_i$
- $\sum_i P(S_i) = 1$

Propiedades MAS

- Cada elemento de la población tiene igual probabilidad de ser seleccionado.
- El proceso de selección debe seguir un esquema de aleatorización de la lista.
- Ley de los grandes números

Para Recordar

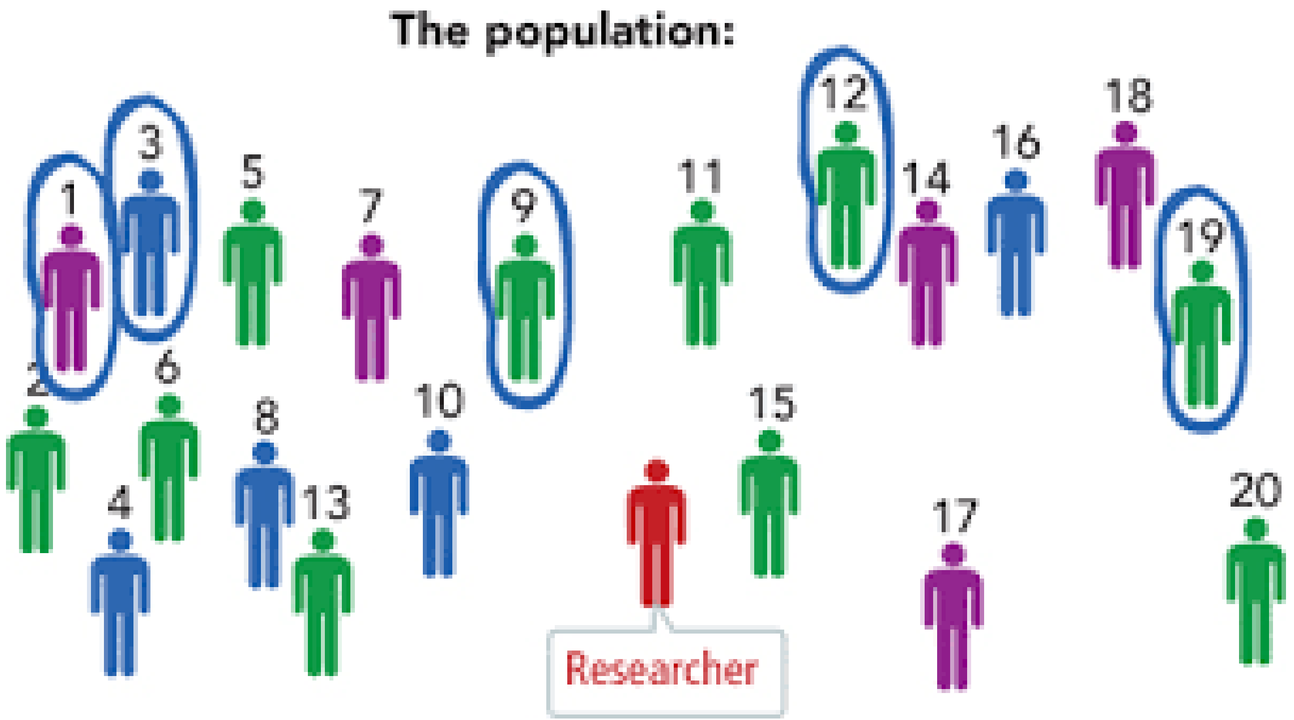
¿Cual es la probablidad de seleccionar una muestra individual de tamaño n en una población de tamaño N?

¿Cuántas muestras de tamaño n pueden seleccionarse de una población de tamaño N?

$$\frac{N!}{n!(N-n)!}$$

$$P(S) = \frac{1}{\binom{N}{n}} = \frac{n!\,(N-n)!}{N!}$$

Propiedades MAS



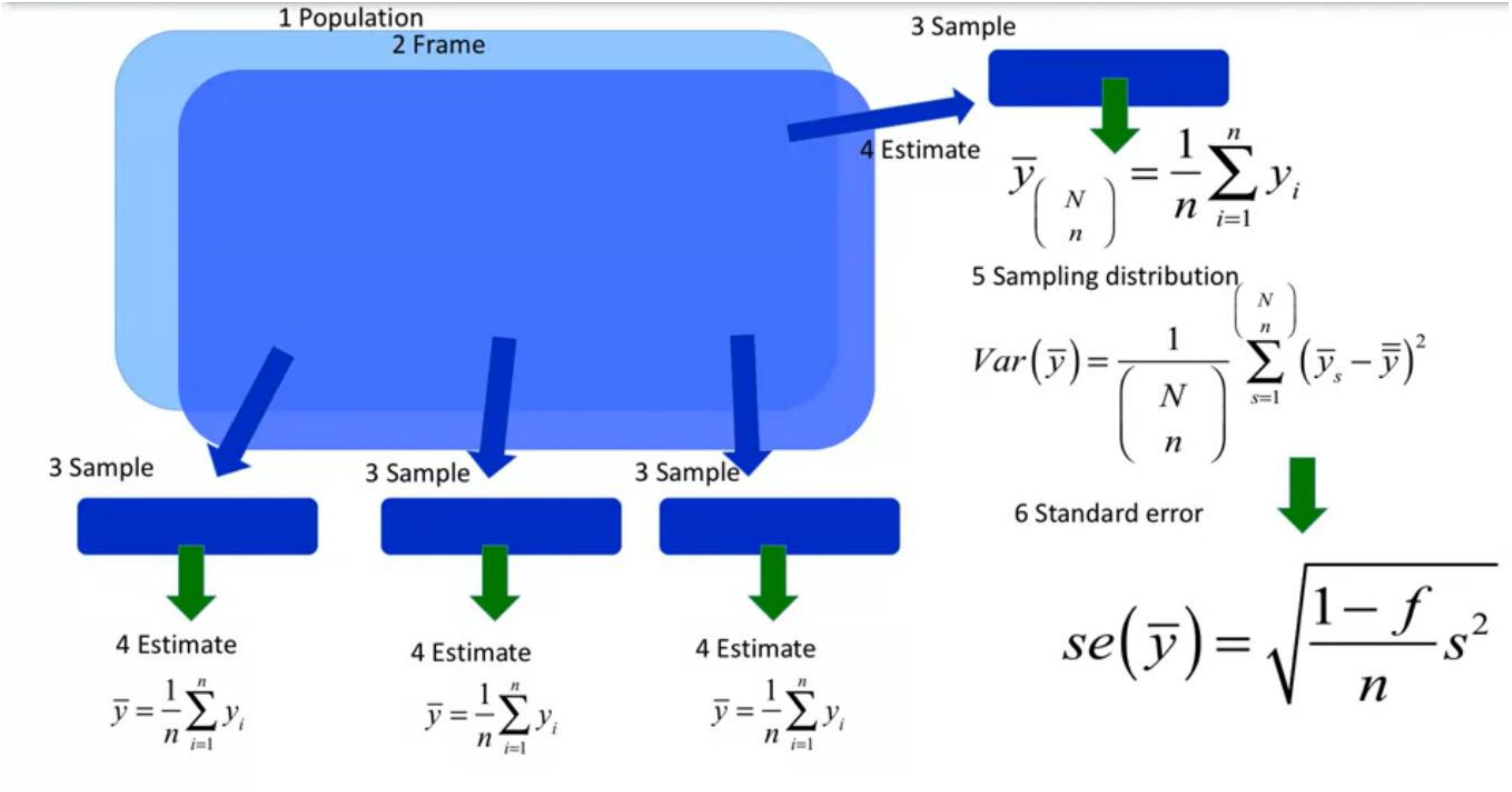
Ventajas del MAS

- Fácil de entender y aplicar.
- Base teórica sólida.
- Permite aplicar inferencia estadística con métodos clásicos.

Limitaciones

- Requiere listado completo y accesible de la población.
- No garantiza representatividad si la población es muy heterogénea (por eso existen otros métodos como estratificado o por conglomerados).

Muestreo Aleatorio Simple - Diseño



Estimadores del MAS

Estimadores de la Media

$$\bar{x} = \frac{\sum x_i}{n}$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Si se conoce la varianza poblacional:

$$\sigma_{\bar{x}} = \frac{(N-n)}{N} \frac{\sigma}{\sqrt{n}}$$

Si se desconoce la varianza poblacional(casi siempre):

$$\sigma_{\bar{x}} = \frac{(N-n)}{N} \frac{S}{\sqrt{n}}$$

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

***factor de corrección** para poblaciones finitas (N – n)/N que puede ignorarse para valores de n ≤ N/20*

Estimadores de la Proporción

Sea $\hat{p}$ la proporción muestral de éxitos en una muestra aleatoria extraída de una población en la que la proporción de éxitos es *P*.

1. La distribución de $\hat{p}$ en el muestreo tiene una media *P*:

$$\hat{p} = \frac{Nro\ exitos}{total}$$

2.La distribución de $\hat{p}$ en el muestreo tiene una desviación típica (error estándar)

$$\sigma_{\hat{p}} = S = \frac{(N-n)}{N} \sqrt{\frac{P(1-P)}{n}}$$

3. Si el tamaño de la muestra es grande, la variable aleatoria $\hat{p}$ está distribuida aproximadamente como una normal estándar.

$$Z = \frac{\hat{p} - P}{\sigma_{\hat{p}}}$$

Muestreo Estadístico

Muestreo Aleatorio Simple
Estimación de parámetros

Andy Domínguez

www.andydominguez.weebly.com

adominguez@utb.edu.co

Junio 2025

FACULTAD DE CIENCIAS BÁSICAS

Especialización en Estadística Aplicada y Ciencia de Datos

Muestreo Aleatorio Simple

EJEMPLO

A continuación se observan los saldos correspondientes a los saldos de cuentas de una muestra aleatoria simple n=15 de la población de cuentas por cobrar:

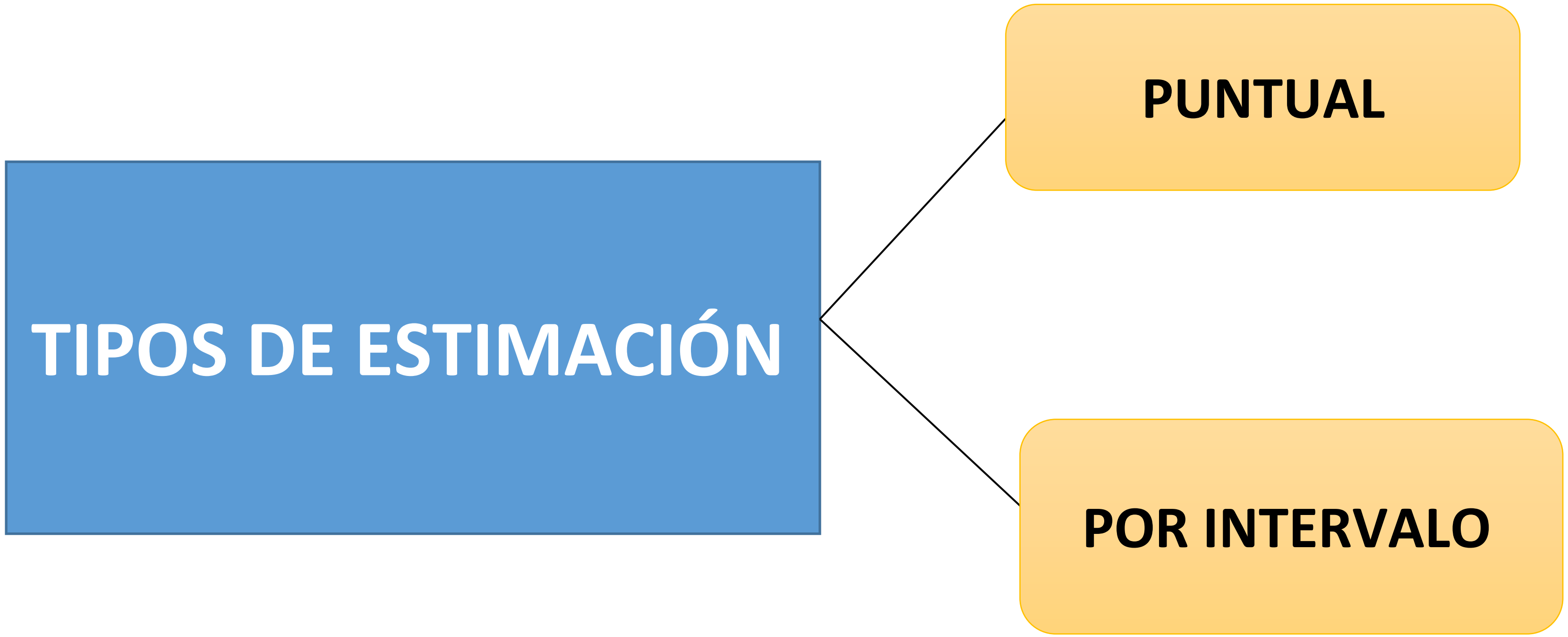
14.5	30.2	17.8	10	8.5
23.49	15.5	27.5	6.9	19.5
42	13.3	23.7	18.4	12.1

Cuadro: Saldos en cuentas por cobrar

Estime el saldo promedio de todas las cuentas ...¿Cómo hacerlo?

ESTIMACIÓN

Es el proceso mediante el cual se busca aproximar el valor de un parámetro poblacional a partir de la información en la muestra.



Estimación Puntual

Estimar un parámetro poblacional mediante un estadístico que predice el valor de dicho parámetro.

$$\mu = \frac{\sum_{i \in N} x_i}{\sum_{i \in N} 1}$$

$$P_A = \frac{casos\ de\ categoría\ A\ en\ N}{N}$$

$$\sigma^2 = \frac{\sum_{i \in N}(x_i - \mu)^2}{N}$$

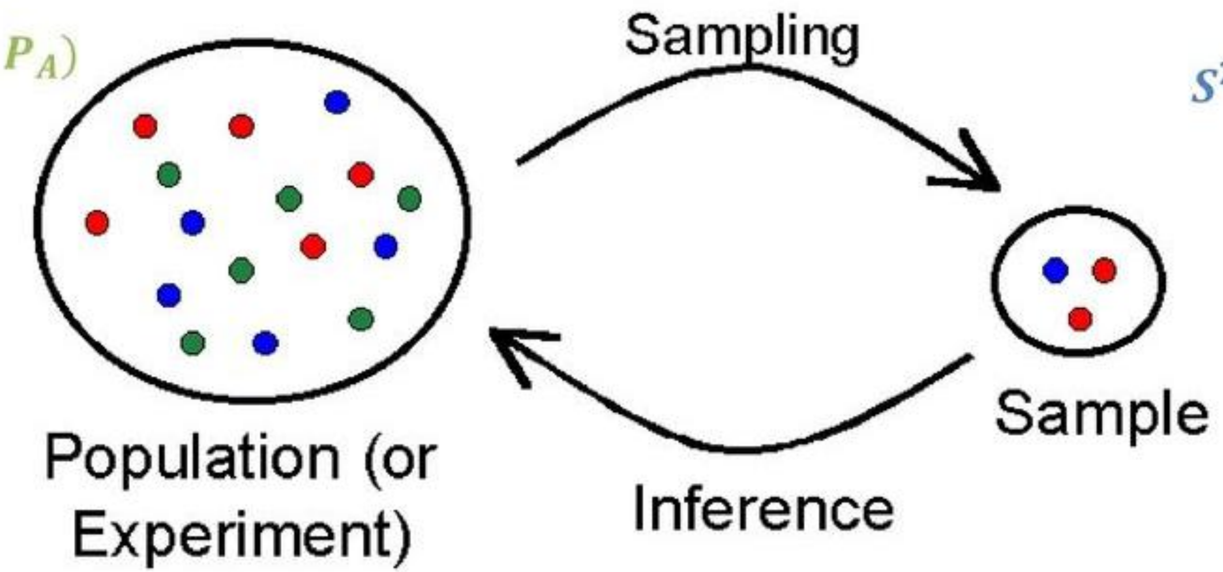
$$\sigma^2 = P_A(1 - P_A)$$

$$\bar{x} = \frac{\sum_{i \in n} x_i}{\sum_{i \in n} 1}$$

$$p_A = \frac{casos\ de\ categoría\ A\ en\ n}{n}$$

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

$$S^2 = \frac{n\widehat{p_A}(1 - \widehat{p_A})}{n-1}$$



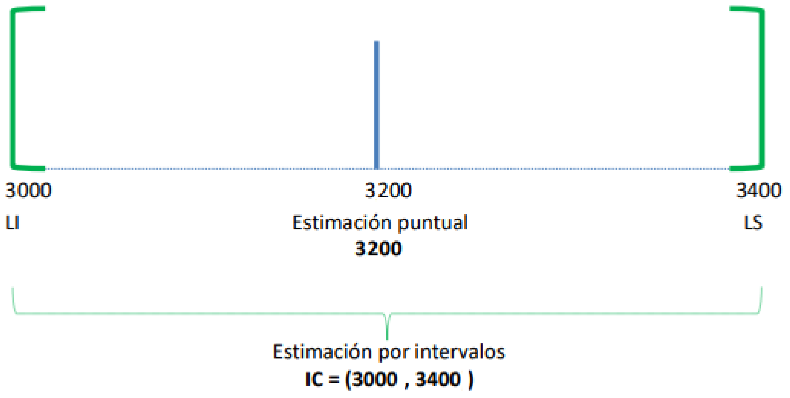
Estimación por Intervalos

Estimar un parámetro poblacional mediante un rango de valores que contiene al parámetro poblacional con una probabilidad conocida.

Diferencias entre los tipos de Estimaciones

Ejemplo: En un evento deportivo desean estimar la cantidad de litros consumido por todos los participantes inscritos. ¿cómo podríamos hacer la estimación?

- Tomar una muestra de 50 posibles asistentes acerca de la cantidad de bebida que consumen en un evento similar
- Estimación Puntual: "en promedio consumen 3200 litros"
- Estimación por Intervalos: "en promedio el consumo será entre 3000 y 3400 litros con un 95% de confiabilidad"



¿Qué necesitamos para construir un IC?

1) El estimador puntual.

2) La varianza de la variable bajo estudio.

3) La función de distribución que rige a la variable bajo estudio.

4) El nivel de confianza (Probabilidad de que el parámetro se encuentre del intervalo) deseado.

Qué necesitamos para Construir un Intervalo de Confianza

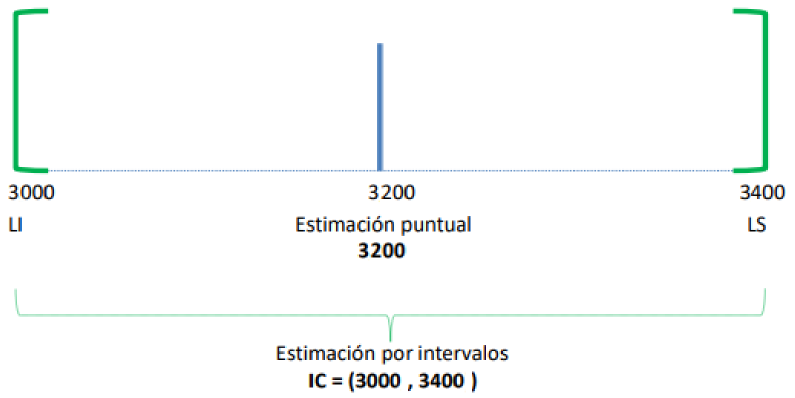
A partir de una muestra tomada de la población, necesitamos:

1) El estimador puntual. → 1) Media muestral, proporción muestral, varianza muestral,….

2) La varianza de la variable bajo estudio. → 2) Varianza poblacional: es o no es conocida?

3) La función de distribución que rige a la variable bajo estudio. → 3) Distribución Normal, Chi-cuadrado, F-Fisher, , t-student (depende de

4) El nivel de confianza (Probabilidad de que el parámetro se encuentre del intervalo) deseado. → 4) 90%, 95%, 99%,…



INTERVALOS DE CONFIANZA PARA LA MEDIA μ

INTERVALO DE CONFIANZA PARA LA MEDIA μ (σ^2 conocida)

- Si la varianza poblacional σ^2 es conocida
- Si n>30 o la población está normalmente distribuida

$$e = error\ de\ estimación$$

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

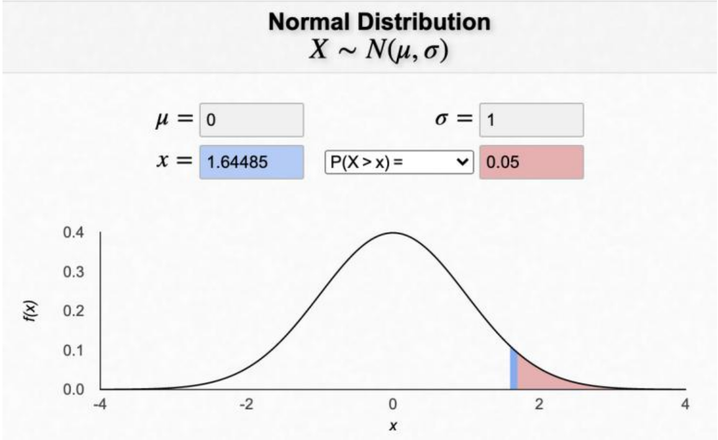
$$\mu = \bar{x} \pm e = [\,\bar{x} - e, \bar{x} + e] = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ , \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

***factor de corrección** para poblaciones finitas (N − n)/N que puede ignorarse para valores de n ≤ N/20*

Donde $(1-\alpha)$ *es el nivel de confianza y* $z_{\alpha/2}$ *es el valor de la variable normal estándar por encima del área* $\boldsymbol{\alpha/2}$

TABLA 8.1 VALORES DE $z_{\alpha/2}$ PARA LOS NIVELES DE CONFIANZA MÁS USADOS

Nivel de confianza	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576



INTERVALO DE CONFIANZA PARA LA MEDIA μ (σ^2 NO conocida)

- Si la varianza poblacional σ^2 es **desconocida**
- Si n>30 o la población está normalmente distribuida

$$e = error\ de\ estimación\ o\ margen\ de\ error$$

$$\mu = \bar{x} \pm e = [\,\bar{x} - e, \bar{x} + e] = \left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}\ , \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

$$e = z_{\alpha/2}\frac{s}{\sqrt{n}}$$

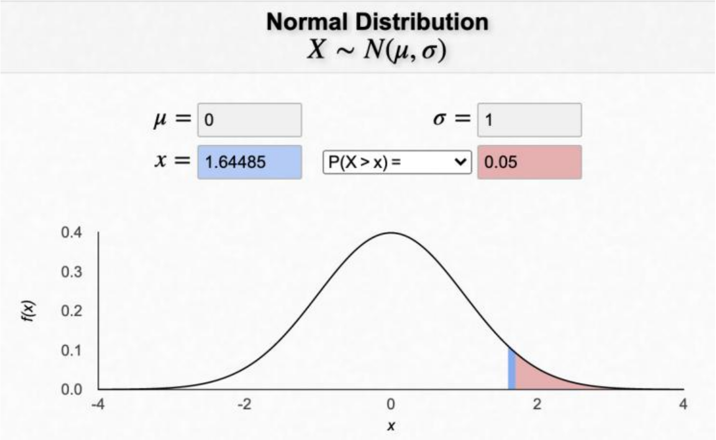
factor de corrección *para poblaciones finitas (N – n)/N que puede ignorarse para valores de n ≤ N/20*

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

$Donde\ (1-\alpha) es\ el\ nivel\ de\ confianza\ y\ z_{\alpha/2}\ es\ el\ valor\ de\ la\ variable\ normal\ estándar\ por\ encima\ del\ área\ \boldsymbol{\alpha/2}$

TABLA 8.1 VALORES DE $z_{\alpha/2}$ PARA LOS NIVELES DE CONFIANZA MÁS USADOS

Nivel de confianza	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576



INTERVALO DE CONFIANZA PARA LA MEDIA μ (σ^2 conocida)

- Si la varianza poblacional σ^2 es conocida
- Si n<30 o la población está normalmente distribuida

$e = error\ de\ estimación$

$$e = t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

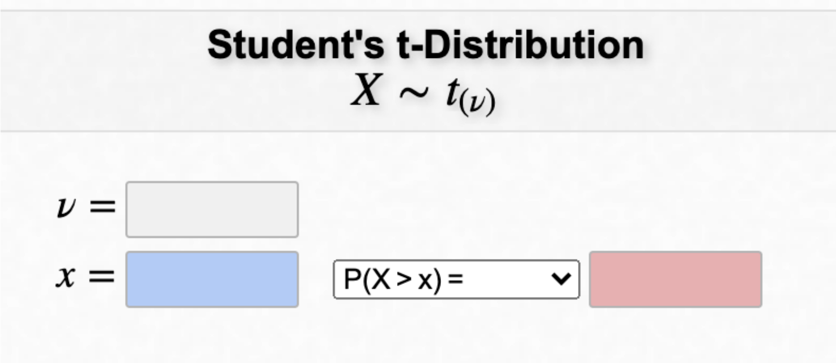
factor de corrección para poblaciones finitas (N − n)/N que puede ignorarse para valores de $n \leq N/20$

$$\mu = \bar{x} \pm e = [\,\bar{x} - e, \bar{x} + e] = \left[\bar{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ , \bar{x} + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Donde $(1-\alpha)$ *es el nivel de confianza y* $t_{\alpha/2}$ *es el valor de la variable* $t-student$ *por encima del área* $\boldsymbol{\alpha/2}$

$$t_{v,\alpha/2}$$

$v = \text{n} - 1$ grados de libertad



En R:
dt(x, df, lower.tail = T)

https://homepage.divms.uiowa.edu/~mbognar/applets/t.html

INTERVALO DE CONFIANZA PARA LA MEDIA μ (σ^2NO conocida)

- Si la varianza poblacional σ^2es **desconocida**
- Si n<30 o la población está normalmente distribuida

$e = error\ de\ estimación$

$$e = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

factor de corrección *para poblaciones finitas (N − n)/N que puede ignorarse para valores de n ≤ N/20*

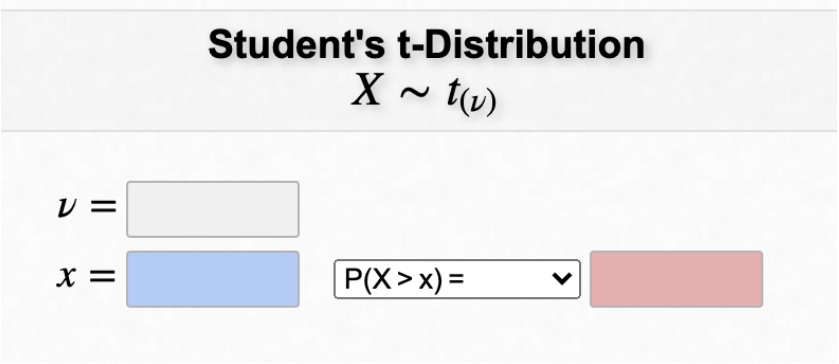
$$\mu = \bar{x} \pm e = [\,\bar{x} - e, \bar{x} + e] = \left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}\ , \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

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

$Donde\ (1-\alpha) es\ el\ nivel\ de\ confianza\ y\ t_{\alpha/2}\ es\ el\ valor\ de\ la\ variable\ t - student\ por\ encima\ del\ área\ \boldsymbol{\alpha/2}$

$$t_{v,\alpha/2}$$

$v = \mathrm{n} - 1$ grados de liberta



En R:
dt(x, df, lower.tail = T)

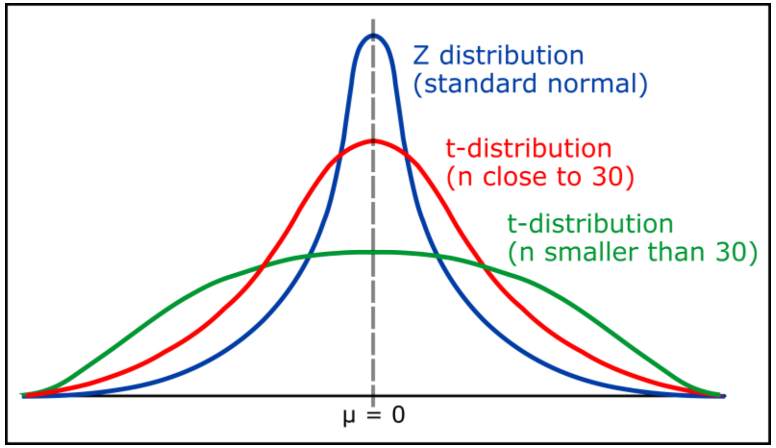
https://homepage.divms.uiowa.edu/~mbognar/applets/t.html

EN RESUMEN….

Simple Random Sampling (Muestreo Aleaorio Simple)

Intervalos de confianza para la media

$$\bar{y} \pm z^* \frac{s}{\sqrt{n}} \qquad \text{and} \qquad \bar{y} \pm t^* \frac{s}{\sqrt{n}}$$



- These formulas are applicable if a sample was to be taken from an infinitely or extremely large population. $n<30$ are not large (t-statistics is used).

- where $z*$ is the upper $\alpha/2$ critical value from the standard normal distribution. For 90%, 95%, and 99%, $z* = 1.645$, 1.96, and 2.576, respectively

factor de corrección *para poblaciones finitas* $(N - n)/N$ *que puede ignorarse para valores de* $n \leq N/20$

INTERVALO DE CONFIANZA PARA LA MEDIA μ

En una muestra aleatoria simple de 50 artículos de una población en la que $\sigma = 6$ la media muestral fue 32.

a. Proporcione un intervalo de confianza de 90% para la media poblacional.

b. Establezca un intervalo de confianza de 95% para la media poblacional.

$a)$

$n = 50$

$\bar{x} = 32$

$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.645)\frac{6}{\sqrt{50}} = 1.40$

$\sigma = 6$

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

$\mu = [32 - 1.40, 32 + 1.40 = [30.60, 33.40]$

$b)$

$n = 50$

$\bar{x} = 32$

$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.96)\frac{6}{\sqrt{50}} = 1.66$

$\sigma = 6$

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

$\mu = [32 - 1.66, 32 + 1.66 = [30.34, 33.66]$

INTERVALO DE CONFIANZA PARA LA MEDIA μ

La vida útil de los focos led para iluminación de escenarios que produce una empresa tiene una desviación estándar de 40 horas, si se toma una **muestra aleatoria simple** de 35 focos led y se encuentra que su vida útil promedio es de 835 horas, construya un intervalo de confianza de 95% para el promedio de vida útil de esos focos.

$n = 35$

$\bar{x} = 835$

$\sigma = 40$

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (1.96)\frac{40}{\sqrt{35}} = 13.25$$

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

$$\mu = [835 - 13.25, 835 + 13.25] = [821.74, 848.25]$$

se estima con una confianza de 95% que el promedio de la vida útil de esos focos para iluminación de escenarios está entre 821.74 *y* 848.25

INTERVALO DE CONFIANZA PARA LA MEDIA μ

En una **muestra aleatoria simple** de 40 comprobantes de compra de un supermercado, la compra promedio fue de $114, con una desviación estándar de $33; si las compras siguen una distribución aproximadamente normal, haga una estimación de intervalo con un nivel de confianza de 99% del promedio de compra del total de compras.

TABLA 8.1 VALORES DE $z_{\alpha/2}$ PARA LOS NIVELES DE CONFIANZA MÁS USADOS

Nivel de confianza	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576

$$n = 40$$
$$\bar{x} = 114$$
$$s = 33$$

$$e = z_{\alpha/2}\frac{s}{\sqrt{n}} = (2.576)\frac{33}{\sqrt{40}} = 13.44$$

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

$$\mu = [114 - 13.44, 114 + 13.44 = [100.55, 127.44]$$

se tiene una confianza de 99% de estar en lo correcto al aseverar que, en la población total de compras, la compra promedio está entre $ 100.55 y $ 127.44

INTERVALO DE CONFIANZA PARA LA MEDIA μ

¿Los comerciales interrumpen constantemente tu programa de TV favorito?. Una consultora estadística presento los datos de un studio sobre la cantidad promedio de minutos de programa en media hora de transmission(sin commercial). Los datos (en minutos) son:

21.06	22.24	20.62
21.66	21.23	23.86
23.82	20.30	21.52
21.52	21.91	23.14
20.02	22.20	21.20
22.37	22.19	22.34
23.36	23.44	

De una estimación puntual y un intervalo de confianza de 95% para la cantidad media de minutos de programa de media hora.

INTERVALO DE CONFIANZA PARA LA MEDIA μ

¿Los comerciales interrumpen constantemente tu programa de TV favorito?. Una consultora estadística presento los datos de un studio sobre la cantidad promedio de minutos de programa en media hora de transmission(sin commercial). Los datos (en minutos) son:

21.06	22.24	20.62
21.66	21.23	23.86
23.82	20.30	21.52
21.52	21.91	23.14
20.02	22.20	21.20
22.37	22.19	22.34
23.36	23.44	

Implementación de estimación MAS con Python

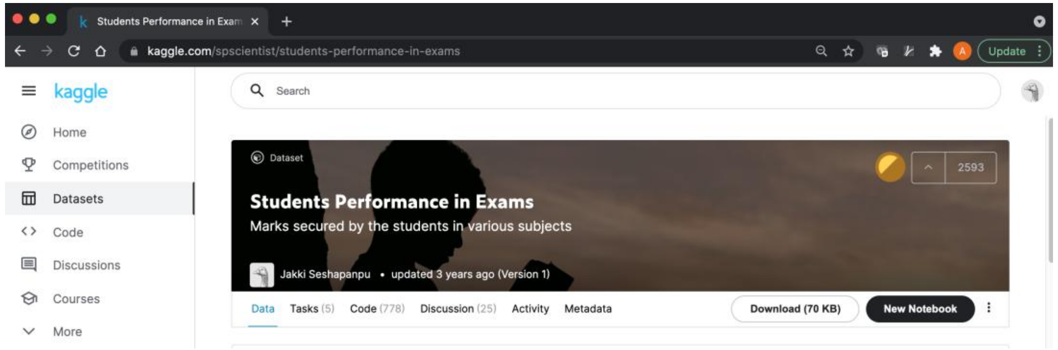
Descargar archivo en SAVIO

Ejercicio Python Estimación con MAS

Objetivo: Aplicar el MAS para realizar inferencias acerca del desempeño estudiantil

Datos: 1000 Estudiantes con resultados de exámenes y otras variables

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
1	female	group B	bachelor's degree	standard	none	72	72	74
2	female	group C	some college	standard	completed	69	90	88
3	female	group B	master's degree	standard	none	90	95	93
4	male	group A	associate's degree	free/reduced	none	47	57	44
5	male	group C	some college	standard	none	76	78	75



https://www.kaggle.com/spscientist/students-performance-in-exams

1) Tome una **Muestra Aleatoria Simple** sin reemplazo de tamaño 100.
2) A partir de la muestra estime un intervalo de confianza al 95% del desempeño en **matemáticas** para la población de estudiantes. Compare el resultado obtenido con la media poblacional en **matemáticas** de los mil estudiantes. Concluya acerca de la estimación.
3) A partir de la muestra estime un intervalo de confianza al 95% del desempeño en **lectura** para la población de estudiantes. Compare el resultado obtenido con la media poblacional de **lectura** de los mil estudiantes. Concluya acerca de la estimación.
4) A partir de la muestra estime un intervalo de confianza al 95% del desempeño en matemáticas para la población de estudiantes. Compare el resultado obtenido con la media poblacional de de **escritura** los mil estudiantes. Concluya acerca de la estimación.
5) ¿Estime en un intervalo al 99% la proporción de los que toman el curso de preparación del examen en la muestra tomada? Compare con la proporción poblacional?
6) ¿De todo lo anterior que podría inferir? Por Ej: ¿En qué area les va mejor?
7) Tome **200 muestras aleaotria simple de tamaño 30** y realice la distribución muestral de la media en **matemáticas. ¿Cuánto es la media de las muestras? ¿Qué tanto se desvía de la media poblacional?**

Ejercicio Python Estimación con MAS

FACULTAD DE CIENCIAS BÁSICAS

Objetivo: Aplicar el MAS para realizar comportamiento de usuarios en Redes Sociales

Datos: Archivo *social_media_engagement.csv* con 1500 usuarios y variables:

followers: número de seguidores.
avg_likes: promedio de "likes" por publicación.
time_in_app_min: tiempo diario en la app (minutos).

1. Toma una **muestra aleatoria simple de tamaño 120** del dataset.
2. Estima un intervalo de confianza al 95% para el promedio de minutos diarios en la app (*time_in_app_min*) de toda la población de usuarios.
3. Estima un intervalo de confianza al 95% para el promedio de likes (*avg_likes*).
4. Compara las medias de las muestras con las medias poblacionales. ¿Las estimaciones parecen precisas?
5. Realiza 300 muestras aleatorias de tamaño 40 para construir la distribución muestral del promedio de likes. Calcula la media y desviación de esa distribución.
6. Reflexiona:

- ¿Cómo evidencia esta simulación el Teorema Central del Límite?
- ¿Qué le dirías a un gerente de marketing que quiere saber cuántos likes promedio esperar en un usuario típico?

TAMAÑO DE LA MUESTRA

Si la población es homogenea, una muestra pequeña puede contener tanta información como una grande:

Ejemplos

Una muestra de sangre

Una muestra de un lote.

Cuando la población es muy disímil en cambio, una muestra pequeña puede contener información muy escasa de la población sobre la que se quiere inferir.

Ejemplos

Altura de los estudiantes (Deportistas)

Potencia del motor de los vehículos de los profesores de la UTB

TAMAÑO DE LA MUESTRA

Para estimar una media

Parámetro	N conocida	N desconocida
Media	$n = \frac{Z^2 S^2 N}{Ne^2 + Z^2 S^2}$	$n = \frac{Z^2 S^2}{e^2}$
Proporción	$n = \frac{Z^2 pqN}{Ne^2 + Z^2 pq}$	$n = \frac{Z^2 pq}{e^2}$

¿Cómo influye el tamaño de la muestra?

A mayor confianza ➔ mayor tamaño de muestra

A menor error ➔ mayor tamaño de muestra

A mayor variabilidad ➔ mayor tamaño de muestra

n = tamaño de la muestra

N = tamaño de la población

S = desviación estándar

p = proporción muestral

q = 1-p
e = error permitido

TABLA 8.1 VALORES DE $z_{\alpha/2}$ PARA LOS NIVELES DE CONFIANZA MÁS USADOS

Nivel de confianza	α	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576

Z = nivel de confianza, expresado como valor del cuantil "z" de una distribución normal estándar, que separa la curva en dos áreas de tamaño 1 - α/2 y α/2 (0 < α < 1)

EN EL CASO DE LA PROPORCIÓN: TAMAÑO DE LA MUESTRA n

Para Estimar una proporción

En una estimación de la proporción poblacional con una precisión determinada el tamaño viene dada por:

$$e = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \Rightarrow \quad \boxed{n = \frac{(z_{\alpha/2})^2\hat{p}(1-\hat{p})}{e^2}}$$

Sin embargo, debido a que NO se conocerá $\hat{p}$ sino hasta que se tome la muestra, no es posible usar esta fórmula para calcular el tamaño de la muestra con el que se obtendrá el margen de error deseado. Usamos un valor planeado $\hat{p}^*$

En la práctica, el valor planeado p^* se determina mediante alguno de los métodos siguientes.

1. Utilizar la proporción poblacional de una muestra previa de las mismas unidades o de unidades similares.
2. Utilizar un estudio piloto y elegir una muestra preliminar. La proporción muestral de esta muestra se usa como valor planeado, p^*.
3. Proponer una "mejor aproximación" para el valor de p^*.
4. Si no aplica ninguna de las alternativas anteriores, emplear como valor planeado $p^* = 0.50$.

$$\boxed{n = \frac{(z_{\frac{\alpha}{2}})^2\hat{p}^*(1-\hat{p}^*)}{e^2}}$$

TAMAÑO DE LA MUESTRA n

Supongamos que queremos realizar un estudio con un intervalo de confianza del 99% y en la estimación no queremos cometer un error mayor de 2. Calcula el tamaño muestral siendo que conocemos de estudios anteriores que σ = 5.

$$\boxed{n = \left(z_{\alpha/2}\frac{\sigma}{e}\right)^2 = \frac{{z_{\alpha/2}}^2\sigma^2}{e^2} = \frac{2.57^2 5^2}{2^2} = 41.28 = 42}$$

TAMAÑO DE LA MUESTRA n

Se realizará un estudio para estimar la proporción de residentes de cierta ciudad y sus suburbios que está favor de que se construya una planta de energía nuclear cerca de la ciudad. ¿Qué tan grande debería ser la muestra, si se desea tener al menos un 95% de confianza en que el estimado esté dentro del 0.04 de la verdadera proporción de residentes que están a favor de que se construya la planta de energía nuclear?

$$\boxed{n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \hat{p}^*(1-\hat{p}^*)}{e^2}}$$

$\hat{p}^*$=0.5 como un valor planeado

$$\boxed{n = \frac{(1.96)^2 0.5(0.5)}{0.04^2} = 600.25 \approx 601}$$