





# Muestreo Probabilístico

Giovany Babativa, PhD

## Sobre Mi

PhD en Estadística, MSc en Big Data, MSc en Estadística. Con 15 años de experiencia, actual director de analítica en el CNC, miembro del comité de expertos en pobreza en el DANE y consultor de la División de Estadística de la CEPAL. Ex-decano de la Facultad de Estadística USTA, ex-director de operaciones en el ICFES,...

Puedes encontrarme en:

-  [Google scholar](#)
-  [GitHub. https://github.com/jgbabativam](https://github.com/jgbabativam)
-  [linkedin](#)
-  [j.babativamarquez@uniandes.edu.co](mailto:j.babativamarquez@uniandes.edu.co)

# MUESTREO PROBABILÍSTICO

Defina a  $U$  un universo<sup>1</sup> de elementos  $\{U_1, \dots, U_N\}$  **finito** y **conocido** de antemano con una variable de interés que toma valores. Sea el parámetro (medida del universo) una función de de esta manera a se denomina parámetro y se denota.

1 En adelante se denominará universo a la población objetivo

Algunos parámetros de interés en un estudio por muestreo:

- Total:
- Promedio:
- Razón:

Se define probabilidad de inclusión de primer orden del elemento

Sea:

Entonces

# Probabilidades de Inclusión

Se define probabilidad de inclusión de segundo orden de los elementos y

Entonces

- 
- Por definición de muestra probabilística .
- En muestreo de elementos para todo son conocidos de antemano.



# Ejercicio

## Tu turno

- Use el espacio muestral del diseño MAS para calcular la probabilidad de inclusión del elemento 1, verifique que .
- Use el espacio muestral del diseño BER para calcular la probabilidad de inclusión del elemento 1, verifique que .

10:00

9

# Estadística y Estimador

Sea una estadística o estimador entonces bajo el diseño muestral define:

- Valor esperado:
- Sesgo:
- Varianza:
- Error Cuadrático Medio:

# Estadística y Estimador

Sea una estadística o estimador entonces bajo el diseño muestral define:

- Error estándar:
- Coeficiente de variación (error relativo):
- Coeficiente de variación estimado:

# Ejercicio

Tu turno

Calcule para:

Use el espacio muestral de un MAS

10:00

12

# Estadística

Ejemplo:

# Estadística

## Ejemplo:

- ejemplo MAS
- ejemplo BER

# Estadística

Ejemplo:





# Estadística

## Ejemplo:

- ejemplo MAS y BER



# Parámetros de interés

Para un universo de tamaño  $N$ , sea la característica de interés, entonces podríamos estar interesados en:

- **Total:** (personas con cierta enfermedad)
- **Media:** (dinero)
- **Proporción:** para (desplazados)
- **Razón:** . Unidades del producto por establecimiento con la intención de venderlo.

Nótese que todos los parámetros pueden ser expresados como función de totales, por tanto hay un particular interés encontrar estimadores para este parámetro.



## Estimador de Horvitz-Thompson (1952)

Para un universo se desea estimar el total de una característica de interés denotado como  $\mu$ . Por ejemplo,

Para  $\mu$  se define:

se denomina

Cada elemento se representa a sí mismo y a una fracción de la población.

# Estimador de Horvitz-Thompson (1952) - Propiedades

## Resultado

- demost.
- 
- 

, además

# Estimador de Horvitz-Thompson (1952)

## Muestreo Aleatorio Simple

- 
- 
-

# Estimador de Horvitz-Thompson (1952)

## Muestreo Bernoulli

- 
- 
- 

**Ejercicio:** Para los ejercicios hechos en clase de MAS y Bernoulli calcule y , este último vía definición y por el estimador dado por la expresión. Para ambas expresiones compruebe el insesgamiento del estimador.





# Estimación de la media poblacional

## Muestreo Aleatorio Simple

- 
- 
- 

Al factor se le conoce como factor de corrección para poblaciones finitas.

# DISEÑOS CON INFORMACIÓN AUXILIAR

# Muestreo Proporcional al Tamaño

Se realizan eventos aleatorios con probabilidad constante, es decir, se selecciona uno de los elementos, se realiza su medición, se reposiciona en el universo y se selecciona de nuevo uno de los elementos con medición y así sucesivamente hasta completar elementos. Sea:

- : *Número de veces que el elemento aparece en la muestra*
- 
- 
- 
- 
-



Una generalización conduce al diseño de Probabilidad Proporcional al Tamaño (PPT) cuando cada elemento se selecciona con probabilidad , tal que . Para este diseño se debe contar con información auxiliar disponible para todos los elementos de , este diseño tienes varios beneficios entre los que se encuentra que reduce costos y que su uso es relativamente simple.

# Algoritmo de Selección: Acumulativo Total

## Ejemplo

Suponga que , para un estudio de mercados se desea seleccionar una muestra de tiendas, pero basado en la información auxiliar se desea darle más peso a las tiendas que tienen una mayor rotación del producto. La información auxiliar para las 9 tiendas es:

Tienda	1	2	3	4	5	6	7	8	9
Ventas_Unid	32	96	65	140	22	78	65	47	106

La muestra ordenada es de tamaño  $n$  mientras que la muestra no ordenada (sin repetición) es de tamaño  $n$ .

El estimador cuenta con sumandos, es decir, que los elementos repetidos solo cuentan una vez dentro del estimador.



# Estimador de Hansen-Hurwitz (1943)

En donde es la probabilidad de selección del elemento . Cada elemento se representa así mismo y al resto del universo promedio.

# Estimador de Hansen-Hurwitz (1943)



# Estimador de Hansen-Hurwitz (1943)

-

•

•

## Ejemplo:

### Tu turno

Para el estudio de mercados suponga que ; y que los valores de las ventas semanales recolectados los establecimientos seleccionados fueron

- Usando el estimador, estime el total de ventas semanales (demanda) del producto en las nueve tiendas.
- Usando el estimador, estime el total de ventas semanales (demanda) del producto en las nueve tiendas y calcule el .

10:00

31

- Es fácil comprobar que si  $\mu$  es constante. Es decir, si es exactamente proporcional a  $\mu$ .

En la práctica eso es imposible !!!... primero es la probabilidad y luego la medición.

- En la práctica, sea una variable auxiliar altamente correlacionada con así entonces se puede determinar

Entre más esté correlacionado con entonces

- Si los no varían casi, me voy con pero si varían mucho me voy con .



## Ejercicio

### Tu turno

Suponga que se tiene  $y$  que se conocen todos los valores de la variable de interés, que están dados por:

Y	79.0	76.00	54.0	39.00	12.0
<hr/>					
pk	0.1	0.15	0.2	0.25	0.3

Para un muestreo con reemplazamiento con  $n$ , si los valores aleatorios son  $u$ , determine la muestra, calcule la estimación y el cve usando el  $\bar{y}$ -estimador. ¿Qué tendría que hacer para determinar si es mejor la estrategia usando  $\bar{y}$ -estimador o  $\bar{y}_p$ -estimador?

12:00

35

# Ejercicio

Tu turno

Resuelva los ejercicios: 3.1.1, 3.3, 3.4, 3.12, 3.14, 3.15 y 3.19 del libro de Gutiérrez, H.A. (2016)

# CONFIABILIDAD

Sea el conjunto de todas las muestras posibles ; para el diseño ; el parámetro y el estimador .

# Confiabilidad

Para cada muestra se construye el intervalo de tal manera que cada muestra tiene su propio intervalo y la longitud del intervalo puede variar dependiendo de la muestra.

Se define **confiabilidad** como la suma de las probabilidades de las muestras de las muestras cuyo intervalo de confianza cubre al parámetro:

Cuando se habla de confiabilidad se habla de la confiabilidad del intervalo NO de la muestra.

# Propuesta de intervalos

- , por TLC se tiene que:

si .

Este intervalo presenta dos inconvenientes:

- Condición de estimador insesgado
- Se necesita la varianza del estimador y esa sólo se conoce en la teoría porque en la práctica solo se conoce cuando es censo y en ese caso no tiene mucho interés.

- 

Este intervalo presenta la cualidad de que en la práctica puede calcularse, lo que interesa es entonces es ¿Qué confiabilidad tiene?.

Puesto que los estimadores vistos son insesgados tanto para el parámetro de interés como para la varianza se espera que para en un de las muestras el intervalo contenga el parámetro. Pero lo anterior no me garantiza que el parámetro esté en el intervalo de la muestra seleccionada.

# Ejercicio

Tu turno

Calcule la confiabilidad para los dos tipos de intervalo usando los datos del gasto recogidos en clase. Hágalo para el caso MAS, de tarea en casa resuelva el diseño BER.

20:00

42



# Sesgo Relativo

La confiabilidad es una medida que se ve directamente afectada por el sesgo<sup>1</sup> NO por el margen de error. Es decir, si la cobertura del intervalo decrece y por más de que la cobertura puede ser muy baja.

<sup>1</sup> medibles y no medibles

# Sesgo Relativo

## Tu turno

Para el ejercicio de clase INCLUYA sesgos de y usando un nivel de confianza del calcule y grafique la confiabilidad real. Use:

10:00

44

**GRACIAS!**

# Referencias

- Gutiérrez, H. A. (2016). Estrategias de muestreo: Diseño de encuestas y estimación de parámetros. Ediciones de la U.
- Lohr, S. L. (2021). Sampling: design and analysis. Chapman and Hall/CRC.
- Särndal, C. E., Swensson, B., & Wretman, J. (2003). Model assisted survey sampling. Springer Science & Business Media.
- Valliant, R., Dever, J. A., & Kreuter, F. (2013). Practical tools for designing and weighting survey samples (Vol. 1). New York: Springer.

# Citación y derechos de autor

Este material ha sido creado por [Giovany Babativa-Márquez](#) y es de libre distribución bajo la licencia [Creative Commons Attribution-ShareAlike 4.0](#).

Si se copia parcial o totalmente, debe citar la fuente como:

Babativa-Márquez, J.G. *Diapositivas del curso de muestreo probabilístico*. URL: <https://jgbabativam.github.io/Muestreo-I/Semana2.html>. 2024.