ESTADISTICA.AR

Muestreo y distribuciones muestrales

Estadística · Semana 8 · Unidad 5: Muestreo y estimación

1 / 25

NOTAS · 1

Pensada para 90 a 120 minutos. Es el puente a la inferencia: hasta acá, la probabilidad iba de la población a lo que puede pasar; desde hoy, se va de una muestra a la población, y para eso hay que saber cuánto puede variar una muestra.

PREGUNTA PARA LA CLASE

La EPH entrevistó a 1.339 personas en Gran Catamarca, y con ellas habla de 236.275 habitantes. ¿Cómo puede alcanzar con tan pocas?

2 / 25

NOTAS · 2

Que arriesguen. La respuesta tiene dos partes: que la muestra sea elegida al azar (partes 1 y 2) y que se sepa cuánto se puede equivocar (partes 3 a 5).

Un dato que sorprende: la precisión depende casi solo del tamaño de la muestra, no de qué fracción de la población es. Una muestra de mil personas sirve igual para una ciudad que para un país.



NOTAS · 3

Las partes 1 y 2 retoman la semana 1 (población, muestra, sesgo) con los diseños concretos. Las partes 3 a 5 usan las distribuciones de la semana 7.



SEMANA 8 · POR QUÉ Y CÓMO MUESTREAR

Por qué se estudia una muestra

Costo y tiempo

Un censo de toda la población cuesta mucho y tarda; la EPH se hace cada trimestre.

Pruebas destructivas

Para saber cuánto dura una lámpara hay que quemarla: no se puede probar todo el lote.

Calidad

Con menos entrevistas, mejor hechas: un censo mal relevado puede tener más error que una buena muestra.

5 / 25

NOTAS · 5

El costo de muestrear es el error de muestreo: la muestra no da exactamente el valor de la población. Lo que hace la estadística es medirlo.

SEMANA 8 · POR QUÉ Y CÓMO MUESTREAR

Población, marco y muestra

Concepto	Qué es	En la EPH
Población	El conjunto sobre el que se quiere concluir	Los hogares y personas de Gran Catamarca
Marco muestral	El listado del que se elige la muestra	Las áreas y viviendas listadas por el INDEC
Unidad de muestreo	Lo que se sortea	Áreas y, dentro de ellas, viviendas
Muestra	La parte que se observa	429 hogares y 1.339 personas

Lo que no está en el marco no puede salir en la muestra: un marco desactualizado sesga los resultados aunque el sorteo sea perfecto.

6 / 25

NOTAS · 6

Un ejemplo de marco malo: una encuesta telefónica a teléfonos fijos deja afuera a los hogares que solo tienen celular, que no son parecidos al resto.

SEMANA 8 · POR QUÉ Y CÓMO MUESTREAR

Probabilístico o no

Muestreo probabilístico

- Cada unidad tiene una probabilidad conocida y mayor que cero de salir.
- El azar elige, no el encuestador.
- Permite **medir el error**: es la base de la inferencia.

No probabilístico

- Por conveniencia, por cuotas, voluntario, «bola de nieve».
- Rápido y barato.
- No hay forma de saber cuánto se equivoca: puede tener un **sesgo** de tamaño desconocido.

7 / 25

NOTAS · 7

La votación por WhatsApp de la semana 1 era una muestra voluntaria: con 10.000 respuestas seguía igual de lejos del valor real. Más casos no arreglan un sesgo.

Los no probabilísticos sirven para explorar, para probar un cuestionario o en investigación cualitativa, pero no para estimar con un margen de error.

2

Cuatro diseños probabilísticos

Simple, sistemático, estratificado y por conglomerados

8 / 25



NOTAS · 9

En el estratificado los estratos tienen que ser homogéneos por dentro y distintos entre sí; en los conglomerados conviene lo contrario: que cada uno sea una pequeña copia de la población.

El práctico tiene la selección con la tabla de números aleatorios y la afijación de un muestreo estratificado.



NOTAS · 10

La suma de PONDERA en Gran Catamarca da 236.275: la población que la muestra representa. Por eso en la semana 1 los porcentajes ponderados y sin ponderar daban distinto.

Casi todas las encuestas oficiales son así. Las fórmulas de hoy son las del muestreo aleatorio simple; con un diseño complejo el error estándar se calcula de otra manera, que excede el curso.

3

La distribución muestral de la media

Qué pasaría si se repitiera la muestra

11 / 25

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Parámetros y estimadores

La media	La proporción	El desvío
$\bar{x}$ estima μ.	$\hat{p}$ estima p.	s estima σ.

Un estimador es una fórmula que se aplica a la muestra; su valor en una muestra concreta es una estimación.

12 / 25

NOTAS · 12

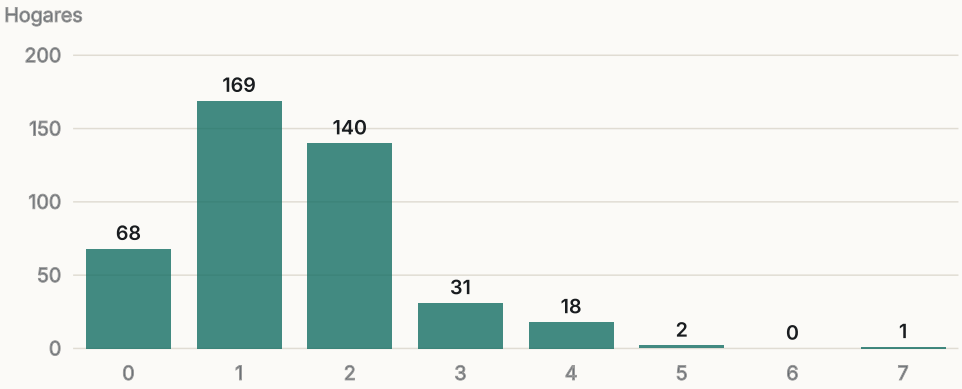
El parámetro es un número fijo y desconocido. El estimador es una variable aleatoria: cambia de muestra en muestra. Por eso tiene una distribución, que es el tema de hoy.

Que un estimador sea insesgado quiere decir que, en promedio sobre todas las muestras posibles, da el parámetro. Por eso s divide por n − 1.

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

La población: 429 hogares

Personas ocupadas en cada hogar encuestado, que hoy hacen de población. μ = 1,471 y σ = 1,043.



Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Gran Catamarca.

13 / 25

NOTAS · 13

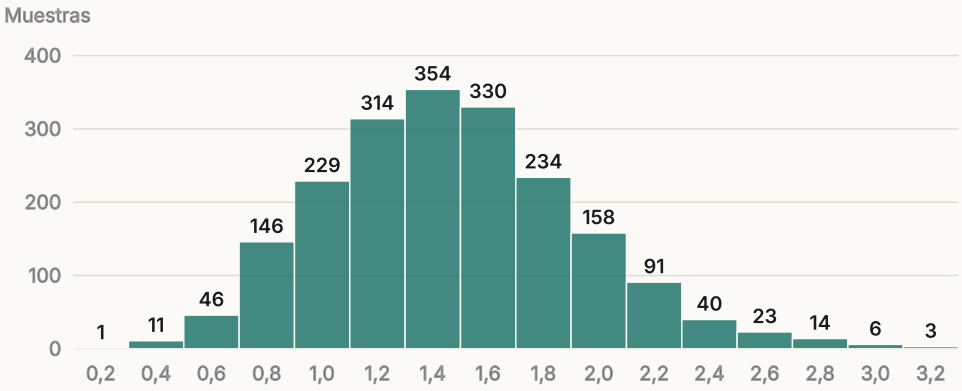
Es la distribución de la semana 7: asimétrica, con una cola hacia la derecha. Desde acá se van a tomar muchas muestras, como si se repitiera la encuesta, para ver cómo se reparte la media de cada una.

σ se calcula dividiendo por N: es el desvío de la población, no el de una muestra.

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

2.000 muestras de 5 hogares

Cuántas muestras dieron cada valor de la media de ocupados.



Muestras aleatorias simples con reposición, simuladas con semilla fija.

14 / 25

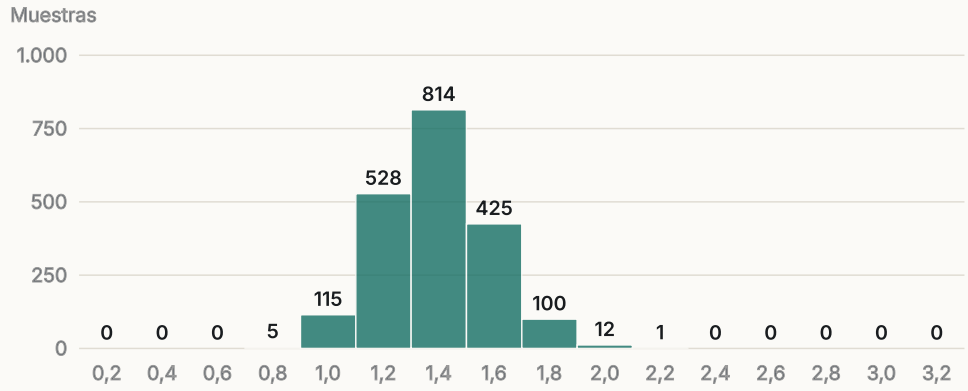
NOTAS · 14

Ninguna muestra da exactamente μ = 1,471: con 5 hogares, la media solo puede ser 0, 0,2, 0,4... Las medias van de 0,2 a 3,2, pero se juntan alrededor de μ. Y la forma ya es menos asimétrica que la de la población.

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

2.000 muestras de 30 hogares

Cuántas muestras dieron la media en cada intervalo, con el mismo eje que el gráfico anterior.



Cada barra va desde su rótulo hasta 0,2 más: de 1,4 a 1,6, por ejemplo.

15 / 25

NOTAS · 15

Con 30 hogares las medias van de 0,90 a 2,20: mucho más juntas, y con forma de campana, aunque la población no la tenga.

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

El centro y la dispersión de las medias

Tamaño de muestra	Media de las medias	Desvío de las medias	σ / √n
La población	1,471	1,043	1,043
n = 5	1,469	0,461	0,467
n = 30	1,473	0,190	0,190

μ = 1,471 y σ = 1,043. Las medias se centran en μ y su desvío se achica como σ / √n.

16 / 25

NOTAS · 16

Dos resultados que valen para cualquier población: $E(\bar{x}) = \mu$, la media muestral es insesgada; y el desvío de $\bar{x}$ es $\sigma / \sqrt{n}$. Ese desvío tiene nombre propio: el error estándar.

Con muestras sin reposición de una población chica, el error estándar se multiplica por $\sqrt{(N - n) / (N - 1)}$, la corrección por población finita. Si la muestra es una parte chica de la población, casi no cambia nada.

SEMANA 8 · LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Error estándar no es desvío estándar

Desvío estándar, σ

Cuánto varían los **hogares** entre sí.

1,043 ocupados: no cambia con n.

Error estándar, σ / √n

Cuánto varían las **medias** de muestra en muestra.

0,190 con 30 hogares: se achica con n.

17 / 25

NOTAS · 17

Es la confusión más común de la unidad, también en los artículos publicados. El desvío describe los datos; el error estándar, la precisión de una estimación.

En la práctica σ no se conoce y se reemplaza por s: el error estándar estimado es s / √n.

PREGUNTA PARA LA CLASE

Si se duplica el tamaño de la muestra, ¿el error estándar se reduce a la mitad?

18 / 25

NOTAS · 18

No: se divide por √2, o sea, queda en el 71 %. Con 30 hogares es 0,190; con 60, 0,135. Para reducirlo a la mitad hay que cuadruplicar la muestra: con 120, 0,095.

Es la razón por la que las encuestas no son enormes: cada ganancia de precisión cuesta cada vez más casos.

4

El teorema central del límite

Por qué aparece la normal

19 / 25

SEMANA 8 · EL TEOREMA CENTRAL DEL LÍMITE

El teorema central del límite

Si n es grande, la media muestral tiene distribución **aproximadamente normal**, con media μ y desvío $\sigma / \sqrt{n}$, **cualquiera sea la forma de la población**.

Cuánto es grande

Con poblaciones no muy asimétricas, $n \geq 30$ suele alcanzar.

Si la población es muy asimétrica, hace falta más.

Si la población es normal

La media es normal con cualquier n.

No hace falta el teorema.

20 / 25

NOTAS · 20

Es el resultado más importante del curso: por él, la normal aparece en casi toda la inferencia aunque los datos no sean normales.

La binomial con n = 100 de la semana pasada ya era un caso: una suma de 100 ensayos independientes.

SEMANA 8 · EL TEOREMA CENTRAL DEL LÍMITE

Lo que permite calcular

¿Qué probabilidad hay de que una muestra de 30 hogares dé una media a menos de 0,2 ocupados de μ? Error estándar: 0,190.

0,7062

con la normal: $P(|Z| < 1{,}05) = 2\Phi(1{,}05) - 1$

0,7145

en las 2.000 muestras simuladas

21 / 25

NOTAS · 21

La normal predice muy de cerca lo que pasó en la simulación, aunque la población sea asimétrica y discreta: es el teorema central del límite en acción.

Es la cuenta que, dada vuelta, da el intervalo de confianza de la semana que viene: en lugar de preguntar qué tan cerca de μ cae $\bar{x}$, se pregunta dónde puede estar μ dado $\bar{x}$.

5

La proporción muestral

El mismo razonamiento para un porcentaje

22 / 25

SEMANA 8 · LA PROPORCIÓN MUESTRAL

La distribución muestral de la proporción

La proporción muestral, x / n, es una media de ceros y unos. Por el teorema central del límite, si n es grande, es aproximadamente normal.

Centro	Error estándar	Condición
$E(\hat{p}) = p$	$\sqrt{(p(1-p)/n)}$	$np \geq 5$ y $n(1-p) \geq 5$
		La misma de la aproximación normal a la binomial.

23 / 25

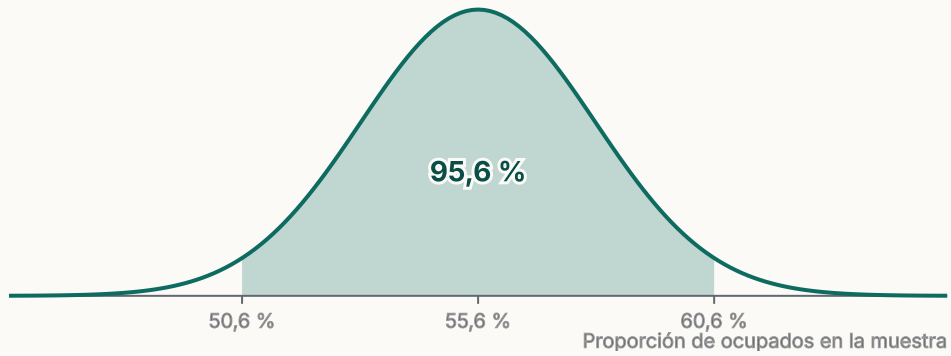
NOTAS · 23

Es la aproximación normal a la binomial de la semana 7, dividida por n: x es binomial y $\hat{p}$ es x / n.

SEMANA 8 · LA PROPORCIÓN MUESTRAL

Una encuesta de 400 personas

Si el 55,6 % de la población está ocupado, el error estándar de $\hat{p}$ es $\sqrt{(0{,}556 \cdot 0{,}444 / 400)} = 0{,}0248$.



$P(|\hat{p} - p| < 0{,}05) = P(|Z| < 2{,}01) = 2 \cdot 0{,}9778 - 1 = 0{,}9556$.

24 / 25

NOTAS · 24

Con 400 personas, en el 95,6 % de las muestras la proporción queda a menos de 5 puntos de la verdadera. Es lo que las encuestas informan como «margen de error», casi siempre con el 95 %.

Las condiciones: np = 222,4 y n(1 − p) = 177,6, las dos muy por encima de 5.

SEMANA 8 · MUESTREO Y DISTRIBUCIONES MUESTRALES

Para llevarse

- Una muestra probabilística permite medir su error; una no probabilística, no.
- El diseño (simple, sistemático, estratificado, por conglomerados) depende del marco, del costo y de la precisión que se busca.
- La media muestral se centra en μ y su error estándar es σ / √n: para reducirlo a la mitad, cuatro veces más casos.
- Por el teorema central del límite, $\bar{x}$ y $\hat{p}$ son aproximadamente normales con n grande.
- El desvío describe los datos; el error estándar, la precisión de la estimación.

Guía de trabajo práctico 8
estadistica.ar/docentes/trabajos-practicos/semana-8

Simulador de muestreo
estadistica.ar/herramientas/sim-muestreo

Simulador del teorema central del límite
estadistica.ar/herramientas/sim-tcl

Ejercicios: error estándar
estadistica.ar/ejercicios/error-estandar

25 / 25

NOTAS · 25

Para el práctico: tabla de números aleatorios, calculadora y, para el ejercicio con software, la base de la EPH.