ESTADISTICA.AR

Muestreo y distribuciones muestrales

Estadística · Semana 8 · Unidad 5: Muestreo y estimación

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?

Hoy

1

Por qué y cómo muestrear

Población, marco y muestra

2

Cuatro diseños probabilísticos

Simple, sistemático, estratificado y por conglomerados

3

La distribución muestral de la media

Qué pasaría si se repitiera la muestra

4

El teorema central del límite

Por qué aparece la normal

5

La proporción muestral

El mismo razonamiento para un porcentaje

1

Por qué y cómo muestrear

Población, marco y muestra

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.

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.

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.

2

Cuatro diseños probabilísticos

Simple, sistemático, estratificado y por conglomerados

Cuatro diseños

Aleatorio simple

Todas las muestras de tamaño n tienen la misma probabilidad.

Hace falta un listado completo.

Sistemático

Una unidad cada k, con un arranque al azar entre 1 y k.

Fácil en el terreno; ojo con los listados periódicos.

Estratificado

Se divide la población en grupos parecidos por dentro y se sortea en cada uno.

Más preciso y da resultados por grupo.

Conglomerados

Se sortean grupos enteros (manzanas, escuelas) y se observa todo o parte de cada uno.

Más barato; menos preciso.

La EPH combina varios

Una muestra **probabilística, estratificada y en etapas**: primero se sortean áreas y después, dentro de cada área, viviendas.

Por qué en etapas

No hay un listado de todas las personas del país, pero sí de las áreas: se listan las viviendas solo en las áreas que salen.

Por qué el ponderador

No todas las viviendas tienen la misma probabilidad de salir. Cada encuestado representa a tantos habitantes como indica PONDERA.

3

La distribución muestral de la media

Qué pasaría si se repitiera la muestra

Parámetros y estimadores

La media

$\bar{x}$ estima μ.

La proporción

$\hat{p}$ estima p.

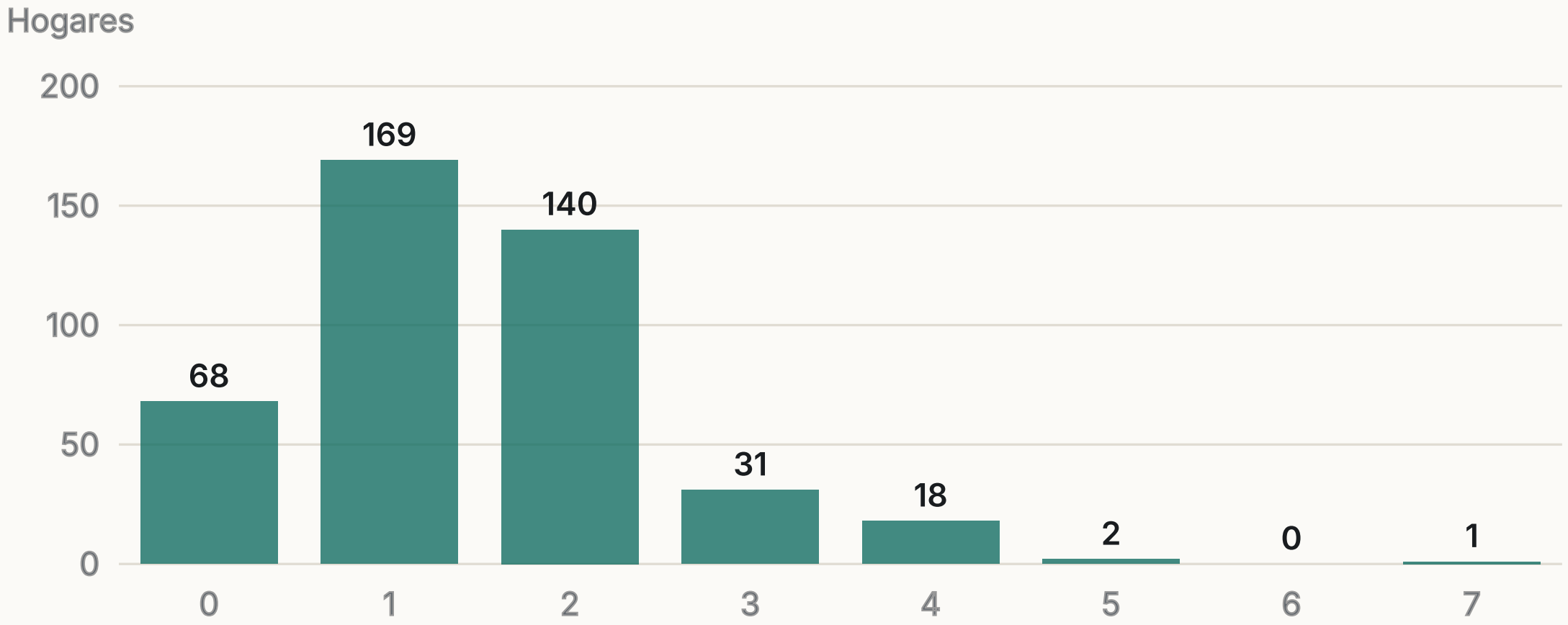
El desvío

s estima σ.

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

La población: 429 hogares

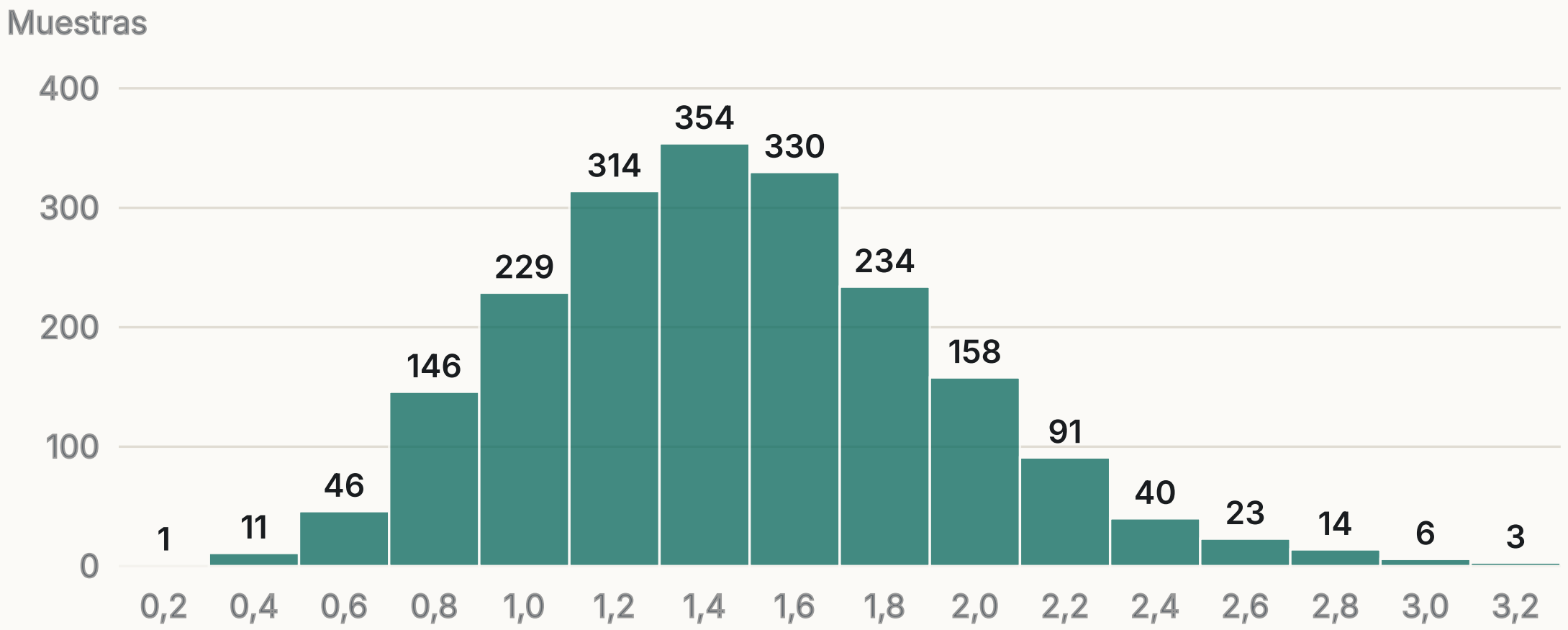
Personas ocupadas en cada hogar encuestado, que hoy hacen de población. μ = 1,471 y σ = 1,043.



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

2.000 muestras de 5 hogares

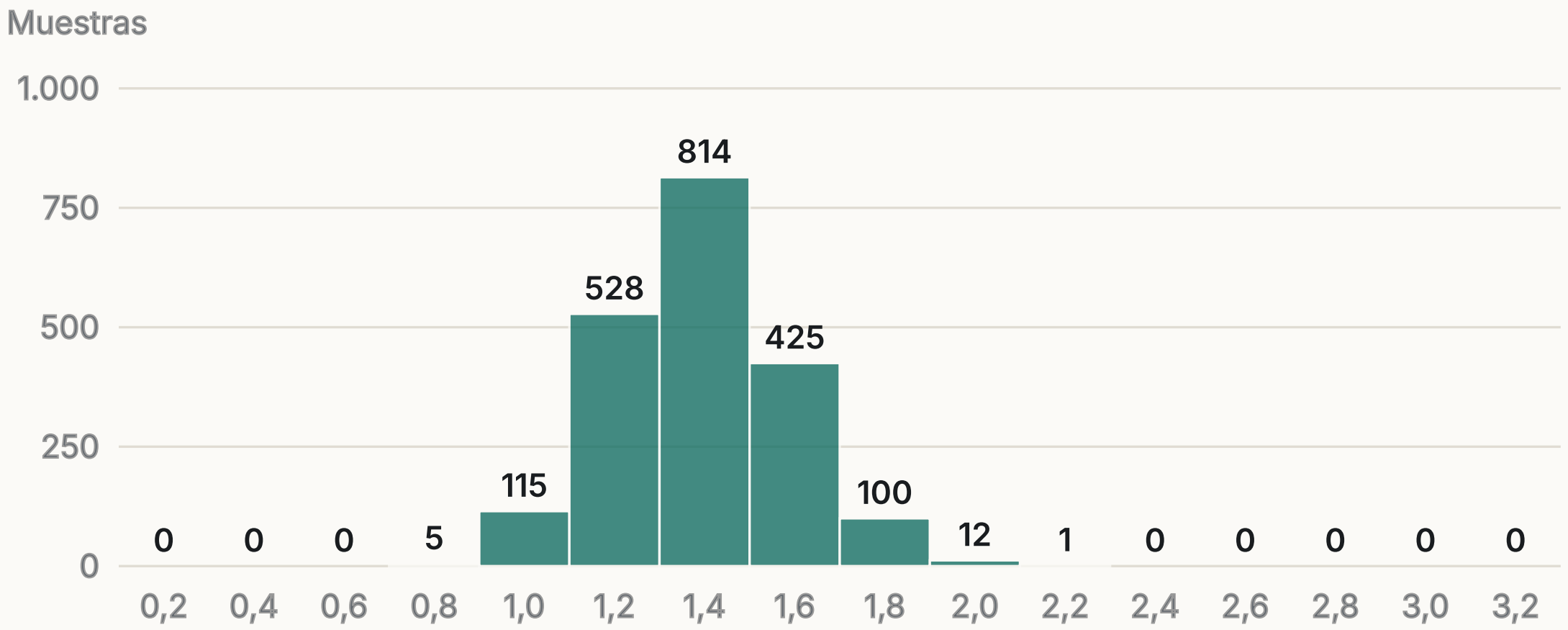
Cuántas muestras dieron cada valor de la media de ocupados.



Muestras aleatorias simples con reposición, simuladas con semilla fija.

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.

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.

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.

PREGUNTA PARA LA CLASE

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

4

El teorema central del límite

Por qué aparece la normal

El teorema central del límite

Si n es grande, la media muestral tiene distribución **aproximadamente normal**, con media μ y desvío σ / √n, **cualquiera sea la forma de la población**.

Cuánto es grande

Con poblaciones no muy asimétricas, n ≥ 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.

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

5

La proporción muestral

El mismo razonamiento para un porcentaje

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

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

Error estándar

$\sqrt{(p(1 - p) / n)}$

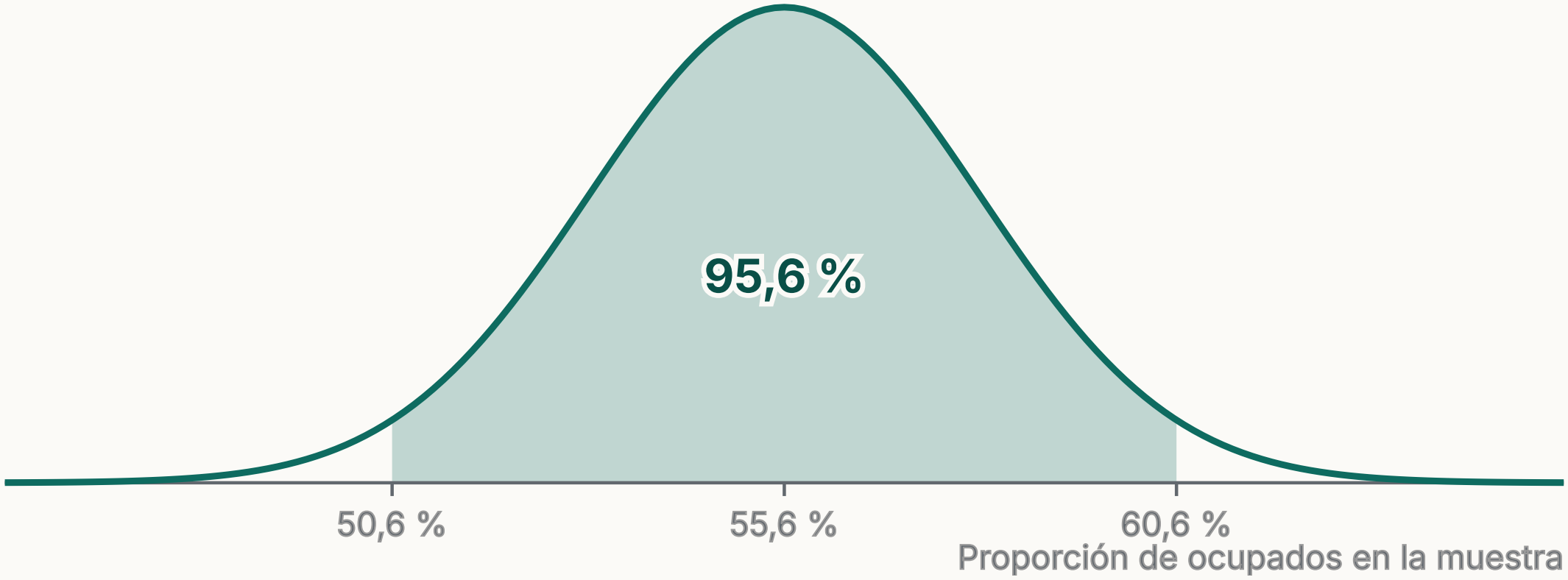
Condición

$np \geq 5$ y $n(1 - p) \geq 5$

La misma de la aproximación normal a la binomial.

Una encuesta de 400 personas

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



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

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 $\sigma / \sqrt{n}$: 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