ESTADISTICA.AR

Los datos y sus variables

Estadística · Semana 1 · Unidad 1: Introducción: los datos

Una encuesta, una ciudad

1.339

personas respondieron la Encuesta Permanente de Hogares en Gran Catamarca

429

hogares visitados por los encuestadores del INDEC

236.275

personas es lo que, entre todas, representan

Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Extracto en estadistica.ar/datos/eph-catamarca-2026t1.

Hoy

1

Para qué sirve la Estadística

Por qué hace falta, y qué hace

2

Población, muestra y unidad de análisis

De quiénes se habla y de qué está hecha cada fila

3

Parámetro y estadístico

Lo que se quiere saber y lo que se calcula

4

Variables y escalas de medición

Qué se registra y qué se puede hacer con cada cosa

5

De dónde salen los datos

Censo, encuesta y registro administrativo

6

Error y sesgo

Por qué un resultado puede estar lejos de la verdad

1

Para qué sirve la Estadística

Por qué hace falta, y qué hace

La Estadística

La disciplina que **recoge, ordena, analiza e interpreta datos** para responder preguntas y tomar decisiones cuando hay incertidumbre.

Por qué existe

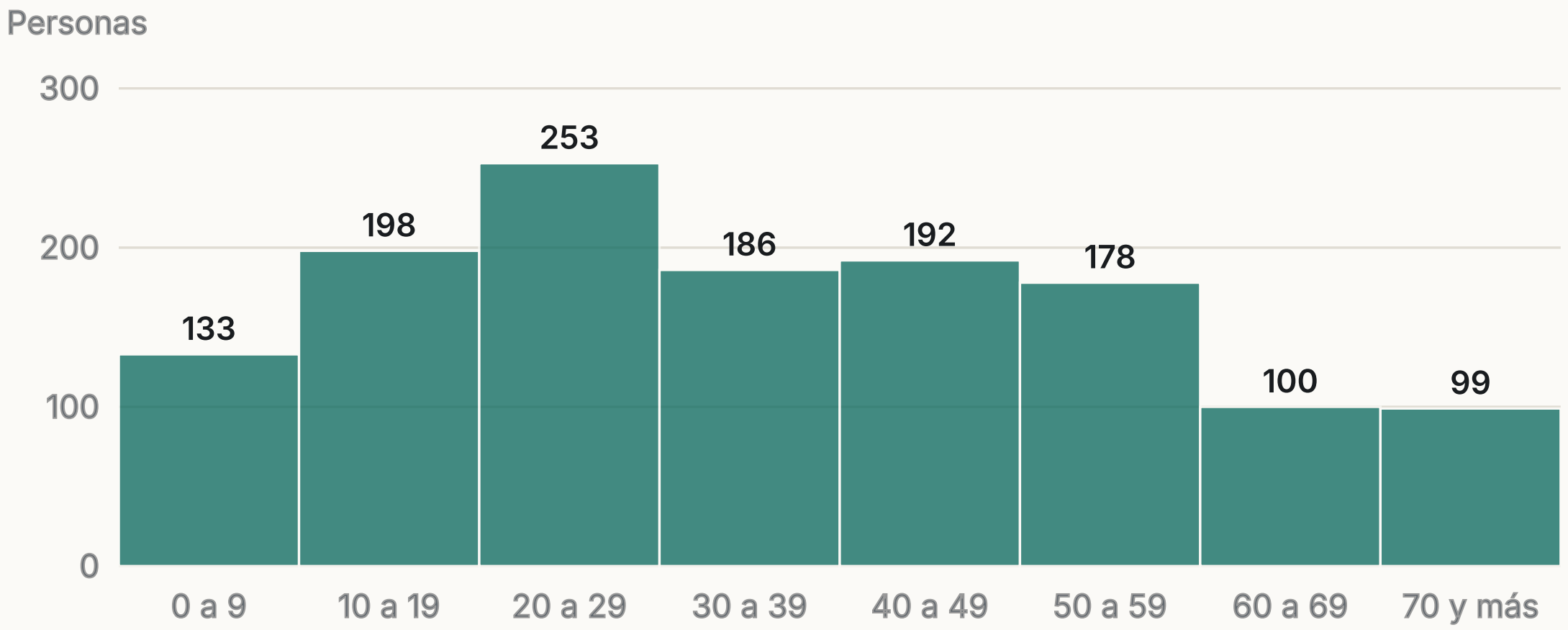
Porque las cosas **varían**: si todas las personas tuvieran la misma edad y el mismo ingreso, alcanzaría con preguntarle a una.

Qué hace

Separa lo que se repite de lo que es azar, y dice cuánto confiar en cada conclusión.

Lo que varía: la edad

Personas encuestadas en Gran Catamarca, por grupo de edad en años cumplidos.



Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Extracto en estadistica.ar/datos/eph-catamarca-2026t1.

Dos tareas

Descriptiva

- Resume y muestra los datos que se tienen.
- Tablas, gráficos, promedios, porcentajes.
- «El 52,5 % de los encuestados de 10 años o más está ocupado.»

Inferencial

- Usa una muestra para decir algo de la población.
- Estimaciones con su margen de error y pruebas de hipótesis.
- «Se estima que en Gran Catamarca está ocupado el 50,9 % de las personas de 10 años o más.»

La primera mitad de la materia es descriptiva; la segunda, inferencial. La probabilidad hace de puente.

2

Población, muestra y unidad de análisis

De quiénes se habla y de qué está hecha cada fila

Las partes de un relevamiento

	Qué es	En la EPH de Gran Catamarca
Población	El conjunto entero sobre el que se pregunta	Las personas que viven en el aglomerado
Muestra	La parte que se observa	Las 1.339 personas de 429 hogares encuestados
Unidad de análisis	Aquello de lo que se dice algo: cada fila de la base	La persona
Variable	Lo que se registra de cada unidad y cambia de una a otra	Edad, sexo, nivel educativo, ingreso
Informante	Quien responde	Un integrante del hogar, que puede responder por los demás

La unidad de análisis no siempre es una persona

Hogar

¿Tuvo cortes de agua el último mes?

Escuela

¿Qué porcentaje de sus alumnos repite?

Departamento

¿Cuántos nacimientos hubo por cada mil habitantes?

Y también

Una consulta médica, una empresa, un municipio, una parcela…

La pregunta que ordena todo: ¿de qué es cada fila?

PREGUNTA PARA LA CLASE

Un hospital registró 3.200 consultas en un mes, de 1.900 pacientes distintos. Queremos saber cuántas veces por mes consulta un paciente. ¿Cuál es la unidad de análisis?

a) La consulta

b) El paciente

c) El hospital

3

Parámetro y estadístico

Lo que se quiere saber y lo que se calcula

Parámetro y estadístico

Parámetro

- Describe a la **población**.
- Es uno solo y casi nunca se conoce.
- Se escribe con letras griegas: μ, σ; la proporción, p.

Estadístico

- Se calcula con la **muestra**.
- Cambia de una muestra a otra.
- Se escribe con letras latinas o con sombrero: $\bar{x}$, s, $\hat{p}$.

Inferir es usar el estadístico para decir algo del parámetro, y decir cuánto puede equivocarse.

¿Qué porcentaje está ocupado en Gran Catamarca?

52,5 %

contando filas: un estadístico que describe a los encuestados

50,9 %

ponderando a cada persona por los habitantes que representa: un estadístico que estima el parámetro

?

el porcentaje en toda la ciudad: el parámetro, que nadie conoce

Personas de 10 años o más con entrevista: 1.202 en la muestra.

PREGUNTA PARA LA CLASE

¿Parámetro o estadístico?

a) «El candidato A obtuvo el 52 % de los votos», según el escrutinio definitivo.

b) «El 48 % votaría al candidato A», según una encuesta a 1.200 personas.

c) La edad media de los estudiantes de esta comisión, si nos interesa esta comisión.

d) La misma media, si nos interesan todos los ingresantes de la facultad.

4

Variables y escalas de medición

Qué se registra y qué se puede hacer con cada cosa

La matriz de datos

Un hogar de la EPH de Gran Catamarca, tal como viene en la base: una fila por persona, una columna por variable.

COMPONENTE	CH04	CH06	NIVEL_ED	ESTADO	P21
1	1	31	4	1	300.000
2	2	29	4	3	0
3	2	10	1	3	0
4	1	8	1	4	0
5	2	4	7	4	0
6	2	2	7	4	0

Sin el diccionario no se lee. CH04: 1 = varón, 2 = mujer. NIVEL_ED: 1 = primario incompleto, 4 = secundario completo, 7 = sin instrucción. ESTADO: 1 = ocupado, 3 = inactivo, 4 = menor de 10 años. P21 es el ingreso de la ocupación principal, en pesos; 0 = no corresponde.

Tipos de variables

Cualitativas: categorías

- **Nominales**, sin orden: sexo, grupo sanguíneo, barrio.
- **Ordinales**, con orden: nivel educativo, satisfacción, estadio de una enfermedad.

Cuantitativas: cantidades

- **Discretas**, se cuentan: hijos, consultas, materias aprobadas.
- **Continuas**, se miden: peso, ingreso, tiempo de espera.

El tipo decide qué tablas, gráficos y medidas tienen sentido. Es lo primero que se pregunta ante una variable.

Las escalas de medición

Escala	Ejemplo	Contar	Ordenar	Restar	Dividir
Nominal	Sexo, barrio	✓			
Ordinal	Nivel educativo	✓	✓		
Intervalo	Temperatura en °C	✓	✓	✓	
Razón	Peso, ingreso, edad	✓	✓	✓	✓

Cada escala permite las operaciones de la anterior y una más (Stevens, 1946).

¿40 °C es el doble de 20 °C?

40 / 20 = 2

en grados Celsius, cuyo cero es arbitrario: el punto de congelamiento del agua

313,15 / 293,15 = 1,07

en kelvin, cuyo cero es la ausencia de calor

En una escala de intervalo tienen sentido las diferencias, no los cocientes. En una de razón, el cero significa «nada».

Un número no siempre es una cantidad

2,13

es el promedio de la columna ESTADO de la EPH. No significa nada: 1, 2, 3 y 4 son etiquetas

7

es el código de NIVEL_ED para «sin instrucción»: el más alto para el nivel más bajo

El tipo de una variable se lee en el diccionario de la base, no en la columna.

PREGUNTA PARA LA CLASE

¿Qué tipo de variable es?

a) Cantidad de materias aprobadas

b) Barrio donde vive

c) Tiempo de viaje hasta la facultad, en minutos

d) Acuerdo con una frase, de «muy en desacuerdo» a «muy de acuerdo»

e) Número de libreta universitaria

f) Ingreso del hogar en tramos de 500.000 pesos

5

De dónde salen los datos

Censo, encuesta y registro administrativo

Tres formas de producir datos

	Censo	Encuesta por muestreo	Registro administrativo
A quiénes cubre	A toda la población	A una muestra	A quienes hacen un trámite
Cada cuánto	Cada diez años	Periódica: la EPH, cada trimestre	Continuo
Error muestral	No tiene	Sí, y se puede calcular	No tiene
Ejemplo	Censo Nacional 2022	Encuesta Permanente de Hogares	Nacimientos y defunciones

Cada fuente responde lo que puede

Censo

Llega a cada localidad, por chica que sea.

Pero se hace cada diez años.

Encuesta

Es rápida y barata, y puede preguntar casi cualquier cosa.

Pero no alcanza para áreas chicas ni para grupos poco frecuentes.

Registro

Es continuo y cubre a todos los que hacen el trámite.

Pero solo trae lo que el trámite pide: el acta de nacimiento no pregunta por el ingreso.

Ninguna es mejor que las otras: la pregunta decide la fuente.

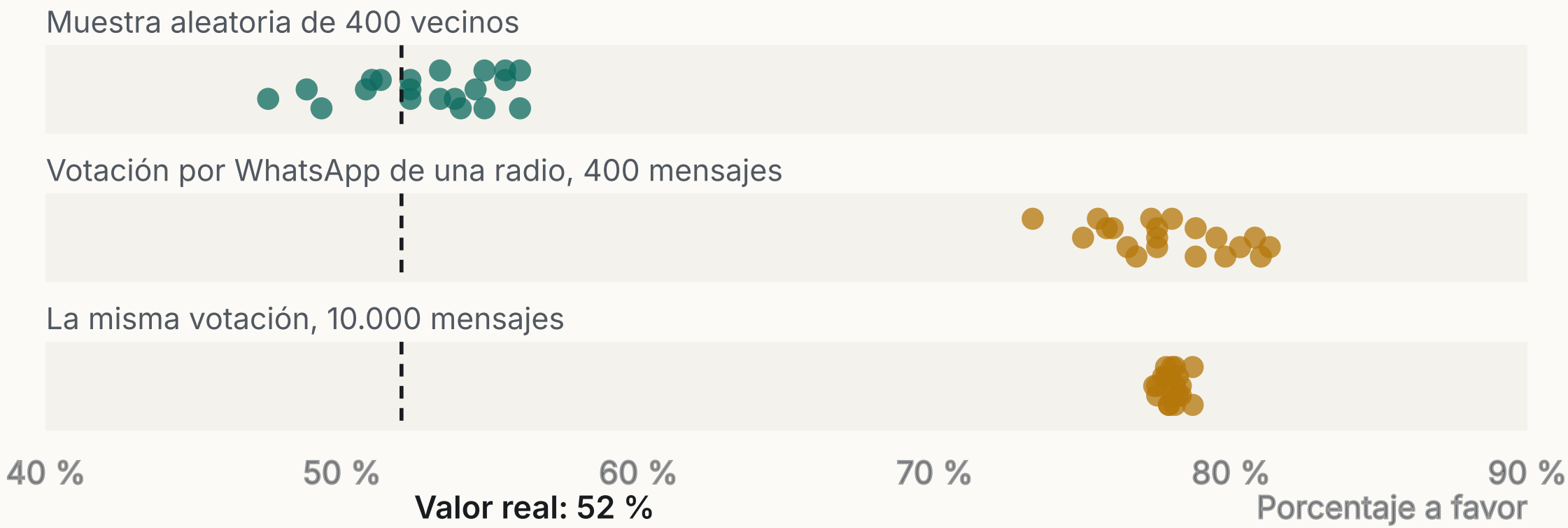
6

Error y sesgo

Por qué un resultado puede estar lejos de la verdad

Error aleatorio y sesgo

Veinte relevamientos simulados de cada tipo, en una ciudad donde el 52 % está a favor de cerrar el centro al tránsito.



Dos clases de error

Error aleatorio

- Viene de mirar solo una parte.
- No empuja para ningún lado.
- Se achica con muestras más grandes.
- Se puede medir: es el margen de error.

Sesgo

- Viene del diseño: a quiénes se pregunta y cómo.
- Empuja siempre para el mismo lado.
- No se arregla con más casos.
- De selección, de cobertura, de medición, de no respuesta.

Para llevarse

- La Estadística existe porque los datos varían.
- Antes de calcular: ¿cuál es la población, cuál la muestra y de qué es cada fila?
- El parámetro describe a la población y casi nunca se conoce; el estadístico se calcula con la muestra.
- El tipo y la escala de una variable deciden qué se puede hacer con ella.
- Más casos achican el error aleatorio, no el sesgo.

Guía de trabajo práctico 1
estadistica.ar/docentes/trabajos-practicos/semana-1

Escalas de medición
estadistica.ar/conceptos/escalas-de-medicion

La base de la EPH de Gran Catamarca
estadistica.ar/datos/eph-catamarca-2026t1

Error y sesgo
estadistica.ar/conceptos/error-y-sesgo