Trabajo práctico 1: Los datos y sus variables

Apellido y nombre: ..... Comisión: ..... Fecha: .....

AL TERMINAR ESTA GUÍA

- Reconocer en un relevamiento la población, la muestra, la unidad de análisis, las variables, el parámetro y el estadístico.
- Clasificar variables por tipo y escala de medición, y saber qué operaciones tienen sentido con cada una.
- Pasar información suelta a una matriz de datos y documentarla en un diccionario de variables, con los datos faltantes marcados.
- Abrir una base real, leer sus códigos antes de calcular y distinguir el error aleatorio del sesgo.

EJERCICIO 1 Cortes de agua en un municipio

A mano

Un municipio del interior tiene 14.000 hogares con conexión a la red de agua. Para saber cuántos sufrieron cortes el último mes, la oficina de estadística sorteó 500 hogares del padrón de la cooperativa de agua y fue a encuestarlos. En cada hogar respondió una persona mayor de edad, que contestó si hubo cortes, cuántos y cuántas horas duró el más largo.

De los 500 hogares, 259 dijeron haber tenido al menos un corte.

a) ¿Cuál es la población? ¿Cuál es la muestra? ¿Cuál es la unidad de análisis, y quién es el que responde?

b) Nombrá las variables que se relevaron y decí de qué tipo es cada una.

c) ¿Qué proporción de los hogares encuestados tuvo cortes? Expresala también en porcentaje.

d) El intendente declara: «El 51,8 % de los hogares del municipio tuvo cortes». ¿Ese número es un parámetro o un estadístico? ¿Cuántos hogares del municipio estimás que tuvieron cortes?

e) ¿Qué fracción de los hogares se encuestó? Si se hubieran sorteado otros 500 hogares, ¿el porcentaje daría lo mismo? ¿Qué cambia de una muestra a otra y qué no?

EJERCICIO 2 La historia clínica

A mano

La ficha de ingreso de un centro de salud registra estos datos de cada paciente:

Variable	Cómo se registra
Edad	Años cumplidos
Sexo	Varón o mujer
Grupo sanguíneo	0, A, B o AB
Temperatura	En °C, con un decimal
Peso	En kg, con un decimal
Consultas en el último año	Cantidad
Nivel educativo	Primario, secundario o superior; completo o no
Dolor	De 0 (nada) a 10 (el peor imaginable)
Obra social	Sí o no
DNI	Número de documento
Satisfacción con la atención	Muy mala, mala, regular, buena o muy buena

a) Clasificá cada variable: cualitativa o cuantitativa; si es cualitativa, nominal u ordinal; si es cuantitativa, discreta o continua. Indicá también su escala de medición: nominal, ordinal, de intervalo o de razón.

b) ¿Por qué la temperatura en grados Celsius no es de razón, y el peso sí? ¿Tiene sentido decir que un paciente con 40 °C tiene «el doble de temperatura» que algo a 20 °C?

c) El DNI se escribe con números y el grupo sanguíneo puede ser «0». ¿Los hace eso cuantitativos? Para cada variable de la tabla, ¿qué operación tiene sentido: contar, ordenar, restar o dividir?

d) La edad también se puede preguntar en tramos: menos de 20, de 20 a 39, de 40 a 59, 60 y más. ¿Qué tipo de variable queda? ¿Qué se pierde? ¿Se puede volver atrás?

EJERCICIO 3 De las fichas a la matriz de datos

A mano

Un equipo de enfermería anotó en papel los datos de seis pacientes, con sus palabras y sin un formato fijo:

Ficha 1. Varón de 41 años. Completó la primaria y no siguió. No tiene hijos. Pesa 69,8 kg. No tiene obra social.

Ficha 2. Mujer de 70 años. Empezó el secundario y no lo terminó. Tiene cuatro hijos. Pesa 61,7 kg. Tiene obra social.

Ficha 3. Mujer de 32 años. Terminó el secundario. Tiene un hijo. Pesa 59,7 kg. No tiene obra social.

Ficha 4. Mujer de 40 años. Completó la primaria y no siguió. No tiene hijos. Pesa 67,9 kg. Tiene obra social.

Ficha 5. Varón de 54 años. Está cursando una carrera en la universidad. Tiene un hijo. Pesa 87,8 kg. Tiene obra social.

Ficha 6. Mujer de 76 años. Completó la primaria y no siguió. Tiene dos hijos. No se registró el peso. Se atiende en el hospital público, sin obra social.

a) Armá la matriz de datos: una fila por paciente y una columna por variable. Decidí cómo cargar cada respuesta, con códigos numéricos donde haga falta.

b) Escribí el diccionario de variables: para cada columna, su nombre, qué mide, el tipo de variable y los valores posibles o la unidad.

c) La ficha 6 no trae el peso. ¿Qué cargás en esa celda? ¿Por qué no un 0?

d) Con la matriz armada, contestá: ¿cuántos pacientes tienen obra social? ¿Cuántos terminaron al menos el secundario? ¿Cuántas de las mujeres tienen hijos?

EJERCICIO 4 Una base real: la EPH de Gran Catamarca

Con software

La base «Ocupación e ingresos en Gran Catamarca» es un extracto real de la Encuesta Permanente de Hogares del INDEC, primer trimestre de 2026. Se baja de estadistica.ar/datos/eph-catamarca-2026t1, donde está la descripción de cada variable y de sus códigos. El separador es punto y coma.

Se puede usar R, Python, Excel, InfoStat o SPSS. Antes de calcular nada, leé la descripción de la base: la mitad del trabajo con datos reales es entender qué significa cada código.

a) ¿Cuántas filas y cuántas columnas tiene la base? ¿Cuál es la unidad de análisis? ¿Cuántas viviendas y cuántos hogares distintos hay?

b) Clasificá las doce variables. Separá los identificadores y los ponderadores, que no son variables para describir.

c) Armá la tabla de frecuencias de ESTADO. ¿Qué significan los códigos 0 y 4? Calculá el promedio de la columna ESTADO: ¿qué quiere decir ese número?

d) Algunos códigos no son valores. ¿Cuántas personas tienen CH06 = −1? ¿Cuántas tienen P21 = 0 y cuántas P21 = −9? ¿Qué significa cada uno?

e) ¿Cuántas personas terminaron al menos el secundario? Un compañero filtró NIVEL_ED ≥ 4. Mirá en la descripción qué significa el código 7 y corregí la cuenta.

f) Entre las personas de 10 años o más con entrevista (ESTADO 1, 2 o 3), ¿qué porcentaje está ocupado? Calculalo contando filas y después ponderando por PONDERA. ¿Cuál de los dos números estima la tasa de Gran Catamarca? ¿Cuántas personas representa la base?

EJERCICIO 5 ¿De dónde sale el dato?

Para discutir

Para cada pregunta, decidí de qué tipo de fuente saldría la respuesta: un censo, una encuesta por muestreo o un registro administrativo. Justificá, y decí qué limitación tiene esa fuente.

a) ¿Cuántos nacimientos hubo en la provincia el año pasado, por departamento de residencia de la madre?

b) ¿Cuál fue la tasa de desocupación de Gran Catamarca en el último trimestre?

c) ¿Cuántas personas viven en cada localidad de la provincia, por edad y sexo?

d) ¿Cuántos estudiantes de esta facultad trabajan más de 20 horas por semana?

e) Con los tres tipos de fuente a la vista: ¿cuál tiene error muestral? ¿Puede tener errores una fuente que cubre a todos?

EJERCICIO 6 ¿Qué opina la ciudad?

Para discutir

Se quiere saber qué proporción de los vecinos de una ciudad está a favor de cerrar el centro al tránsito de autos. En cada caso, decidí si el resultado sirve para estimar esa proporción y, si no sirve, qué falla.

a) Una radio pide a los oyentes que voten por WhatsApp. Llegan 2.400 mensajes y el 81 % está a favor. ¿Ese 81 % estima la opinión de la ciudad? ¿Se arreglaría con 10.000 mensajes?

b) Una consultora sortea 400 números de la guía de teléfonos fijos y llama.

c) El cuestionario pregunta: «¿No le parece que el centro sería más seguro y más lindo sin autos?».

d) Dos encuestas bien hechas, cada una con una muestra aleatoria de 400 vecinos, dan 54 % y 49 % a favor. ¿Alguna de las dos está sesgada?

---

estadistica.ar/docentes/trabajos-practicos/semana-1 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.