Trabajo práctico 4: Relación entre dos variables

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

AL TERMINAR ESTA GUÍA

- Construir una tabla de contingencia y calcular los porcentajes en el sentido que contesta la pregunta.
- Reconocer una variable de confusión y la paradoja de Simpson.
- Calcular e interpretar la covarianza, la correlación de Pearson, la recta de mínimos cuadrados y el R^2, y no extrapolar.
- Calcular las mismas medidas con software sobre datos reales, y distinguir correlación de causalidad.

EJERCICIO 1 Actividad física y edad

A mano

En una encuesta de salud a personas adultas de una ciudad se preguntó si hacen actividad física al menos tres veces por semana. Las respuestas, por grupo de edad:

Edad	Hace actividad	No hace
18 a 39	60	43
40 a 59	39	47
60 y más	18	55

a) Completá la tabla con los totales de cada fila y de cada columna. ¿Qué porcentaje del total hace actividad física?

b) ¿Qué porcentaje de las personas de 60 años y más hace actividad física? ¿Qué porcentaje de las que hacen actividad física tiene 60 años y más? ¿Por qué no da lo mismo?

c) ¿La actividad física cambia con la edad? Calculá los porcentajes que permiten contestarlo y decí en qué sentido de la tabla los calculaste.

EJERCICIO 2 ¿Qué hospital opera mejor?

A mano

Dos hospitales informan cuántos pacientes operaron de la misma cirugía el último año y cuántos sobrevivieron. El A es el hospital de referencia de la provincia, al que se derivan los casos complicados:

Hospital	Estado al ingresar	Operados	Sobrevivieron
A	Leve	52	49
A	Grave	209	141
B	Leve	209	185
B	Grave	44	26

a) Sin separar por estado, ¿qué porcentaje de los operados sobrevivió en cada hospital? ¿Cuál parece mejor?

b) Ahora calculalo por separado para los pacientes leves y para los graves. ¿Qué hospital tiene mejor resultado en cada grupo?

c) ¿Cómo puede ser? Mirá qué proporción de pacientes graves recibió cada hospital.

d) Si tuvieras que operarte, ¿qué hospital elegirías? ¿Qué comparación es la justa?

EJERCICIO 3 Horas de estudio y nota

A mano

Ocho estudiantes anotaron cuántas horas estudiaron para el primer parcial. Después se registró la nota que sacaron:

Estudiante	1	2	3	4	5	6	7	8
Horas de estudio	6	1	10	15	2	5	10	7
Nota	4	3	8	10	2	6	8	7

a) Dibujá el diagrama de dispersión, con las horas en el eje horizontal. Antes de calcular nada, ¿cuánto dirías que vale el coeficiente de correlación?

b) Calculá la covarianza y el coeficiente de correlación de Pearson. Armá la tabla de desvíos y productos.

c) Calculá la recta de mínimos cuadrados para predecir la nota a partir de las horas. ¿Qué significan la pendiente y la ordenada al origen?

d) Calculá el coeficiente de determinación e interpretalo. ¿Qué nota predice la recta para alguien que estudió 8 horas? ¿Y para alguien que estudió 40?

EJERCICIO 4 Informalidad e ingreso en 32 aglomerados

Con software

La base «Treinta y dos aglomerados por diez indicadores» resume la EPH del primer trimestre de 2026: una fila por aglomerado urbano del país, con tasas de empleo y desocupación, ingresos, educación, informalidad y edad. Se baja de estadistica.ar/datos/indicadores-aglomerados. El separador es la coma y el punto es el decimal.

Interesan no_registrado, el porcentaje de asalariados sin descuento jubilatorio, e ingreso_mediano, el ingreso mediano de la ocupación principal en pesos. Se puede usar R, Python, Excel, InfoStat o SPSS.

a) Hacé el diagrama de dispersión del ingreso mediano (eje vertical) contra el porcentaje de asalariados no registrados (eje horizontal). Calculá el coeficiente de correlación y describí la relación.

b) Calculá la recta de mínimos cuadrados y el R^2. Interpretá la pendiente en pesos.

c) Gran Catamarca es el aglomerado 22. ¿Qué ingreso mediano predice la recta para su nivel de informalidad? ¿Está por encima o por debajo de lo esperado?

d) Calculá también la correlación entre secundario_mas y desocupacion, y entre edad_media y prop_65_mas. ¿Qué dice cada una? ¿Alguna de las tres correlaciones dice algo sobre las personas?

EJERCICIO 5 Varones, mujeres y trabajo en Gran Catamarca

Con software

Con la base de la EPH de Gran Catamarca (estadistica.ar/datos/eph-catamarca-2026t1), quedate con las personas de 10 años o más con entrevista: ESTADO igual a 1 (ocupado), 2 (desocupado) o 3 (inactivo). Interesa la relación entre CH04, el sexo, y ESTADO.

a) Armá la tabla de contingencia de sexo por condición de actividad, con frecuencias absolutas y totales.

b) Calculá los porcentajes por fila. ¿Qué diferencia hay entre varones y mujeres?

c) La tasa de desocupación es el porcentaje de desocupados entre los activos (ocupados más desocupados), no entre todos. Calculala para cada sexo. ¿Por qué no se divide por el total de la fila?

d) Un informe dice: «el 62 % de los inactivos son mujeres, así que las mujeres trabajan menos». ¿Qué porcentaje calculó el informe? ¿Es el que corresponde para esa conclusión?

EJERCICIO 6 ¿Una cosa causa la otra?

Para discutir

En cada caso hay una correlación real. Decidí si alcanza para afirmar que una variable causa la otra y, si no, qué otra explicación hay.

a) En los meses en que se venden más helados hay más personas ahogadas en ríos y piletas.

b) En los incendios a los que acuden más bomberos, los daños son mayores.

c) Los estudiantes que desayunan todos los días tienen mejores notas que los que no.

d) En un ensayo clínico, se sorteó qué pacientes recibían un medicamento y cuáles un placebo. Los que recibieron el medicamento mejoraron más.

e) Con los cuatro casos a la vista: ¿cuándo una correlación permite hablar de causa?

---

estadistica.ar/docentes/trabajos-practicos/semana-4 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.