Trabajo práctico 3: Posición, dispersión y forma

AL TERMINAR ESTA GUÍA

- Calcular a mano las medidas de posición, dispersión y forma de un conjunto chico de datos, y controlar las cuentas.
- Elegir entre la media y la mediana según la forma de la distribución y la presencia de atípicos.
- Comparar variabilidades con el coeficiente de variación y posiciones relativas con el puntaje z.
- Calcular las mismas medidas con software sobre una base real, y decidir con criterio qué hacer con un valor atípico.

PARA EL DOCENTE

Los ejercicios 1 a 5 son para el encuentro práctico, con calculadora; el 1 y el 2 alcanzan para una hora. El 6 necesita computadora: sirve para un encuentro virtual o como tarea.

El 7 funciona mejor como cierre, discutido en grupos, después de haber hecho el 2 y el 6.

EJERCICIO 1 Tiempos de espera en una guardia

A mano

En la guardia de un hospital se registró cuántos minutos esperó cada uno de los diez pacientes que llegaron entre las 8 y las 10 de la mañana, desde que lo anotaron hasta que lo atendieron:

Minutos de espera	11	46	5	41	6	22	8	10	6	25

a) Calculá la media, la mediana y el modo.

RESOLUCIÓN

Ordenados: 5, 6, 6, 8, 10, 11, 22, 25, 41, 46.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{180}{10} = 18 \text{ minutos}$$

$$\text{Me} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{10 + 11}{2} = 10{,}5 \text{ minutos}$$

El modo es 6 minutos: es el único valor que se repite.

b) Calculá el primer y el tercer cuartil, y el rango intercuartílico.

RESOLUCIÓN

Con el método de R y de Excel (CUARTIL.INC), el cuantil p está en la posición 1 + (n − 1)p de los datos ordenados:

$Q_1:\ 1 + 9 \cdot 0{,}25 = 3{,}25 \ \Rightarrow\ 6 + 0{,}25\,(8 - 6) = 6{,}50$

$Q_3:\ 1 + 9 \cdot 0{,}75 = 7{,}75 \ \Rightarrow\ 22 + 0{,}75\,(25 - 22) = 24{,}25$

$\text{RIC} = Q_3 - Q_1 = 24{,}25 - 6{,}50 = 17{,}75 \text{ minutos}$

SPSS y la mayoría de los manuales usan la posición (n + 1)p, que da Q_1 = 6,00, Q_3 = 29,00 y RIC = 23,00. Las dos convenciones son correctas; con pocos datos se nota la diferencia, con muchos casi no.

c) Calculá el rango, la varianza y el desvío estándar. Armá la tabla de desvíos respecto de la media.

RESOLUCIÓN

$\text{Rango} = 46 - 5 = 41 \text{ minutos}$

x	x − x̄	(x − x̄)²
5	−13	169
6	−12	144
6	−12	144
8	−10	100
10	−8	64
11	−7	49
22	4	16
25	7	49
41	23	529
46	28	784
Σ = 180	**0**	**2.048**

La columna del medio suma cero, como siempre: sirve de control antes de seguir.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{2.048}{9} = 227{,}56 \text{ minutos}^2 \qquad s = \sqrt{227{,}56} = 15{,}08 \text{ minutos}$$

d) Calculá el coeficiente de variación e interpretalo.

RESOLUCIÓN

$$\text{CV} = \frac{s}{\bar{x}} = \frac{15{,}08}{18} = 0{,}8381 \approx 83{,}8\ \%$$

El desvío es 83,8 % de la media: los tiempos de espera son muy dispares entre pacientes. Por encima de 30 % se suele hablar de una dispersión alta.

e) ¿Qué medida de posición informarías para describir la espera típica? ¿Qué dice la diferencia entre la media y la mediana sobre la forma de la distribución?

RESOLUCIÓN

La media (18) supera a la mediana (10,5): unas pocas esperas largas tiran de la media hacia arriba. La distribución es asimétrica a la derecha, que es lo esperable en tiempos de espera: no pueden bajar de cero pero sí estirarse mucho.

Para la espera típica conviene la mediana: la mitad de los pacientes esperó 10,5 minutos o menos. La media sigue siendo útil si lo que interesa es el total (diez pacientes por 18 minutos es el tiempo total de espera acumulado).

EJERCICIO 2 Retiros mensuales en una cooperativa

A mano

Una cooperativa de trabajo de nueve integrantes publica lo que retiró cada uno el último mes, en miles de pesos. Uno de ellos es el presidente, que además coordina la venta a mayoristas:

Retiro (miles de $)	780	750	740	2.800	850	870	800	680	880

a) Calculá la media y la mediana de los nueve retiros.

RESOLUCIÓN

Ordenados: 680, 740, 750, 780, 800, 850, 870, 880, 2.800.

$$\bar{x} = \frac{9.150}{9} = 1.016{,}67 \qquad \text{Me} = x_{(5)} = 800$$

b) Calculá los cuartiles y los límites de la regla de 1,5 RIC. ¿Hay valores atípicos?

RESOLUCIÓN

Con nueve datos, el método de R y de Excel cae justo sobre un dato: la posición 1 + 8 · 0,25 = 3 para Q_1 y 1 + 8 · 0,75 = 7 para Q_3.

$$Q_1 = 750 \qquad Q_3 = 870 \qquad \text{RIC} = 120$$

$$Q_1 - 1{,}5\,\text{RIC} = 570{,}00 \qquad Q_3 + 1{,}5\,\text{RIC} = 1.050{,}00$$

El retiro de 2.800 queda por encima del límite superior: es atípico. Ningún otro valor sale del cerco.



En miles de $. La caja va de Q_1 a Q_3 y la línea es la mediana; los bigotes llegan al último dato dentro de 1,5 RIC y los círculos son los atípicos.

c) Sacá el retiro más alto y volvé a calcular la media y la mediana. ¿Cuál de las dos cambió más? ¿Por qué?

RESOLUCIÓN

$$\bar{x}_{\text{sin}} = \frac{6.350}{8} = 793{,}75 \qquad \text{Me}_{\text{sin}} = \frac{x_{(4)} + x_{(5)}}{2} = \frac{780 + 800}{2} = 790{,}00$$

> La media bajó 222,92 mil pesos; la mediana, 10,00. La media usa el valor de cada dato, así que un solo retiro muy alto la arrastra; la mediana solo usa el orden, y al dato más alto le da lo mismo valer 2.800 o 890.

d) Si tuvieras que informar el «retiro típico» de la cooperativa en una línea, ¿qué número darías? ¿Corresponde sacar el retiro del presidente del cálculo?

> RESOLUCIÓN
>
> La mediana, 800 mil pesos: es lo que retira el integrante del medio, y no depende de cuánto retire el presidente.
>
> Sacarlo no corresponde: no es un error, es un integrante real de la cooperativa. Un atípico se saca cuando es un error de registro o no pertenece a la población que se quiere describir; si es un dato válido, se lo deja y se eligen medidas que no dependan de él. Lo honesto es informar la mediana y aclarar el rango, o la media con y sin ese caso.

EJERCICIO 3 Peso y talla al nacer

A mano

En una maternidad, los recién nacidos de término del último año tuvieron estas medidas:

Variable	Media	Desvío estándar
Peso (gramos)	3.230	490
Talla (centímetros)	50,5	1,9

a) El desvío del peso es mucho mayor que el de la talla. ¿Alcanza eso para decir que el peso varía más? ¿Por qué?

> RESOLUCIÓN
>
> No. Los desvíos están en unidades distintas, gramos y centímetros, y no se pueden comparar: el desvío del peso pasaría a ser mil veces menor con solo expresarlo en kilos. Para comparar la variabilidad de variables con distinta unidad, o con medias muy distintas, hace falta una medida sin unidad.

b) Calculá el coeficiente de variación de cada variable. ¿Cuál es relativamente más variable?

> RESOLUCIÓN
>
> $$\mathrm{CV}_{\text{peso}} = \frac{490}{3.230} = 0{,}1517 \approx 15{,}2\,\% \qquad \mathrm{CV}_{\text{talla}} = \frac{1{,}9}{50{,}5} = 0{,}0376 \approx 3{,}8\,\%$$
>
> El peso es relativamente 4,0 veces más variable que la talla. Tiene sentido: la talla al nacer está más acotada por la edad gestacional que el peso.

c) Si el peso se expresara en kilos, ¿qué pasaría con la media, el desvío y el coeficiente de variación?

> RESOLUCIÓN
>
> La media y el desvío se dividen por mil (3,23 kg y 0,49 kg) y el coeficiente de variación no cambia, porque es un cociente: el factor de escala se cancela. Por eso sirve para comparar.

EJERCICIO 4 Dos comisiones, el mismo parcial

A mano

Dos comisiones rindieron el mismo parcial con docentes distintos. Ana está en la comisión A y Bruno en la B:

Comisión	Media	Desvío estándar	Nota
A	6,7	1,5	Ana: 8,5
B	4,7	1,9	Bruno: 8,0

a) Calculá el puntaje z de cada uno. ¿Quién se destacó más dentro de su comisión?

> RESOLUCIÓN
>
> $$z_{\text{Ana}} = \frac{8{,}5 - 6{,}7}{1{,}5} = 1{,}20 \qquad z_{\text{Bruno}} = \frac{8{,}0 - 4{,}7}{1{,}9} = 1{,}74$$
>
> Ana sacó más nota, pero Bruno se destacó más: está a 1,74 desvíos por encima de la media de su comisión, contra 1,20 de Ana.

b) ¿Qué nota tendría que haber sacado un estudiante de la comisión B para quedar en la misma posición relativa que Ana?

> RESOLUCIÓN
>
> $$x = \bar{x}_B + z_{\text{Ana}} \cdot s_B = 4{,}7 + 1{,}20 \cdot 1{,}9 = 6{,}98$$

c) ¿Qué supuesto hay detrás de comparar puntajes z de dos comisiones distintas?

> RESOLUCIÓN
>
> Que los dos grupos sean comparables: que la diferencia de medias y desvíos venga de cómo se corrigió o se enseñó, y no de que una comisión tenga estudiantes con otra preparación. El z dice dónde está alguien dentro de su grupo; si los grupos no son equivalentes, estar arriba en uno no significa lo mismo que estar arriba en el otro.

EJERCICIO 5 El promedio de toda la materia

A mano

La misma materia se dicta en tres comisiones. Cada docente informó cuántos estudiantes rindieron el primer parcial y el promedio que obtuvieron:

Comisión	Rindieron	Promedio
Mañana	16	6,8
Tarde	44	5,0
Noche	42	5,3

a) Calculá el promedio de todos los estudiantes que rindieron.

RESOLUCIÓN

$$\bar{x} = \frac{\sum n_i \bar{x}_i}{\sum n_i} = \frac{108{,}8 + 220{,}0 + 222{,}6}{102} = \frac{551{,}4}{102} = 5{,}41$$

b) Un compañero propone promediar los tres promedios y le da 5,70. ¿Por qué no es el promedio de la materia? ¿En qué caso coincidiría?

RESOLUCIÓN

Porque le da el mismo peso a cada comisión, tenga los estudiantes que tenga: un estudiante de la comisión más chica cuenta más que uno de la más grande. El promedio de la materia pondera cada promedio por la cantidad de estudiantes.

Las dos cuentas coinciden solo si las tres comisiones tienen la misma cantidad de estudiantes, o si los tres promedios son iguales.

EJERCICIO 6 Ingresos en 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: 1.339 personas. Se baja de estadistica.ar/datos/eph-catamarca-2026t1, con la descripción de cada variable. El separador es punto y coma.

Interesan P21, el ingreso de la ocupación principal en pesos; CH04, el sexo (1 = varón, 2 = mujer), y ESTADO, la condición de actividad (1 = ocupado). En P21, el 0 significa «no corresponde» y el −9, «no respondió». Se puede usar R, Python, Excel, InfoStat o SPSS.

a) Calculá el promedio de la columna P21 entera. Después quedate con las personas ocupadas que declararon su ingreso (ESTADO = 1 y P21 > 0). ¿Cuántas quedan? ¿Por qué el primer promedio no describe a nadie?

RESOLUCIÓN

El promedio de la columna entera da $ 391.264. Mezcla 715 ceros de personas sin ocupación —que no ganan cero: la pregunta no les corresponde— y 30 códigos −9 de no respuesta, que no son montos.

Con el filtro quedan 594 personas: 323 varones y 271 mujeres.

b) Para varones y mujeres por separado, calculá n, media, mediana, primer y tercer cuartil, rango intercuartílico, desvío estándar, coeficiente de variación y coeficiente de asimetría.

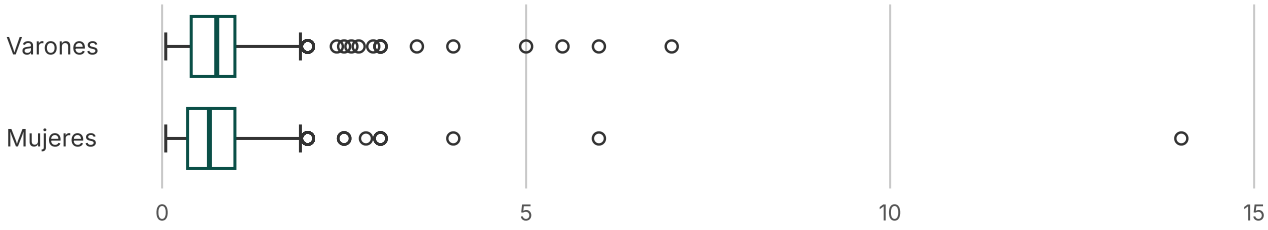
RESOLUCIÓN

	Varones	Mujeres
n	323	271
Media	$ 894.560	$ 867.011
Mediana	$ 750.000	$ 650.000
Q_1	$ 400.000	$ 350.000
Q_3	$ 1.000.000	$ 1.000.000
RIC	$ 600.000	$ 650.000
Desvío estándar	$ 816.306	$ 1.062.148
CV	91,3 %	122,5 %
Asimetría	3,61	7,76

Los cuartiles dan lo mismo con el método de R y Excel que con el de SPSS: con tantos ingresos declarados en números redondos, los dos caen sobre el mismo valor. La asimetría es la que informan Excel y SPSS, con la corrección por tamaño de muestra.

c) Hacé un diagrama de caja de los ingresos de cada sexo, uno al lado del otro. ¿Cuántos atípicos marca la regla de 1,5 RIC en cada grupo? ¿Qué dice el gráfico sobre la forma de la distribución?

RESOLUCIÓN



En millones de pesos. La caja va de Q_1 a Q_3 y la línea es la mediana; los bigotes llegan al último dato dentro de 1,5 RIC y los círculos son los atípicos.

Por encima del límite superior hay 23 varones (más de $ 1.900.000) y 23 mujeres (más de $ 1.975.000). Las dos distribuciones son muy asimétricas a la derecha: la caja está apretada abajo, el bigote superior es más largo y los atípicos se estiran hasta varios millones. La asimetría, 3,61 y 7,76, lo confirma.

Tantos atípicos no son errores: es la forma típica del ingreso. La regla del 1,5 está pensada para distribuciones más o menos simétricas, y en una con cola larga marca como atípicos a valores que son parte normal de la distribución.

d) El ingreso más alto de la base es $ 14.000.000. ¿Es un error? Recalculá la media y la mediana del grupo al que pertenece sin ese caso. ¿Qué cambia?

RESOLUCIÓN

Es el ingreso de una persona del grupo de mujeres. No hay manera de saber desde la base si es un error, y no es imposible: un ingreso así existe en Catamarca. Sin una razón para pensar que está mal cargado, se queda.

$\bar{x}$: 867.011 $\rightarrow$ 818.370 Me : 650.000 $\rightarrow$ 650.000

Un solo caso entre 271 mueve la media $ 48.641 y la mediana nada. Es el mismo fenómeno del ejercicio de la cooperativa, con datos reales.

e) ¿Cuánto más ganan los varones que las mujeres? Contestá con la media y con la mediana. ¿Por qué dan distinto, y cuál informarías?

RESOLUCIÓN

Con la media, los varones ganan 3,2 % más; con la mediana, 15,4 % más.

Dan distinto porque la media de las mujeres está empujada por unos pocos ingresos muy altos, que achican la diferencia. La mediana describe a la persona del medio de cada grupo y no depende de esos casos: es la más adecuada para comparar ingresos típicos. Lo mejor es informar las dos, porque contar solo una cuenta media historia.

Ninguna de las dos alcanza para hablar de discriminación salarial: las mujeres ocupadas pueden trabajar menos horas o en otras ramas. Eso se discute en la unidad de relación entre variables.

f) Para ir más lejos: la EPH es una muestra y cada persona representa a otras. Calculá la media de cada grupo ponderada por PONDIIO, el ponderador del ingreso. ¿Cambian las conclusiones?

RESOLUCIÓN

Ponderadas, las medias son $ 907.483 para varones y $ 862.288 para mujeres, contra $ 894.560 y $ 867.011 sin ponderar. En R: weighted.mean(P21, PONDIIO); en Excel: SUMAPRODUCTO(P21; PONDIIO) / SUMA(PONDIIO).

La diferencia es chica en esta base, pero el camino correcto para estimar el ingreso de la población es el ponderado: la media sin ponderar describe a las personas encuestadas, no a Gran Catamarca.

EJERCICIO 7 ¿Lo saco o lo dejo?

Para discutir

En cada caso apareció un valor que se aleja del resto. Decidí qué harías con él y justificá. No hay una regla única: la decisión depende de qué es el dato y de qué se quiere describir.

a) En una encuesta a estudiantes universitarios, una persona tiene cargada una edad de 230 años.

RESOLUCIÓN

Es un error: nadie tiene 230 años. Si se puede, se vuelve al cuestionario y se corrige (quizás era 23); si no, ese valor se trata como dato faltante. Lo que no se hace es dejarlo, ni borrar la fila entera: el resto de sus respuestas puede ser válido.

b) En un estudio de tiempos de reacción, un participante respondió en 0,02 segundos, cuando el resto tardó entre 0,25 y 0,60.

> RESOLUCIÓN
>
> Ningún ser humano reacciona a un estímulo en 0,02 segundos: el participante apretó antes de ver el estímulo. No mide lo que el estudio quiere medir, así que se excluye. Lo correcto es que el criterio (por ejemplo, descartar respuestas de menos de 0,10 s) esté escrito antes de mirar los datos, y que el informe diga cuántos casos se excluyeron.

c) En la base de la EPH, una persona declara un ingreso de varios millones, muy por encima del resto.

> RESOLUCIÓN
>
> Es un dato posible y no hay evidencia de error: se queda. Sacarlo haría que la estadística describa una población sin ingresos altos, que no existe. Lo que cambia es qué medidas se informan: mediana y rango intercuartílico en lugar de media y desvío, o las dos con la aclaración.

d) Con los tres casos a la vista, ¿qué regla general propondrías para decidir qué hacer con un valor atípico?

> RESOLUCIÓN
>
> Primero preguntar qué es el dato. Si es un error de carga, se corrige o se trata como faltante. Si no pertenece a la población que se estudia, se excluye con un criterio explícito. Si es un valor real, se queda, y se eligen medidas resistentes o se informan los resultados con y sin él. La regla de 1,5 RIC sirve para encontrar los casos que hay que mirar, no para decidir por sí sola cuáles se borran.

---

estadistica.ar/docentes/trabajos-practicos/semana-3 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.