Trabajo práctico 2: Frecuencias y gráficos

AL TERMINAR ESTA GUÍA

- Construir tablas de frecuencias absolutas, relativas y acumuladas, para datos sin agrupar y agrupados en intervalos.
- Leer en una tabla proporciones como «a lo sumo» y «al menos», y saber qué se pierde al agrupar.
- Elegir el gráfico según el tipo de variable y lo que se quiere mostrar, y dibujarlo sin distorsionar los datos.
- Armar tablas e histogramas con software sobre una base real, respetando sus códigos.

PARA EL DOCENTE

Los ejercicios 1 a 3 son para el encuentro práctico, con papel cuadriculado para los gráficos; el 1 y el 2 alcanzan para una hora. El 4 necesita computadora: sirve para un encuentro virtual o como tarea.

El 5 funciona bien si los estudiantes traen un gráfico publicado en un diario o una red social para analizar entre todos. El 6 se puede hacer de a dos, en diez minutos.

EJERCICIO 1 Hijos por hogar

A mano

En una encuesta a 30 hogares de un barrio se preguntó cuántos hijos menores de 18 años viven en el hogar. Las respuestas, en el orden en que se registraron:

Hogares 1 a 10	1	1	2	3	4	2	2	3	1	0
Hogares 11 a 20	2	5	1	0	1	0	3	1	2	1
Hogares 21 a 30	3	2	1	2	2	3	1	1	3	2

a) Armá la tabla de frecuencias: absoluta, relativa, absoluta acumulada y relativa acumulada.

RESOLUCIÓN

Hijos	f	fr	F	Fr
0	3	0,100	3	0,100
1	10	0,333	13	0,433
2	9	0,300	22	0,733
3	6	0,200	28	0,933
4	1	0,033	29	0,967
5	1	0,033	30	1,000
Total	**30**	**1,000**		

Dos controles: las frecuencias absolutas suman 30, las relativas suman 1, y la última acumulada es el total. Contar con palotes sobre la lista desordenada es más seguro que ordenar primero.

b) ¿Qué porcentaje de los hogares tiene a lo sumo dos hijos? ¿Y al menos tres? ¿Qué cantidad de hijos es la más frecuente?

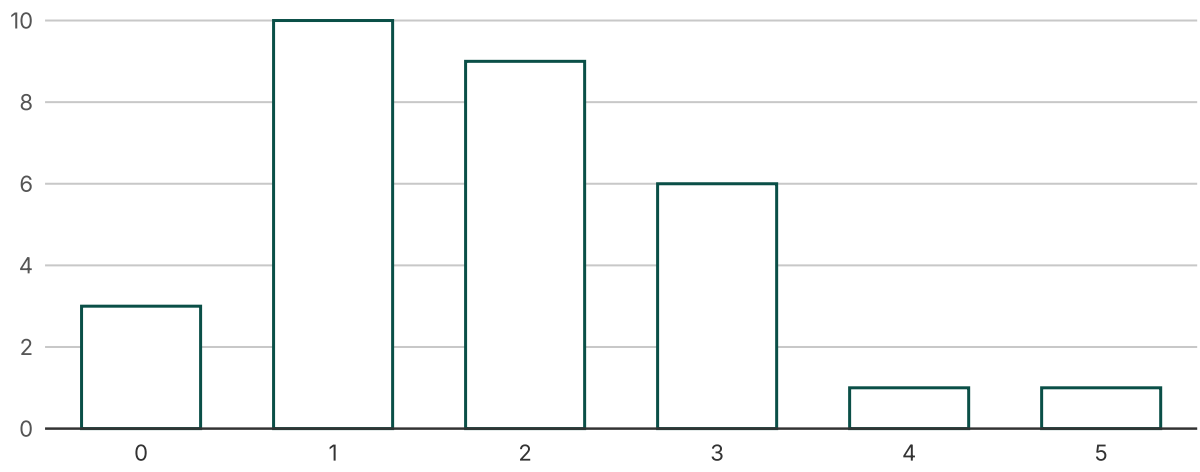
RESOLUCIÓN

A lo sumo dos es la acumulada en el 2: 22 de 30, 73,3 %. Al menos tres es el resto: 8 hogares, 26,7 %. Lo más frecuente son 1 hijo, con 10 hogares.

«A lo sumo» incluye el valor (≤ 2) y «al menos» también (≥ 3): son complementarios, por eso suman 100 %.

c) Graficá la distribución. ¿Por qué las barras van separadas y no pegadas como en un histograma?

RESOLUCIÓN



Eje vertical: cantidad de hogares.

La variable es discreta: entre 1 y 2 hijos no hay valores posibles, y las barras separadas lo muestran. El histograma, con las barras pegadas, es para variables continuas, donde cada barra cubre un intervalo sin huecos entre uno y otro.

EJERCICIO 2 Tiempo de viaje a la facultad

A mano

Se les preguntó a 40 estudiantes de primer año cuántos minutos tardan en llegar a la facultad desde su casa:

Estudiantes 1 a 10	21	8	8	8	67	11	27	70	29	12
Estudiantes 11 a 20	25	9	56	11	13	16	11	27	11	18
Estudiantes 21 a 30	21	34	19	16	11	27	30	21	19	11
Estudiantes 31 a 40	55	17	12	30	9	44	16	50	6	13

a) Agrupá los datos en intervalos de 10 minutos, desde 0: [0, 10), [10, 20), etcétera. Armá la tabla con la marca de clase, la frecuencia absoluta, la relativa y las dos acumuladas.

RESOLUCIÓN

Minutos	Marca	f	fr	F	Fr
[0, 10)	5	6	0,150	6	0,150
[10, 20)	15	17	0,425	23	0,575
[20, 30)	25	8	0,200	31	0,775
[30, 40)	35	3	0,075	34	0,850
[40, 50)	45	1	0,025	35	0,875
[50, 60)	55	3	0,075	38	0,950
[60, 70)	65	1	0,025	39	0,975
[70, 80)	75	1	0,025	40	1,000
Total		**40**	**1,000**		

El corchete incluye el límite y el paréntesis lo excluye: un estudiante que tarda justo 20 minutos va en [20, 30). Sin esa convención, un valor que cae en el borde se cuenta dos veces o ninguna.

¿Por qué 8 intervalos? La regla de Sturges sugiere $1 + \log_2(40) \approx 6{,}3$, o sea unos 6 o 7; con intervalos de 10 minutos, que se leen fácil, quedan 8. La regla es una orientación, no una obligación.

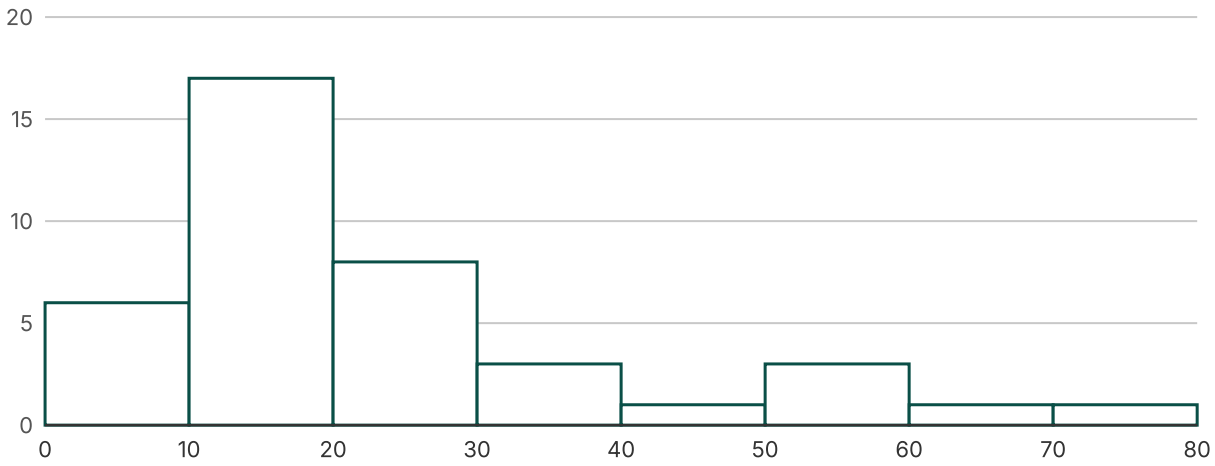
b) ¿Qué porcentaje de los estudiantes tarda menos de 30 minutos? ¿Y 50 minutos o más?

RESOLUCIÓN

Menos de 30 es la acumulada hasta [20, 30): 31 de 40, 77,5 %. Cincuenta o más es lo que queda después de [40, 50): 5 de 40, 12,5 %.

c) Dibujá el histograma. ¿Cómo es la forma de la distribución?

RESOLUCIÓN



Eje vertical: cantidad de estudiantes.

Asimétrica a la derecha: la mayoría tarda poco y unos pocos, que viven lejos, estiran la cola hacia los tiempos largos. Es la forma típica de los tiempos: no pueden bajar de cero pero sí estirarse mucho.

d) Solo con la tabla agrupada, estimá cuántos estudiantes tardan menos de 25 minutos. Después contalo en los datos originales. ¿Por qué no coinciden?

RESOLUCIÓN

Con la tabla se supone que los datos se reparten parejos dentro de cada intervalo: 25 es la mitad de [20, 30), así que se suma la mitad de esa clase.

$$F(25) \approx F(20) + \frac{25 - 20}{10} \cdot f_{[20,\,30)} = 23 + 0{,}5 \cdot 8 = 27{,}0$$

En los datos son 26. No coinciden porque dentro de [20, 30) los valores no están repartidos parejos. Al agrupar se pierde el detalle: la tabla es más fácil de leer, pero ya no se puede volver al dato individual.

EJERCICIO 3 Motivos de consulta en dos centros de salud

A mano

Dos centros de atención primaria registraron el motivo principal de cada consulta durante una semana:

Motivo	Centro A	Centro B
Control de salud	63	24
Respiratoria	58	27
Hipertensión	40	10
Gastrointestinal	17	8
Traumatismo	28	8
Otros	22	3

a) Calculá el total de cada centro y la frecuencia relativa de cada motivo, en porcentaje.

RESOLUCIÓN

Motivo	A	% A	B	% B
Control de salud	63	27,6	24	30,0
Respiratoria	58	25,4	27	33,8
Hipertensión	40	17,5	10	12,5
Gastrointestinal	17	7,5	8	10,0
Traumatismo	28	12,3	8	10,0
Otros	22	9,6	3	3,8
Total	**228**	**100,0**	**80**	**100,0**

Los porcentajes redondeados pueden sumar 99,9 o 100,1: no es un error, y se aclara o se ajusta el más grande.

b) ¿En qué centro hubo más consultas respiratorias? ¿En cuál pesan más dentro del total? ¿Por qué las dos respuestas no son la misma?

RESOLUCIÓN

En el A hubo más: 58 contra 27. Pero en el B pesan más: 33,8 % de sus consultas contra 25,4 % en el A.

El A atiende muchas más consultas (228 contra 80), así que tiene más de casi todo. Para comparar grupos de distinto tamaño se usan las frecuencias relativas; las absolutas dicen cuánto trabajo hay, no cómo se reparte.

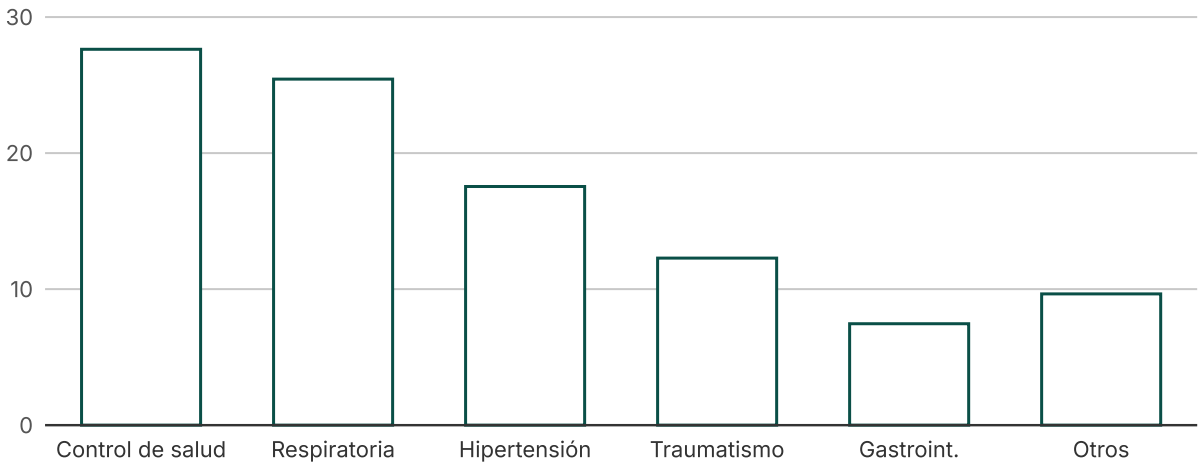
c) ¿Tiene sentido calcular frecuencias acumuladas para el motivo de consulta? ¿Por qué?

RESOLUCIÓN

No. Acumular es sumar «hasta esta categoría», y eso supone un orden. El motivo de consulta es una variable nominal: el orden de las filas es arbitrario, y con otro orden las acumuladas cambiarían. Las acumuladas tienen sentido en variables ordinales y cuantitativas.

d) Graficá los porcentajes del centro A. ¿En qué orden conviene poner las barras? ¿Usarías un gráfico de sectores?

RESOLUCIÓN



Eje vertical: porcentaje de las consultas.

En una variable nominal, de la más frecuente a la menos frecuente, con «Otros» al final aunque sea grande: así se lee el ranking de un vistazo. Un gráfico de sectores también vale para seis categorías que suman el 100 %, pero el ojo compara peor ángulos que alturas: si dos motivos tienen porcentajes parecidos, en las barras se nota cuál es mayor y en los sectores no. Para comparar los dos centros, barras agrupadas o apiladas al 100 %.

EJERCICIO 4 Educación, edad e ingreso en Gran Catamarca

Con software

Se usa la misma base de la semana pasada: «Ocupación e ingresos en Gran Catamarca», extracto de la EPH del primer trimestre de 2026, que se baja de estadistica.ar/datos/eph-catamarca-2026t1. Interesan NIVEL_ED (máximo nivel educativo), CH06 (edad) y P21 (ingreso de la ocupación principal).

Se puede usar R, Python, Excel, InfoStat o SPSS. Recordá los códigos especiales: en NIVEL_ED el 7 es «sin instrucción»; en CH06 el −1 es «menos de un año»; en P21 el 0 y el −9 no son ingresos.

a) Armá la tabla de frecuencias de NIVEL_ED con las categorías en el orden de los niveles, del más bajo al más alto, con porcentaje y porcentaje acumulado. ¿Qué porcentaje terminó al menos el secundario?

RESOLUCIÓN

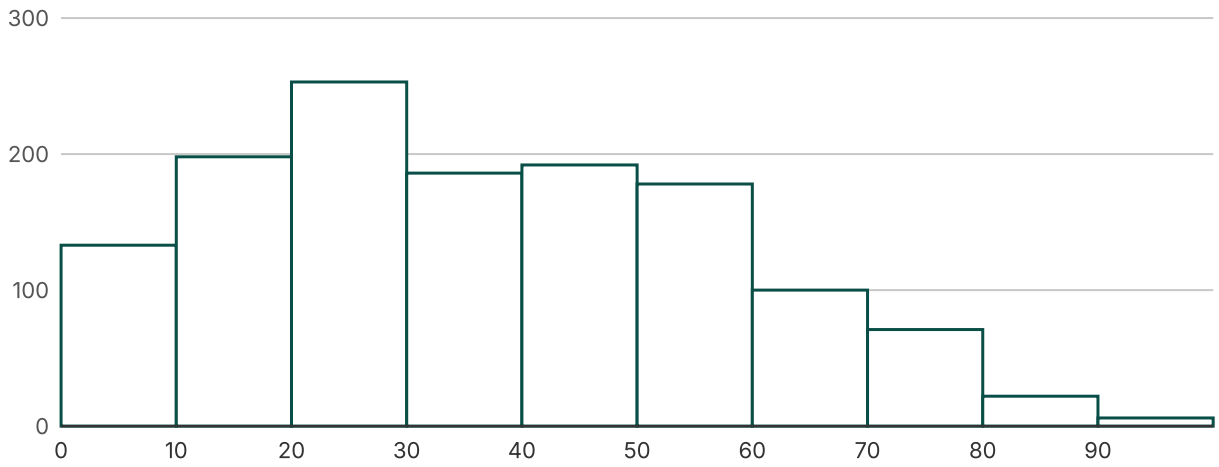
Nivel educativo	Personas	%	% acumulado
7 · Sin instrucción	54	4,0	4,0
1 · Primario incompleto	142	10,6	14,6
2 · Primario completo	126	9,4	24,0
3 · Secundario incompleto	252	18,8	42,9
4 · Secundario completo	349	26,1	68,9
5 · Superior incompleto	192	14,3	83,3
6 · Superior completo	224	16,7	100,0
Total	**1.339**	**100,0**	

Hasta secundario incompleto acumula 42,9 %; el resto, 57,1 %, terminó al menos el secundario. Si la tabla sale en el orden de los códigos, con el 7 al final, la columna acumulada no significa nada: hay que reordenar antes de acumular. En R: factor(NIVEL_ED, levels = c(7, 1:6)).

b) Agrupá la edad en intervalos de 10 años, de [0, 10) a «90 y más», y dibujá el histograma. ¿Qué porcentaje tiene menos de 20 años? ¿Qué forma tiene la distribución?

RESOLUCIÓN

Edad	Personas	%
[0, 10)	133	9,9
[10, 20)	198	14,8
[20, 30)	253	18,9
[30, 40)	186	13,9
[40, 50)	192	14,3
[50, 60)	178	13,3
[60, 70)	100	7,5
[70, 80)	71	5,3
[80, 90)	22	1,6
90 y más	6	0,4
Total	**1.339**	**100,0**

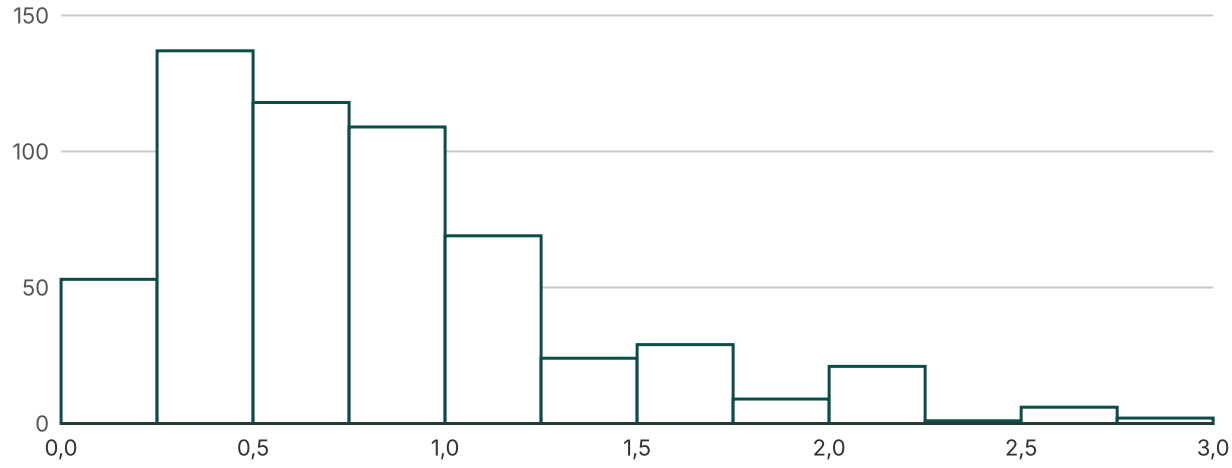


Eje vertical: cantidad de personas.

Tiene menos de 20 años el 24,7 %. La distribución es bastante pareja entre los 10 y los 60 años y cae después: es la estructura por edad de una población joven, que se angosta en las edades mayores. Antes de agrupar, los 7 casos con −1 pasan a 0 años: si no, quedan fuera de todo intervalo o caen en uno que no existe.

c) Para las personas ocupadas que declararon ingreso (ESTADO = 1 y P21 > 0), armá un histograma con intervalos de 250.000 pesos, de 0 a 3 millones. ¿Cuántas personas quedan por encima de 3 millones? ¿En qué intervalo hay más casos? Describí la forma.

RESOLUCIÓN



Eje vertical: cantidad de personas.

En millones de pesos. De las 594 personas, 16 tienen ingresos de 3 millones o más y no entran en el gráfico. El intervalo con más casos es el de 250 a 500 mil pesos, con 137 personas.

La distribución es muy asimétrica a la derecha: muchos ingresos bajos y medios, y una cola larga de ingresos altos. Por eso se corta el eje: con intervalos iguales hasta el máximo de la base, casi todas las personas quedarían en las dos primeras barras y el resto del gráfico sería vacío. Al cortar, hay que decir cuántos casos quedaron afuera.

d) La EPH es una muestra: repetí el porcentaje de personas con al menos secundario completo ponderando por PONDERA. ¿Cambia mucho?

RESOLUCIÓN

Ponderado da 50,9 %, contra 57,1 % contando filas. En R: sum(PONDERA[NIVEL_ED %in% 4:6]) / sum(PONDERA); en Excel, SUMAR.SI.CONJUNTO sobre PONDERA dividido por SUMA(PONDERA).

La diferencia es de 6,2 puntos: no es un detalle. Entre los encuestados hay, en proporción, más personas con secundario completo que en la población, y el ponderador lo corrige. El porcentaje que describe a Gran Catamarca es el ponderado; el otro describe a la muestra.

EJERCICIO 5 Gráficos que engañan

Para discutir

Cada uno de estos gráficos salió publicado. Decidí qué está mal, qué impresión deja en quien lo mira rápido y cómo lo harías bien.

a) Un noticiero muestra la intención de voto de dos candidatos en barras: 44 % y 47 %. El eje vertical empieza en 42 %, y la barra del segundo parece el triple de alta que la del primero.

RESOLUCIÓN

El eje truncado. En un gráfico de barras la altura es la cantidad: si el eje no empieza en cero, la proporción entre las barras miente. 47 no es el triple de 44: la diferencia es de 3 puntos. Las barras arrancan siempre en cero; si lo que importa es la diferencia, se la informa con un número o con un gráfico de puntos.

b) Una encuesta pregunta «¿Qué redes sociales usás?» y presenta las respuestas en un gráfico de sectores: los porcentajes de las porciones suman 163 %.

RESOLUCIÓN

Un gráfico de sectores es un todo repartido en partes que no se superponen: tiene que sumar 100 %. Con respuesta múltiple cada persona elige varias redes, y los porcentajes suman más. El gráfico correcto son barras, una por red, con el porcentaje de personas que la usa, y una aclaración de que una persona puede estar en varias.

c) Un informe muestra los ingresos con un histograma de intervalos desiguales: hasta 200 mil, de 200 a 500 mil, de 500 mil a 1 millón y de 1 a 5 millones. La altura de cada barra es la cantidad de personas, y la última barra es la más ancha y la más alta.

RESOLUCIÓN

En un histograma lo que representa la frecuencia es el área, no la altura. Si los intervalos tienen distinto ancho y la altura es la frecuencia, la barra ancha queda con un área enorme y parece que los ingresos altos son la mayoría. Hay dos arreglos: usar intervalos iguales, o dibujar la densidad, que es la frecuencia dividida por el ancho del intervalo.

d) Una línea muestra la cantidad de nacimientos en 2010, 2015, 2020, 2022, 2023 y 2024, con los seis años igualmente espaciados en el eje horizontal.

RESOLUCIÓN

El eje del tiempo no es proporcional: cinco años y un año ocupan el mismo espacio, así que la pendiente de los últimos tramos parece más suave o más brusca de lo que es. El tiempo se dibuja a escala, dejando los huecos de los años que faltan, o se aclara que la serie no es anual.

EJERCICIO 6 ¿Qué gráfico usarías?

Para discutir

Para cada situación, elegí un gráfico y justificá. Empezá por preguntarte qué tipo de variable es y qué se quiere mostrar.

a) La distribución del grupo sanguíneo de 300 donantes de un banco de sangre.

RESOLUCIÓN

Una variable cualitativa nominal con cuatro categorías: barras ordenadas de mayor a menor. Un sector también es aceptable, porque son pocas categorías y suman el 100 %.

b) La edad de los 500 pacientes internados en un hospital durante el año.

RESOLUCIÓN

Una variable cuantitativa continua con muchos datos: un histograma. Si además interesa ver la mediana y los atípicos, un diagrama de caja, que se ve en la semana que viene.

c) La cantidad de nacimientos de cada mes, de los últimos cinco años.

RESOLUCIÓN

Una serie de tiempo: un gráfico de líneas, con el tiempo en el eje horizontal a escala. Las barras también sirven, pero la línea muestra mejor la tendencia y la estacionalidad.

d) El nivel educativo de varones y de mujeres en la base de la EPH.

RESOLUCIÓN

Dos variables cualitativas, una de ellas ordinal: barras agrupadas por sexo con los porcentajes de cada nivel, o barras apiladas al 100 %, con los niveles en orden. Los porcentajes se calculan dentro de cada sexo, para que la comparación no dependa de cuántos varones y mujeres hay. Es la tabla de contingencia, que se ve en la semana 5.

---

estadistica.ar/docentes/trabajos-practicos/semana-2 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.