

Práctico 1

Bases de datos

En las siguientes actividades se trabajará principalmente con información extraída de las siguientes bases de datos.

- Base de datos de una encuesta de evaluación de CEPRAM. CEPRAM (Centro de Promoción del Adulto Mayor) es una fundación destinada a la promoción de la salud de los adultos mayores. Una de las actividades principales desarrollada por CEPRAM es el dictado de cursos en distintas áreas de conocimiento. (Para más información ver www.cepram.org.ar) Su utilización se realiza con permiso de CEPRAM.
- Base de datos de la Ecuesta Permanente de Hogares (EPH) realizada por el Instituto Nacional de Estadística y Censos (INDEC) correspondiente al primer trimestre de 2017. EPH es un programa nacional de producción sistemática y permanente de indicadores sociales, que permite conocer las características sociodemográficas y socioeconómicas de la población. Es una base de datos de acceso público. (Para más información, acceda al sitio de Indec).

Adicionalmente se consideraron situaciones ficticias a fines de ejemplificar algunos conceptos teóricos.

Actividades

1. Se desea realizar un estudio para determinar que variables están relacionadas y predicen mejor el rendimiento escolar en alumnos de una escuela primaria de la Ciudad de Córdoba. Para ello se seleccionó aleatoriamente una muestra de alumnos de dicha escuela. Se evaluaron las siguientes variables: condición socioeconómica de los padres, nivel educativo de los padres, región de procedencia, índice de IQ. Identifique la unidad de análisis y las variables en dicho estudio.
2. Se desea realizar un estudio para comparar el rendimiento académico entre escuelas públicas y privadas en la ciudad de Córdoba. Para ello se seleccionaron al azar 25 colegios privados y 25 colegios públicos. Se evaluó el promedio histórico y el puntaje obtenido en el Plan Aprender (Plan Nacional de Evaluación Educativa) en las diferentes áreas de conocimiento. Identifique unidad de análisis y la variables en dicho estudio.
3. Determine los niveles de medición de las siguientes variables. Indique, adicionalmente, que valores puede tomar cada una (p ej. la variable año de nacimiento no tiene un límite inferior definido, sin embargo el límite superior no puede ser mayor al año actual)

Base EPH:

- Sexo
 1. Varón
 2. Mujer
- Año de nacimiento
- Nivel educativo:
 1. Primaria Incompleta (incluye educación especial)
 2. Primaria Completa
 3. Secundaria Completa
 4. Superior Universitaria Incompleta
 5. Superior Universitaria Completa
- Cuanto tiempo hace que trabaja en su empleo actual
- Cuánto cobró en el último mes

Base CEPRAM:

- Cómo está usted de satisfecho con su vida:
 1. Nada
 2. Algo
 3. Bastante
 4. Mucho
 5. Totalmente
- ¿Cuántas veces ha ido al médico en el último mes?
- Indique su nivel de acuerdo. Uno de los principales motivos para acercarse al CEPRAM fue hacer amigos, insertarse en un grupo:
 1. Nada importante
 2. Poco importante
 3. Importante
 4. Muy importante
- 4. Si somos estrictos, las siguientes variables están mal definidas. Justifique por qué, considerando las nociones de **exhaustividad** y **exclusión mutua**.
 - Nivel Educativo Alcanzado
 1. Primario
 2. Universitario
 - Ocupación actual
 1. Estudia
 2. Trabaja
 3. Desocupado
- 5. De la muestra correspondiente a la EPH se seleccionaron las personas mayores a 18 años que no tuvieron ningún nivel de instrucción educativo en la region del gran Córdoba. A continuación observamos las edades para esos casos

caso	Edad	Lee
1	61	No
2	48	No
3	56	No
4	65	No
5	38	No
6	62	No
7	68	No
8	63	Si
9	63	No
10	76	No
11	63	Si
12	33	No
13	57	No
14	60	No
15	82	Si
16	62	No

- Calcule la **media**, la **varianza** y la **desviación estándar** para la variable Edad.
- Realice una **tabla de distribución de frecuencias** para la variable **Edad**. Tome como referencia la siguiente tabla

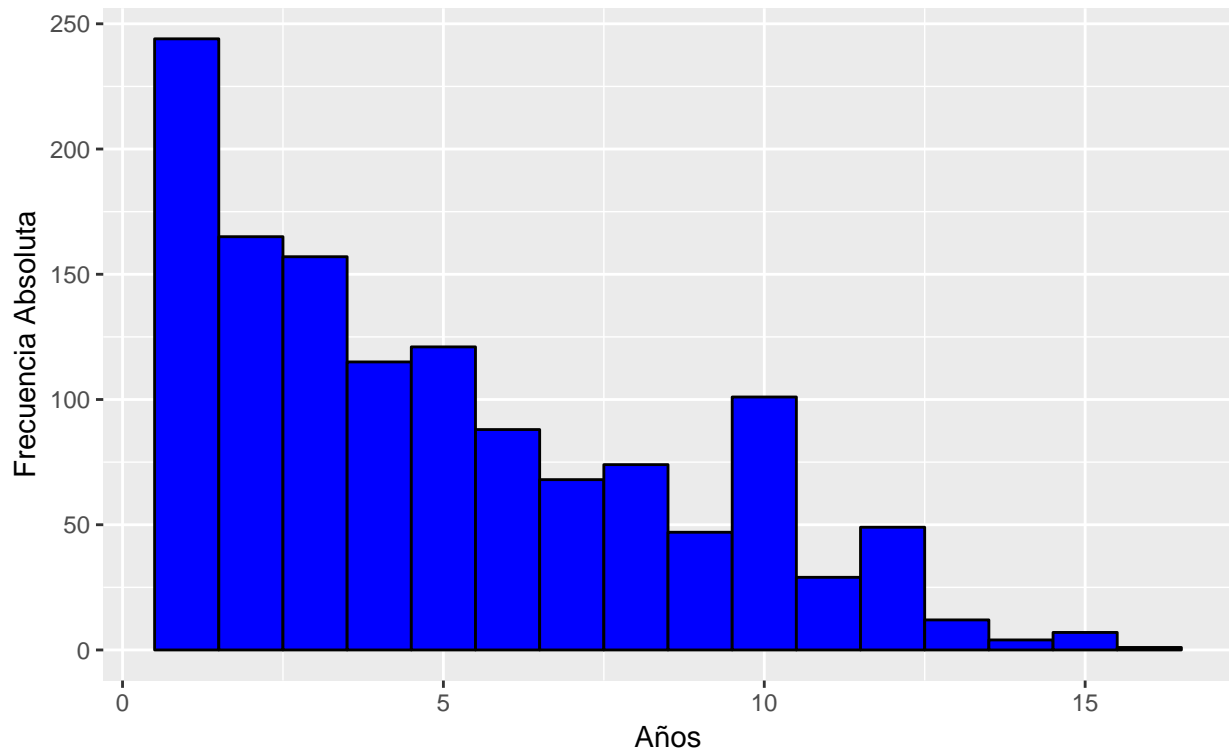
Edad	f	F	f'	F'
[30, 40]				
(40, 50]				
(50, 60]				
(60, 70]				
(70, 80]				

- Identifique los intervalos que contienen a la **mediana** y la **moda** en la tabla anterior. Realice alguna descripción de lo observado. Por ejemplo: La mediana se encuentra entre los valores ... Esto quiere decir que el 50% de los casos se encuentra aproximadamente por debajo de ...
 - Realice un **tabla de distribución de frecuencias** para la variable **Lee**. En este caso, no es necesario que calcule las frecuencias acumuladas (F y F').
6. La siguiente es una tabla de distribución de frecuencias para la variable nivel educativo (nivel máximo alcanzado) para personas mayores de 18 años de la EPH:

Nivel	f	Fa	Fr	Fra
Preescolar	60	60	0.0014712	0.0014712
Primario	10172	10232	0.2494238	0.2508950
EGB	489	10721	0.0119906	0.2628856
Secundario	16725	27446	0.4101074	0.6729930
Polimodal	498	27944	0.0122113	0.6852043
Terciario	5169	33113	0.1267471	0.8119514
Universitario	7366	40479	0.1806189	0.9925703
PosgradoUniversitario	303	40782	0.0074297	1.0000000

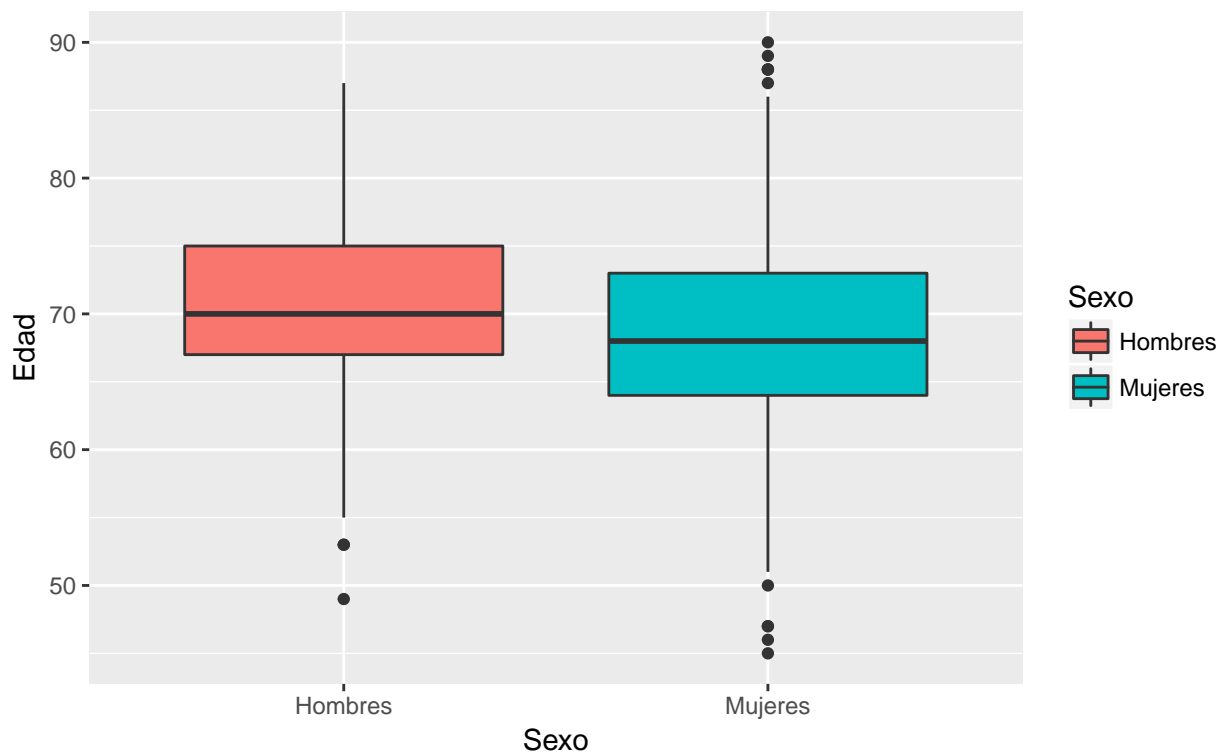
¿Qué nivel de medición tiene la variable? Identifique el **modo**, la **mediana**. ¿Qué porcentaje de la muestra tiene un nivel educativo inferior o igual al Primario? ¿Qué porcentaje tiene un nivel Terciario o Superior? ¿Cual es la suma de frecuencias de personas que completaron el Secundario y el Polimodal?

7. La siguiente imagen es un histograma para los años que llevan asistiendo a la institución los participantes de la encuesta de CEPRAM.



Identifique el **modo**. Especifique si la distribución es **simétrica**, **asimétrica a la izquierda** o **asimétrica a la derecha**.

8. El siguiente gráfico es un boxplot (también llamado gráfico de cajas o bigotes) para la Edad de los participantes del CEPRAM, particionado por sexo.



Describe el gráfico. Además, indique qué representan los extremos de las cajas, los extremos de los bigotes, la

línea horizontal al medio de la caja y los puntos extremos.