

# 7. Datos - variables

Los datos numéricos son valores de variables numéricas.  
Los datos categóricos son valores de variables categóricas.

Las **variables** son **características** que pueden tomar valores diferentes de una unidad a otra, como la edad de las personas, la cantidad de habitantes de cada ciudad, la duración o el consumo de una lamparita.



¿Datos y variables? ¿Son o no son lo mismo?

¿Entonces que son los datos?

Los **datos** son los **valores** observados de las variables.

Para ilustrar los conceptos, consideremos la siguiente tabla. Muestra una parte de la libreta donde la maestra registra datos de sus alumnos.

| Alumno                 | Lengua | Matemática | Ciencias Naturales | Participación | Certificado de Vacunas |
|------------------------|--------|------------|--------------------|---------------|------------------------|
| Cortez María           | 8,25   | 6,12       | 9,51               | Buena         | Si                     |
| García Lobos, Federico | 6,59   | 9,06       | 8,47               | Regular       | Si                     |
| Gordon, Susana         | 9,07   | 7,39       | 9,72               | Buena         | Si                     |
| Medignone, Horacio     | 7,55   | 6,42       | 8,64               | Mala          | No                     |
| Vázquez, Florencia     | 6,25   | 9,63       | 7,59               | Buena         | Si                     |

Las unidades son los alumnos del grado, identificados mediante la variable “Alumno” cuyos valores son el nombre y apellido de cada uno de ellos (primera columna de la tabla). Las cinco columnas restantes contienen el **nombre** y los valores de las demás **variables**.

Los nombres encabezan las columnas: Lengua, Matemática, Ciencias Naturales, Participación, Certificado de Vacunas, y en el cuerpo de la tabla (filas a continuación) aparecen los **valores** de cada una de ellas.

**Nombres de las variables**

**Valores observados de las variables (datos)**

|  |  |  |  |
|--|--|--|--|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Las variables tienen un **nombre** y un **valor** para cada individuo de la población.

Los **datos** son los **valores** observados -medidos- de las **variables** para los individuos de una muestra.

Los datos solos dicen muy poco, si no sabemos a qué variables corresponden.

## □ 7.1 Variables numéricas y variables categóricas

Los **datos numéricos** son valores de variables numéricas. Los **datos categóricos** son valores de variables categóricas.

En el ejemplo de la libreta de anotaciones de una maestra, las columnas 2, 3 y 4 dan el promedio de notas en cada una de las asignaturas, se trata de **variables numéricas**. La primera, muestra el nombre y apellido de cada alumno; la quinta, el grado de participación en clase registrado en 3 categorías, y la sexta, si la/el alumna/o presentó o no presentó su certificado de vacunas. Todas ellas son **variables categóricas**.

La estadística trata con números, pero **no todas las variables son numéricas**. En este ejemplo, la primera y las dos últimas son **categóricas**. Para resumir los valores de este tipo de variables utilizamos **cantidades y porcentajes**. Por ejemplo, podemos calcular la cantidad de alumnos que se llaman “Juan”, o que entregaron el certificado de vacunas, o el porcentaje de alumnas/os que tienen una participación “Buena”.

La mayoría de las variables (y por consecuencia también de los datos) se pueden clasificar en **numéricas y categóricas**. También se los denominan **cuantitativos y cualitativos** respectivamente.

Para analizar variables categóricas se utilizan **cantidades, proporciones y porcentajes**.

### Ejemplo:

En el censo de población de la República Argentina del año 2001, una de las preguntas fue: ¿Cuál es el grado de educación de las personas con 15 años y más? La tabla 7.1 responde a esa pregunta. Su título permite ver, inmediatamente, de qué se tratan los datos. Se consigna el año porque estos datos cambian con el tiempo.

Al pie figura, la fuente de los datos: el INDEC. En la primera columna de la tabla se presentan los nombres de las categorías de la variable “Nivel de Educación”; en la segunda y tercera su distribución. En la segunda columna, la distribución se expresa en cantidades, con el encabezamiento indicando “Cantidad de personas”. En la tercera columna, la distribución se expresa en porcentajes como también lo muestra su encabezamiento. Suele ser más sencillo pensar en porcentajes. Es más fácil decir el 48,9% tiene estudios primarios completos, que decir que 12.720.081 personas tienen estudios primarios completos.

**DISTRIBUCIÓN DEL NIVEL DE EDUCACIÓN**  
**DE LA POBLACIÓN DE 15 AÑOS Y MÁS. 2001** TABLA 7.1

| Nivel de Educación  | Cantidad de personas | Porcentaje |
|---------------------|----------------------|------------|
| Sin instrucción (1) | 962.460              | 3,7        |
| Primario incompleto | 3.693.766            | 14,2       |
| Primario completo   | 12.720.081           | 48,9       |
| Secundario completo | 6.373.046            | 24,5       |
| Terciario completo  | 2.263.082            | 8,7        |
| Total               | 26.012.435           | 100        |

(1) incluye nunca asistió, jardín e inicial.

**Fuente:** INDEC. Dirección Nacional de Estadísticas Sociales y de Población. Dirección de Estadísticas Sectoriales en base a procesamiento especiales del Censo Nacional de Población, Hogares y Viviendas 2001.

**Distribución de una variable:** La distribución de una variable nos dice cuáles son sus posibles valores y con qué frecuencia aparecen.

La tabla 7.1 muestra la distribución de la **variable categórica “Nivel de educación”**, máximo nivel de educación alcanzado por las personas de 15 años o más. Tiene 5 categorías: “Sin instrucción”, “Primario incompleto”, “Primario completo”, “Secundario completo” y “Terciario completo”. La columna encabezada por “Cantidad de personas” muestra **la frecuencia de cada una de las 5 categorías**, esto es, la **cantidad de personas** que pertenecen a esa categoría. Se trata de **frecuencias absolutas**. La **suma de las frecuencias** da como resultado la **cantidad total de datos**, 26.012.435, es la cantidad de personas de 15 años ó más en el año 2001.

**La frecuencia relativa** es el cociente entre la frecuencia absoluta y la cantidad total de datos. Su suma es 1. Cuando las frecuencias relativas están expresadas en porcentaje, la suma es 100, como vemos en la tercera columna de la tabla 7.1.

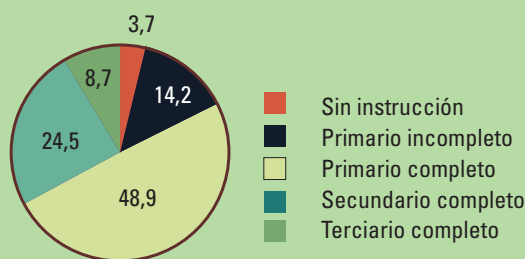
## 7.1.1 Gráficos para datos categóricos

### 7.1.1.1 Gráficos circulares

Utilizaremos un gráfico circular, también llamado gráfico de torta, para visualizar la distribución de la variable “nivel de educación” (tabla 7.1). Podremos visualizar los porcentajes de personas que pertenecen a cada una de las 5 categorías.

**Gráfico circular:** Se utiliza para representar la distribución de los valores de una variable categórica. El círculo representa el total de los datos. Cada sector dentro del círculo representa una categoría con el ángulo proporcional a su tamaño (cantidad o porcentaje que pertenece a dicha categoría).

Para realizar un gráfico circular, primero se dibuja un círculo. Los  $360^\circ$  representan el total, en este caso todas las personas de 15 años o más de la República Argentina en el 2001. Cada sector dentro del círculo representa una categoría con el ángulo proporcional a su tamaño (cantidad o porcentaje). El sector correspondiente a la categoría “Secundario completo” tendrá un ángulo de  $0,245 \times 360 = 88,2$  grados.

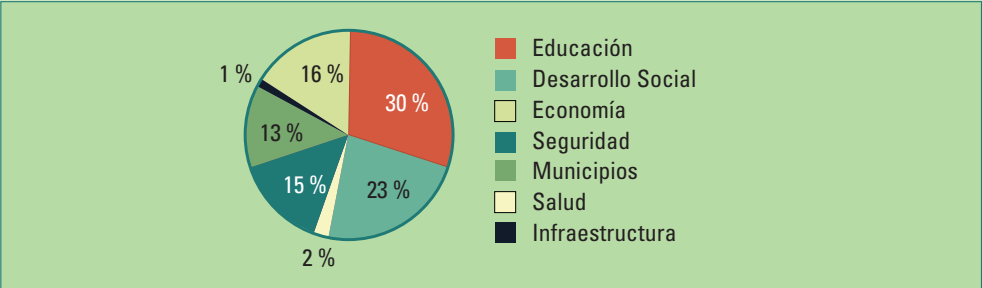


**Figura 7.1.** Gráfico circular de la distribución del nivel de educación de las personas de 15 años y más de la República Argentina. Año 2001

Los gráficos circulares permiten visualizar cómo las partes forman el total, aunque es más difícil comparar ángulos que longitudes. Estos gráficos no son buenos para comparar con precisión los tamaños de las diferentes partes, para eso los gráficos de barras son mejores.

Los gráficos circulares muestran sectores de área proporcional al porcentaje del total correspondiente a cada grupo o categoría, pero generalmente no muestran la cantidad total en cada grupo, en términos de unidades originales (pesos, número de personas, etc.). Este enfoque se traduce en una pérdida de información.

Para ilustrar esa situación consideremos los datos proporcionados por la Lotería de la Provincia de Buenos Aires en Junio de 2008 <http://www.loteria.gba.gov.ar/> sobre como reparte sus ganancias entre diferentes organismos de la provincia.



**Figura 7.2.** Gráfico circular de la distribución las ganancias de la Lotería de la Provincia de Buenos Aires junio de 2008

Vemos los porcentajes destinados a los diferentes organismos. Se destinó más del 50% entre Educación y Desarrollo Social. Pero, ¿cuánto fue realmente, en pesos? Veamos esa información en la tabla siguiente.

Siempre se puede pasar de cantidades a porcentajes. En la página de la Lotería de la provincia de Buenos Aires aparecen las cantidades totales y las destinadas a educación por mes, para el período enero-julio de 2008, pero aunque no están los porcentajes podemos calcularlos:

| Año 2008 | Educación  | Total mensual | Porcentaje |
|----------|------------|---------------|------------|
| Enero    | 37.307.382 | 143.225.097   | 26%        |
| Febrero  | 45.541.083 | 164.313.370   | 28%        |
| Marzo    | 34.872.907 | 130.834.379   | 27%        |
| Abril    | 32.646.300 | 116.425.710   | 28%        |
| Mayo     | 25.241.707 | 96.293.288    | 26%        |
| Junio    | 35.416.187 | 117.960.104   | 30%        |
| Julio    | 45.553.614 | 139.475.636   | 33%        |

No se puede ir de los porcentajes a los valores originales sin el conocimiento del total. Esta falta de información puede ser un verdadero problema, por ejemplo, cuando los gráficos muestran los resultados de una encuesta de opinión. Para evaluar el margen de error del porcentaje de personas que respondieron a la pregunta de una manera determinada es necesario saber cuántas personas respondieron la encuesta.

| DISTRIBUCIÓN DE LAS GANANCIAS DE LA LOTERÍA DE LA PROVINCIA DE BUENOS AIRES DE JUNIO DE 2008 POR ORGANISMO. TABLA 7.2 |             |
|---|-------------|
| Organismo   | Junio 2008  |
| Educación   | 35.416.187  |
| Desarrollo Social   | 27.370.667  |
| Salud   | 2.843.829   |
| Seguridad   | 17.224.945  |
| Municipios  | 15.141.832  |
| Infraestructura   | 1.413.519   |
| Economía  | 18.549.125  |
| Total   | 117.960.104 |

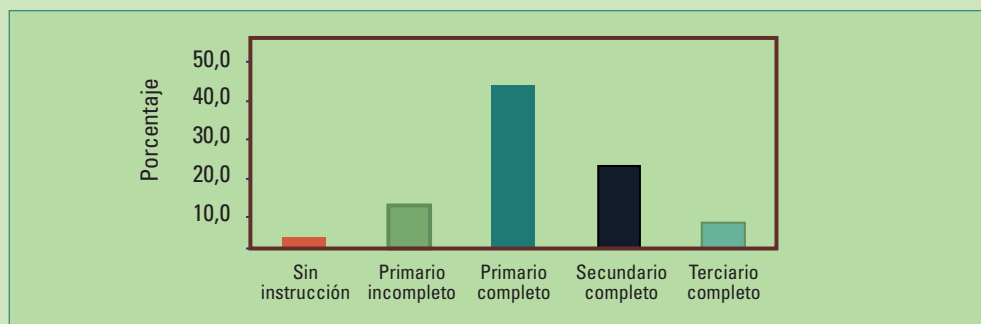
### 7.1.1.2 Gráficos de barras

Las categorías se representan en el eje horizontal y la cantidad, o el porcentaje, de datos en el eje vertical. La **altura de las barras** sobre cada categoría representa la cantidad de datos de cada una de ellas. Tal como ocurre con los gráficos

**Los gráficos de barras** se utilizan para representar la distribución de los valores de una variable categórica.

circulares, divide a los datos en grupos correspondientes a las categorías y muestra cuántos, o qué porcentaje de individuos pertenecen a cada categoría. Mientras que los gráficos circulares utilizan fundamentalmente porcentajes para indicar el tamaño de cada clase, los gráficos de barras utilizan tanto cantidades como porcentajes.

La figura 7.3 muestra un gráfico de barras de la distribución de los valores de la variable “Nivel de Educación”. La altura de cada barra representa los porcentajes de las personas de más de 15 años con nivel de educación mostrado en su base. La barra sobre la categoría “Primario Completo” es la más alta, es la categoría con la mayor cantidad de personas. Podemos comparar categorías: vemos que son más los individuos que tienen el secundario completo, que aquellos que no completaron su educación primaria.



**Figura 7.3.** Gráfico de barras de la distribución de la población de 15 años y más de la República Argentina, según máximo nivel educativo. Año 2001

El gráfico de barras tiene un interés adicional cuando las categorías tienen un orden natural como ocurre en este caso. Vemos que la categoría central “nivel primario completo” es la más poblada y que la caída es más abrupta hacia las categorías correspondientes a menores niveles de educación que hacia los mayores.

Tanto en los gráficos de barras como en los gráficos circulares, los porcentajes de las categorías tienen que sumar 100%:

$$3,7 \% + 14,2 \% + 48,9 \% + 24,5 \% + 8,7 \% = 100 \%$$

# 7.1.2 Dos variables categóricas

Retomando el tema del nivel de educación, el INDEC incluye los totales y los porcentajes por nivel de educación y género en la presentación de la información de la distribución de la población de 15 años y más de la República Argentina. La tabla nos muestra cómo se distribuyen en forma conjunta dos variables categóricas, nivel de educación y género.

Podemos calcular las cantidades de todas las casillas que nos interesen.

DISTRIBUCIÓN DE LA POBLACIÓN DE 15 AÑOS O MÁS SEGÚN NIVEL DE EDUCACIÓN DE Y GÉNERO. AÑO 2001 TABLA 7.3

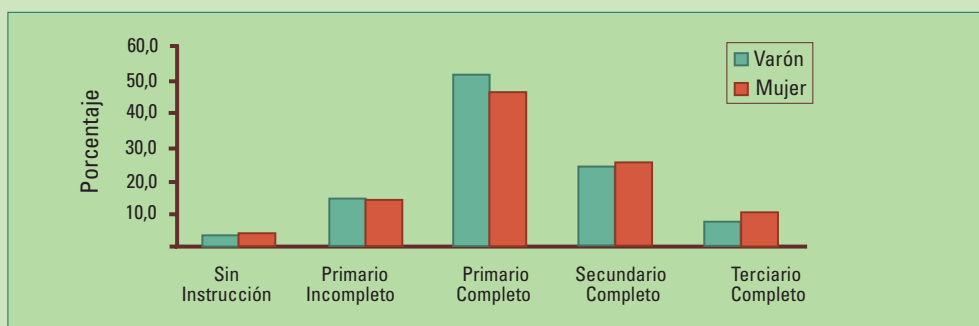
| Nivel de educación  | Total | Total      | Género     |            |
|---------------------|-------|------------|------------|------------|
|                     |       |            | Varón      | Mujer      |
|                     |       | 26.012.435 | 12.456.479 | 13.555.956 |
| Sin instrucción (1) | 3,7%  |            | 3,5%       | 3,9%       |
| Primario incompleto | 14,2% |            | 14,3%      | 14,1%      |
| Primario completo   | 48,9% |            | 51,5%      | 46,5%      |
| Secundario completo | 24,5% |            | 23,7%      | 25,2%      |
| Terciario completo  | 8,7%  |            | 7,0%       | 10,3%      |

(1) incluye nunca asistió, jardín e inicial.

**Fuente:** INDEC. Dirección Nacional de Estadísticas Sociales y de Población. Dirección de Estadísticas Sectoriales en base a procesamiento especiales del Censo Nacional de Población, Hogares y Viviendas 2001

A menudo los **gráficos de barras** se utilizan para **comparar dos grupos**, dividiendo la barra de cada categoría en dos y mostrándolas una al lado de la otra.

Un gráfico de barras conjunto nos permite comparar las distribuciones de la variable “Nivel de Educación” en varones y mujeres.



**Figura 7.4.** Porcentaje de personas con más de 15 años de acuerdo al nivel de educación y género.  
Datos tabla 7.3.

Vemos que en el nivel primario hay más varones que mujeres, pero en el secundario y terciario la relación se invierte, aunque todas las diferencias son pequeñas.



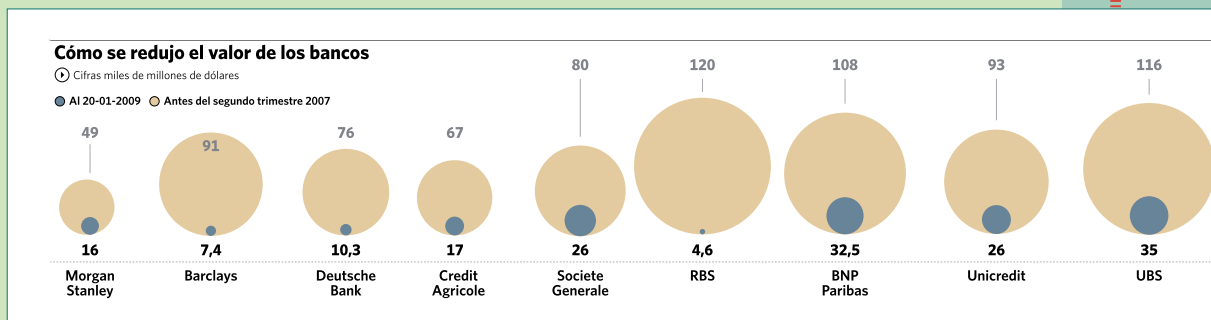
## 7.2 Actividades y ejercicios

1. Un pictograma es un gráfico de barras que se reemplaza por figuras. Las figuras representan las cantidades o los porcentajes. En forma intencional o no intencional, muchas veces los gráficos exageran las relaciones entre las categorías.
  - a. Se utilizó el siguiente pictograma para ilustrar una reducción cercana al 50% de los abandonos de mascotas en la vía pública de una ciudad después de una campaña oficial de concientización. Para reflejar esa reducción sin distorsionar la figura, el artista redujo tanto el alto como el ancho en un 50%:



Explique por qué la sensación visual de la reducción es bastante mayor que 50%. ¿Cómo debería haber sido la reducción de la figura para reflejarla en forma adecuada?

- b. Un artículo referido a las consecuencias de la crisis financiera de Estados Unidos en 2008 ilustra la reducción de los valores de los bancos mediante el siguiente pictograma. El valor del banco se calcula multiplicando la cantidad total de acciones por su cotización en la Bolsa de Nueva York.



**Fuente:** Diario Clarín, 22 de Febrero 2009

Indique si el pictograma muestra en forma correcta la reducción. Observe que los diámetros de los círculos son proporcionales a los valores.

- Los siguientes datos son parte de los resultados del primer censo general de la Provincia de Santa Fe (1887). <http://www.digitalmicrofilm.com.ar/censos/estadisticas.php>

| Localización de la vivienda |         | Nacionalidades |        |                |     | Alfabetización   |        |
|-----------------------------|---------|----------------|--------|----------------|-----|------------------|--------|
| Urbana                      | 90.764  | Argentina      | 92.170 | Inglaterra     | 753 | Sí sabe escribir | 62.608 |
| Rural                       | 116.712 | Italia         | 46.268 | Paraguay       | 673 | No sabe escribir | 87.042 |
| Fluvial                     | 2.250   | Suiza          | 5.232  | Chile          | 211 |                  |        |
| Otros                       | 382     | Francia        | 2.944  | Brasil         | 192 |                  |        |
|                             |         | España         | 2.397  | Bélgica        | 142 |                  |        |
|                             |         | Alemania       | 2.070  | Portugal       | 76  |                  |        |
|                             |         | Austria        | 1.131  | Estados Unidos | 74  |                  |        |
|                             |         | Uruguay        | 903    |                |     |                  |        |

Obtenga un diagrama de barras y un gráfico circular para distribución de los habitantes de la provincia de Santa Fe en 1887 de acuerdo con cada una de las siguientes tres variables categóricas: 1) Alfabetización, 2) Nacionalidades y 3) Localización de la vivienda.

- Utilice el gráfico que considere adecuado para representar los datos de la tabla siguiente.

**PRODUCTO BRUTO NOMINAL EN DÓLARES PER  
CÁPITA PARA 10 PAÍSES DE AMÉRICA DEL SUR,  
DURANTE 2008 SEGÚN EL FMI**

|           |        |           |        |
|-----------|--------|-----------|--------|
| Argentina | 8.522  | Ecuador   | 3.927  |
| Bolivia   | 1.889  | Paraguay  | 2.658  |
| Brasil    | 8.676  | Perú      | 4.610  |
| Chile     | 10.814 | Uruguay   | 8.860  |
| Colombia  | 5.174  | Venezuela | 11.828 |

**Fuente:** [http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(nominal\)\\_per\\_capita](http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita)