



Presentación de la información estadística

M. Laura Dodino | Magíster. Sala de Matemática de los Institutos Normales de Montevideo. PROYECTO "CON MATEMÁTICA SÍ PUEDO".

En homenaje a la Prof.^a Marita Gamio

«Estudiar es desocultar, es alcanzar la comprensión más exacta del objeto, es percibir sus relaciones con los otros objetos. Implica que el estudioso, sujeto del estudio, se arriesgue, se aventure, sin lo cual no crea ni recrea.»

Paulo Freire (1994)

El objetivo de este trabajo es continuar con el estudio de la Estadística Descriptiva en la formación de maestros con referencia a las formas básicas de presentación de la información.

Trataremos las características de cada representación, sus ventajas y desventajas, destacando especialmente la relación entre el tipo de variable y la representación gráfica que se usa.

► Queremos destacar inicialmente *un rasgo esencial del análisis estadístico: Busca describir propiedades de los conjuntos de datos y no de cada dato en particular de forma aislada, implicando una mirada global a la distribución.*

Carmen Batanero (2000) destaca: *«La idea de representante de un conjunto de datos es importante en las aplicaciones prácticas, por ejemplo, al comparar dos conjuntos de datos respecto a una misma variable de interés. Como indican Mokros y Russell (1995) hasta que los niños no conciben el conjunto de datos como*

un todo, y no como un agregado de valores, no podrán comprender las ideas de resumen de los datos o representante de los datos, que se refiere al conjunto global y no a ninguno de sus valores aislados.»

Presentación de la información relevada

La forma en que se presenta la información estadística es importante básicamente por dos razones: el análisis de los datos es más fácil de hacer y nos permite poner la información a disposición de otras personas.

Podemos considerar cuatro formas básicas de presentación de la información:

1. Tabular
2. Gráfica
3. Numérica
4. Textual.

1. Presentación tabular

Corresponde a una ordenación sistemática de datos en filas y columnas, de acuerdo a criterios de clasificación que interesen, utilizando números, de forma tal que:

- los datos puedan ser interpretados rápidamente
- se puedan extraer conclusiones y hacer comparaciones.

Para ello se requiere completar la tabla o el cuadro con títulos, notas que aclaren el significado de los datos y su origen, faciliten la interpretación de la información.

Un cuadro está compuesto de un título, encabezado, columna matriz, y un cuerpo o contenido. Se pueden agregar además notas preliminares y al pie, la fuente de información y el número de tabla para facilitar su referencia.

Cuadro N° 1		
DISTRIBUCIÓN PORCENTUAL DE VIVIENDAS PARTICULARES POR CONDICIÓN DE OCUPACIÓN, SEGÚN DEPARTAMENTO*		
Departamento	Ocupadas	Desocupadas
Total País	81,7	18,3
Montevideo	90,5	9,5
Artigas	86,6	13,4
Canelones	78,7	21,3
Cerro Largo	81,1	18,9
Colonia	79,5	20,5
Durazno	81,7	18,3
Flores	83,4	16,6
Florida	79,9	20,1
Lavalleja	78,3	21,7
Maldonado	53,0	47,0
Paysandú	82,7	17,3
Río Negro	82,3	17,7
Rivera	85,2	14,8
Rocha	56,5	43,5
Salto	86,5	13,5
San José	85,5	14,5
Soriano	85,1	14,9
Tacuarembó	80,9	19,1
Treinta y Tres	80,1	19,9

* Corresponden a resultados preliminares de Censos 2011.

Fuente: http://www.ine.gub.uy/censos2011/adelantos_breves.html#viviendas (consultada 22/3/12)¹

- ▶ El título debe dar idea del tipo de información que contiene el cuadro, debe ser claro y llamativo, y si se incluye en el informe más de un cuadro es recomendable incluir un número que lo identifique.

En este caso, “Distribución porcentual de viviendas particulares por condición de ocupación, según departamento”.

Por ejemplo, a partir de la información de la fila destacada, podemos leer que en Soriano, el 85,1% de las viviendas particulares están ocupadas y el 14,9% están desocupadas.

Un título bien confeccionado debe decir:

1. a qué refieren los datos
2. dónde se recogieron los datos, a qué lugar corresponden
3. cuándo ocurrió el fenómeno al que se refieren los datos.

- ▶ Encabezado/s, aparece/n en la primera fila, que sigue al título en este caso, y se indican aquí otras clasificaciones de los datos (otra variable en este ejemplo, la *condición de ocupación de la vivienda* que toma dos valores: ocupadas y desocupadas).

¹ El Instituto Nacional de Estadística aclara en dicha página: «Las cantidades que se presentan pueden verse modificadas en las próximas semanas una vez que: (a) se analicen los datos recogidos que presentaron problemas de consistencia o que quedaron incompletos en el momento de la entrevista; (b) se incorporen a la base de datos las planillas especiales de establecimientos de reclusión, hogares de guarda, hogares de ancianos, hospitales psiquiátricos, personas en situación de calle y en refugios, y aquellas utilizadas para agilizar el trabajo de campo durante las dos últimas semanas de 2011».

► La columna matriz es la primera columna del cuadro (a la izquierda) y en ella aparece la clasificación principal de los datos; en este caso, la distribución de las viviendas particulares *por departamento*.

► Las notas al pie generalmente se colocan antes de la fuente y tienen como objetivo hacer alguna aclaración, observación o advertencia. Se ubica una señal, * en este caso, puede ser también un número, una letra, y a continuación el comentario. En este ejemplo se indica que los datos corresponden a resultados preliminares de Censos 2011, dando cuenta de que pueden cambiar con el tiempo a medida que avance el procesamiento de los datos.

Cuando los datos que se presentan en el cuadro no fueron obtenidos directamente por quien confecciona el cuadro, se debe indicar el origen de los mismos. Esto es un principio ético asociado al trabajo estadístico, que permite a los usuarios de esta información evaluar la calidad de los datos de base, hacer verificaciones y acceder eventualmente a los datos de base directamente. La fuente debe contener los requisitos básicos de una cita: autor, nombre de la publicación, editorial, fecha, número de cuadro o página donde se encuentran los datos. En nuestro ejemplo (Fuente: http://www.ine.gub.uy/censos2011/adelantos_breves.html#viviendas [consultada 22/3/12]) también se indica la fecha de consulta considerando la nota al pie de página.

► El cuerpo o contenido del cuadro lo constituyen los números que aparecen en cada casillero determinado por la columna matriz y el encabezamiento. Dichos números dan cuenta de los datos estadísticos realmente observados, y que luego son ordenados y distribuidos de acuerdo a las características predeterminadas.

En este caso aparecen las frecuencias relativas expresadas en porcentajes, tal como lo dice el título (distribución porcentual); así podemos leer a partir de la primera fila del contenido del cuadro que en nuestro país el 81,7% de las viviendas particulares están ocupadas.

Recordemos que la información cuantitativa básica corresponde a frecuencias absolutas (provenientes directamente del recuento de datos) y a la frecuencia relativa. Tiene interés a veces agregar las frecuencias acumuladas.²

Desde el punto de vista de su diseño, se recomienda que el cuadro no sea ni muy ancho y corto, ni muy largo y angosto.

Un buen cuadro nos permite responder las preguntas: qué son los datos presentados; dónde se recogieron; cuándo, cómo y bajo qué criterios de clasificación.

¿Qué es un censo?

Según el INE:

«Es una investigación estadística en la que los datos e información se obtienen de la totalidad de las unidades de información que componen el “universo o población” por investigar.

Las claves para definir una población son: (1) que los sujetos o unidades de información tengan características comunes (son viviendas u hogares o personas o empresas, etc.) y (2) que comprende absolutamente a todas esas unidades.»

2. Presentación gráfica

**«Una imagen vale más que mil palabras»
(pero no más que mil números)**

Los gráficos son una de las formas más usadas de presentación de información estadística y ayudan a su análisis. Las ideas presentadas son en general entendidas con mayor rapidez que las explicaciones numéricas, verbales o escritas. Generalmente, los lectores no se saltan la observación de un gráfico atractivo y bien construido. Los gráficos presentan mejor la idea en

² Se llama frecuencia absoluta (f.a) de un valor de la variable a la cantidad de veces que aparece, es producto del recuento de los datos. Se llama frecuencia relativa (f.r) de cada valor de la variable al cociente entre su f.a y la cantidad total de datos recogidos, en general se muestra en porcentajes.

general, a expensas de los detalles, pues no aparecen o no se destacan los números exactos tal como se presentan en un cuadro.

Es necesario considerar que:

- ▶ no presentan tanta información como una tabla, ya que una gran variedad de información menoscaba la utilidad de un gráfico y
- ▶ presentan además cifras aproximadas, mientras que los cuadros permiten la inclusión de los datos con toda la exactitud que se desee.

Este tipo de presentación es complementario de la tabular, y por lo tanto no debe sustituir las tablas que contienen la información básica.

Un gráfico es un instrumento que presenta datos numéricos por medio de cantidades de magnitud, por ejemplo, longitudes, áreas, volúmenes.

Como las comparaciones son fundamentalmente visuales, debe procurarse que los gráficos estén contruidos de tal forma que no exageren o disminuyan las tendencias de los datos, ya que podría producirse que las conclusiones o deducciones del observador resulten distorsionadas.

Los gráficos tienen en común los siguientes componentes estructurales (*ápu*d Friel, Curcio y Bright, 2001:126):

- ▶ Un *marco* que proporciona información acerca de las medidas usadas y los datos medidos, y que está formado por ejes, escalas, marcas...
- ▶ Los *especificadores* que suelen ser líneas, barras u otras marcas, que indican las relaciones entre los datos representados.
- ▶ Las *etiquetas* que indican el tipo de medida usada, los datos a los que se aplica esa medida o el título del gráfico.
- ▶ El *fondo*, que incluye los colores, la cuadrícula e imágenes, sobre el que el gráfico puede ser superpuesto.

Considerando estos componentes presentamos algunos principios para construir un gráfico adecuado y atractivo que, a su vez, propicie una interpretación global correcta de la información estadística:

- ▶ Debe tener proporciones adecuadas, ni muy ancho ni muy alto, pues pueden influir en la percepción del lector. Para el rectángulo que contiene al gráfico, una proporción adecuada entre su base y su altura puede ser de 1,5 a 1,

por ejemplo, 15 cm de base y 10 cm de altura. Es conveniente también centrar el gráfico en la hoja o en el espacio donde se presente.

- ▶ Debe explicarse por sí mismo. Debe tener título; leyenda; escalas; referencias de colores, por ejemplo; fuente.
- ▶ Debe incluir pocas series de datos, de lo contrario puede volverse confuso.
- ▶ Las escalas utilizadas no deben desfigurar los hechos o relaciones que se quieren mostrar, debiendo indicarse, por ejemplo, la base 0 (cero) para que sirva de punto de referencia en el eje vertical y debe señalarse con claridad si se ha cortado la escala.
- ▶ Debe ser sencillo, fácil de interpretar y adecuado al tipo de datos que se quieren presentar y al público al cual va dirigido.

Tipos de gráficos y su relación con la variable que se muestra

Es fundamental que el tipo de gráfico sea coherente, por sus características, con el tipo de información o el objetivo que se persigue al representarla, de otra manera puede suceder que la representación gráfica se convierta en un instrumento ineficaz, que produce más confusión que otra cosa, innecesaria o productora de malas interpretaciones.

Haremos referencia a los gráficos: diagramas circulares – diagramas de barras – histogramas – pictogramas.

¿Según el tipo de variable a representar (cualitativa: nominal, ordinal; cuantitativa: discreta, continua) hay por lo menos un gráfico que se adecua mejor?

Recordemos que llamaremos:

- ▶ VALOR de una variable a cada uno de los posibles estados con los que puede presentarse dicha variable.
- ▶ DATO es el valor que toma la variable en cada individuo de nuestra población o muestra.

Valor y dato no son sinónimos. Por ejemplo, una vivienda particular puede estar ocupada o desocupada, estos son los dos valores en los que puede presentarse la variable “estado de ocupación de una vivienda particular”, pero cuando el censista realiza su trabajo de campo constata que en el departamento de Soriano, en su capital Mercedes, la vivienda de la calle 18 de julio 478 se encuentra ocupada (dato). A un elemento concreto de nuestra población se le asigna un

estado específico de la variable (ocupada). Cada dato corresponde a un valor de la variable, pero no necesariamente todo valor de la variable es efectivamente un dato (esto ocurre especialmente para las variables cuantitativas).

En Censos 2011 se relevó el estado de ocupación de una vivienda particular determinada por su dirección/ciudad/departamento y esto es lo que se envía para conformar la matriz de datos, que luego será procesada.

Es necesario tener presente **qué tipo de comparaciones se pueden hacer con los datos, según el tipo de variable** a la que referencian, pues esto determina también el tipo de gráfico que mejor está en sintonía con dicha naturaleza.

Dados dos datos, si estos son:

- ▶ *nominales*, solo podemos decir si estos dos datos son iguales o diferentes (es decir, si ambos pertenecen a la misma clase o a dos clases diferentes);
- ▶ *ordinales*, podemos establecer si estos dos datos son iguales o diferentes (lo mismo que podemos hacer con los datos nominales) y además podemos establecer un orden entre los mismos (es decir que existe un cierto orden entre las clases);
- ▶ *numéricos*, correspondientes a variables discretas o continuas, podemos decidir si son iguales o diferentes, en qué orden están y además podemos realizar cálculos con ellos, esto último implicará la obtención de ciertos parámetros estadísticos como el promedio, la mediana, etc.

A partir de estas consideraciones se establece la pertinencia del uso de cada tipo gráfico a utilizar según el tipo de variable que se quiere representar.

Cuadro Nº 2

Tipo de variable		Ejemplo, indicando valores que puede tomar la variable	Tipo de gráfico a utilizar
CUALITATIVA	Nominal	Sexo: femenino, masculino	Circular/gráfico de barras/pictogramas
	Ordinal	Estado de mantenimiento una vivienda: malo, regular, bueno, muy bueno, excelente	Gráfico de barras
CUANTITATIVA	Discreta	Cantidad de hijos: 0,1,2,...15	Gráfico de barras/pictogramas
	Continua	Estatura:	Histogramas

Para una variable nominal, es decir, para aquella que identifica un atributo o característica no medible de un individuo (no es posible asignar un número-medida) y, por lo tanto, establece categorías o clases mutuamente excluyentes y exhaustivas entre las que no existe un orden, para relevar un dato se establecerá entonces a cuál de las categorías pertenece el individuo. Cada respuesta del individuo (dato) debe poder ser incluida en una de las clases definidas y solamente en una.

Un gráfico circular representa la frecuencia de cada clase usando una cantidad de superficie (en este caso, de una figura que es un sector circular; el círculo queda dividido en tantos sectores como valores diferentes tenga la variable, y cada uno de los cuales tendrá una amplitud proporcional a las frecuencias de los valores), permitiendo así la comparación visual de las frecuencias a partir de esas superficies.

Estos gráficos no muestran valores en orden lineal.

En general se utilizan, además, si la variable estudiada tiene **pocos valores distintos**. Si los sectores fueran muchos, la comparación visual se dificulta y puede ser más conveniente usar diagrama de barras.

Tengamos presente que para construir manualmente un diagrama circular, debemos obtener el **ángulo al centro que determina cada sector** que, como ya dijimos, es **proporcional a la frecuencia absoluta (o relativa)** correspondiente.

Ejemplo: el 100% de los datos corresponde con 360° (medida del ángulo al centro completo); así, un valor de la variable que tenga una frecuencia relativa de 25% debe representarse con un ángulo al centro de 90°.

Debido a que la lectura del gráfico se basa en hacer comparaciones proporcionales se recomienda, luego de haber calculado la medida de la amplitud del ángulo al centro correspondiente a cada valor de la variable, que se representen los ángulos al centro siguiendo su orden de amplitud, comenzando con el mayor, ubicando consecutivamente los otros ángulos en sentido horario.

Un gráfico circular permite apreciar dicha proporción y cuál de las categorías es la mayor o cómo se relacionan entre

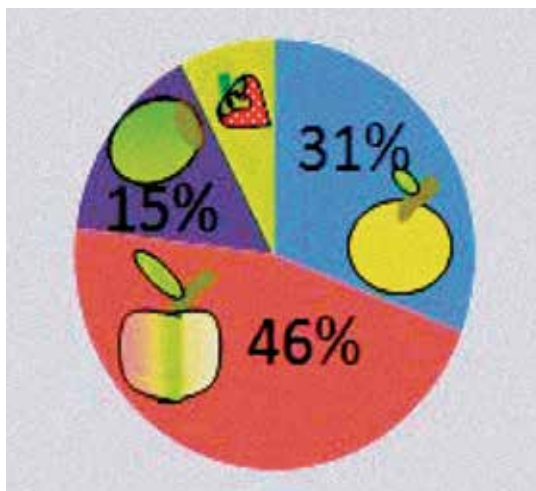
ellas las diversas categorías (tamaño relativo de cada una de las partes).

En cambio, los diagramas que utilizan ejes cartesianos pueden destacar además de las categorías implicadas, el orden lineal de los valores de la variable según algún criterio. Por ejemplo, en una variable ordinal como el nivel de instrucción de las madres de alumnos de nivel 3 de Educación Inicial ubicar, como primera categoría, primaria incompleta, y siguiente primaria completa, ciclo básico incompleto, ciclo básico completo, etc., da cuenta de la cantidad de años de educación formal recibidos.

Diagramas circulares

Veamos algunas propuestas para alumnos escolares en las que aparecen representaciones gráficas circulares.

A) El gráfico circular representa las cantidades en % de algunas variedades de fruta vendida por el señor Silva durante el mes de abril de 2011 (manzanas, naranjas, kiwis y frutillas).



- ¿Cuál fue la variedad de fruta más vendida?
- ¿Cuál fue la variedad de fruta menos vendida? (fundamenta tu respuesta)
- Estima el porcentaje correspondiente a las frutillas.
- Sumando el porcentaje de kiwis al porcentaje de ..., se obtiene una suma igual al porcentaje de ...
- Sabiendo que el gráfico se refiere a un peso total de 130 kg de fruta, estima el número de kilos de cada variedad de fruta vendida.

Problema adaptado a partir de http://www.ajudaalunos.com/Quiz_mat/percent_html/cloze_grafico.htm (consultada 23/3/12)

B) ¿Cuáles son las lenguas más habladas del mundo?



Fuente: <http://www.saberespractico.com/ordenador/como-hacer-una-grafica-circular-en-excel/> (consultada 23/3/12)

Este gráfico se construyó usando Microsoft Office Excel, a partir de una tabla que indicaba los porcentajes correspondientes a cuántas personas del mundo hablan cada una de las 5 lenguas más habladas. Se construyó una tabla en el libro 1, y luego se abrió la pestaña INSERTAR para seleccionar dentro la opción CIRCULAR. Se eligió a continuación el gráfico circular 3D.

- ▶ Confecciona una tabla que pueda dar lugar al gráfico presentado.
- ▶ Fundamenta.

Nota complementaria para consideración del docente:

Tabla original

IDIOMA	FRECUENCIA RELATIVA (%)
Chino	23,6
Inglés	11,3
Hindi	8,2
Español	6,9
Ruso	4,7

C) La rueda de los alimentos



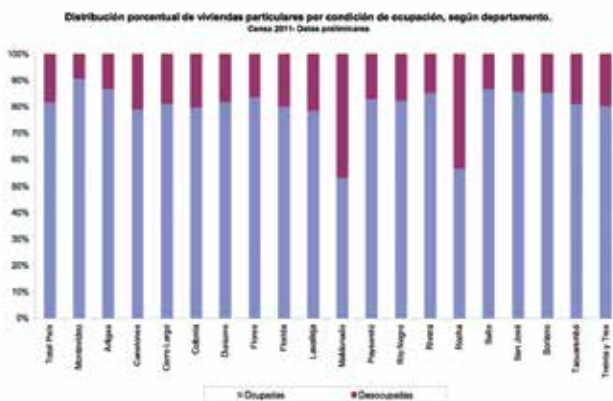
Fuente: http://mat7anoest.no.sapo.pt/ex1_pag21.html

Desearíamos construir un gráfico similar a este en nuestro cuaderno, que cumpliera con todos las sugerencias de este artículo. Explica cómo lo realizarías.

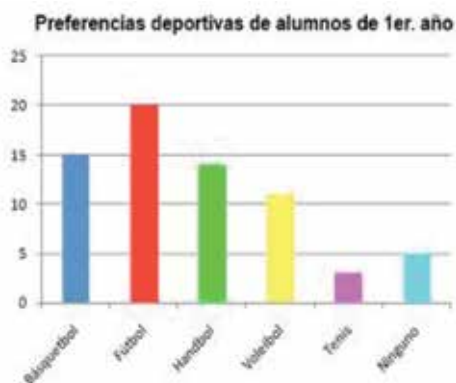
Diagramas de barras

Veamos diferentes tipos de diagramas de barras.

A continuación aparece un ejemplo de **diagrama de barra compuesto**, hay dos series de datos que se representan, uno encima de otro para cada departamento, que indican una distribución porcentual. Se consideran frecuencias relativas, usando como unidad de referencia (100%) la cantidad de viviendas de cada departamento. Cada barra, por lo tanto, tiene una altura de 100%, que no en todos los casos corresponde a la misma frecuencia absoluta. Con esta información, por ejemplo, no podemos asegurar que las viviendas desocupadas en Maldonado sean efectivamente más que las desocupadas en Montevideo. Para afirmar eso deberíamos contar con más información: la cantidad de viviendas encuestadas en cada uno de los departamentos. Aquí podemos leer que en Maldonado la relación de las viviendas ocupadas y las desocupadas es aproximadamente de 1 a 1, o que la mitad de las viviendas de Maldonado están desocupadas, que en Flores las viviendas ocupadas son aproximadamente 4 veces más que las desocupadas, o que por cada dos viviendas desocupadas, hay 8 ocupadas.



Veamos otro ejemplo:



En el **eje horizontal** o de **abscisas** se representan los **datos o modalidades**; en el **eje vertical** o de **ordenadas** se representan las **frecuencias de cada dato o modalidad**. Las frecuencias que en general se usan son las *absolutas*, *relativas* y *relativas acumuladas*.

Conviene tomar en cuenta ciertas **consideraciones** al momento de construir una representación gráfica, con el objetivo de que la información sea comunicada de la manera más eficaz posible y sin distorsiones. Por ejemplo, para aquellos gráficos que usan ejes de coordenadas: el **eje** en el que se representan las **frecuencias** de los datos (el vertical o de las ordenadas usualmente) debe **comenzar en cero (0)**, pues de otra manera podrían darse impresiones erróneas al comparar la longitud de las columnas, la posición de los puntos o las líneas que representan dichas frecuencias.

Fuente: <http://bibliotecadeinvestigaciones.wordpress.com/matematicas/estadistica-descriptiva-conceptos-generales/>

En los diagramas de barras, la **longitud de los espacios** que representan **un valor de una variable de tipo cuantitativa discreta** debe ser igual, así como la distancia entre las barras.

Las barras no se tocan entre sí lo que representa el hecho de que los valores de la variable son discretos para el caso de las variables cuantitativas o clases disjuntas para las variables cualitativas ordinales.



Ejemplo de **diagrama de barras simples**.

Este gráfico muestra las barras con longitudes proporcionales a las cantidades de temperatura que representan, pero es necesario observar que se ha cortado la escala en el eje vertical sin indicarlo explícitamente; la distancia entre el supuesto 0 del eje y el 70, es casi igual a la distancia entre el 70 y el 85. Su eje horizontal no tiene una escala numérica, pero sí pone en evidencia el orden de los días de la semana.

Fuente: <http://www.mathematicsdictionary.com/spanish/vmd/full/v/verticalbargraph.htm>

Actividad para el lector:

El siguiente gráfico muestra las compras de resmas de papel para la impresora de la adscripción del Instituto, durante los primeros meses del año 2007.



Fuente: <http://javalangnullpointer.wordpress.com/2007/02/22/graficas-con-java/>

La secretaria informa que en el segundo trimestre del año se compró la misma cantidad de resmas que en el primer trimestre y fundamenta su respuesta basándose en información que obtiene del gráfico. Detallar una posible fundamentación para su afirmación.

- ▶ Si se está planificando el presupuesto para el segundo semestre del año, y se asume un aumento global del 60% con respecto al primer semestre, calcular el monto anual total destinado a este rubro, suponiendo que cada resma vale \$ 240.
- ▶ Indicar si la siguiente proposición es verdadera o falsa. “Se gastó en el primer cuatrimestre la tercera parte del monto total destinado a este rubro”. Argumentar.

Este **diagrama de barras múltiples** permite comparar dos o más distribuciones estadísticas, pues las mismas se representan en un mismo sistema de ejes.

Las barras de la misma serie se representan con el mismo color y luego para cada categoría se pegan las barras.




Histogramas

Este tipo de gráfico se usa para variables cuantitativas continuas o para aquellas cuantitativas discretas que toman muchos valores distintos y su rango de variación es muy grande; por lo tanto se tratan como continuas. Para este segundo caso se hace necesario agrupar los datos recogidos en intervalos o clases.

Este tipo de gráfico se realiza sobre un sistema de coordenadas cartesianas de modo esencialmente igual a los diagramas de barras. En el eje abscisas se construyen unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia absoluta de cada intervalo si estos son todos de la misma longitud. El área de cada rectángulo es proporcional a la frecuencia de los valores representados. Los rectángulos así construidos estarán “pegados” pues los intervalos son contiguos, lo que muestra la continuidad de los valores de la variable.

Por ejemplo, si queremos analizar los números que han salido a la quiniela, en la página aparece el siguiente cuadro:

 LISTADO COMPLETO DE LAS VECES QUE HA SALIDO CADA NUMERO (DEL 000 AL 99) DESDE EL 17/12/1997 a 2011

000	001	002	003	004	005	006	007	008	009
132	131	116	96	114	121	127	132	122	136
010	011	012	013	014	015	016	017	018	019
132	128	140	131	139	131	120	142	138	126
020	021	022	023	024	025	026	027	028	029
143	118	120	103	149	113	116	130	136	124
030	031	032	033	034	035	036	037	038	039
113	140	100	115	125	132	126	111	140	130
040	041	042	043	044	045	046	047	048	049
150	150	131	111	124	114	139	111	102	141
050	051	052	053	054	055	056	057	058	059
123	125	112	123	132	132	133	120	104	116
060	061	062	063	064	065	066	067	068	069
117	119	120	153	115	137	104	117	138	135
070	071	072	073	074	075	076	077	078	079
120	132	127	126	148	121	125	134	112	127
080	081	082	083	084	085	086	087	088	089
115	128	123	120	132	108	157	120	116	132
090	091	092	093	094	095	096	097	098	099
130	121	99	127	110	134	133	127	113	113

Fuente: http://www.labanca.com.uy/jkdocs/hm/hm_jue_est.html

Este listado tiene dos características que no hacen sencilla la manipulación de la información así como se presenta: el rango de variación de la variable es muy grande –del 0 al 99– y la cantidad de datos recabados es mucha, pues abarca un plazo de tiempo que va desde el año 1997 hasta 2011.

Esto implica que si deseamos tener una imagen más simple de la distribución de datos, sería conveniente agruparlos por intervalos. Para la construcción de histogramas se recomienda usar entre 4 y 20 intervalos. Con muy pocos se pierde mucha información, y con muchos se dificulta la lectura y el análisis de los datos. Tenemos que decidir entonces cuántos intervalos construiremos: una posible opción sería, considerando la tabla de la que partimos (y teniendo en cuenta nuestro SND), trabajar por decenas, construyendo así 10 intervalos o clases de igual amplitud.

La nueva tabla con los datos agrupados es:

Números que salieron en la quiniela durante el período 1997-2011, por intervalo		
Números	Intervalo o clase	Cantidad de veces que salió
0 al 9	[0-10]	1227
10 al 19	[10-20]	1327
20 al 29	[20-30]	1252
30 al 39	[30-40]	1232
40 al 49	[40-50]	1273
50 al 59	[50-60]	1220
60 al 69	[60-70]	1255
70 al 79	[70-80]	1272
80 al 89	[80-90]	1251
90 al 99	[90-99]	1207

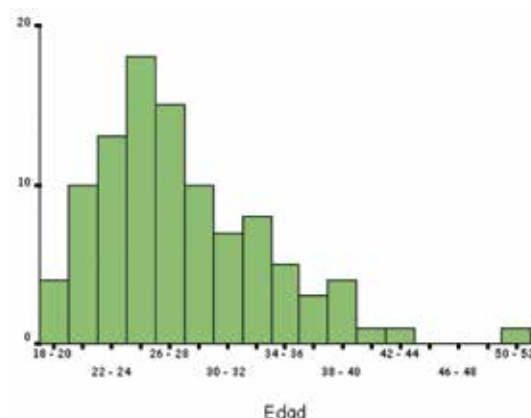
Observación: La frecuencia del primer intervalo corresponde a la suma de las frecuencias correspondientes a los primeros 10 números naturales de la tabla presentada originalmente.

Debemos recordar que al trabajar con intervalos se pierde información, cuantos menos intervalos más pérdida. Ahora, considerando solo la información de la tabla de datos agrupados no podremos asegurar si salió alguna vez el 78 o cuántas veces salió.

Para construirlo a mano, necesitamos comenzar representando en el eje de las abscisas el rango de variación de los valores de la variable (el valor menor y el valor mayor de la variable) y subdividir luego este segmento para representar los intervalos de clase definidos. Esencialmente, un histograma se construye como un diagrama de barras. Una diferencia fundamental es que en los diagramas de barras, las frecuencias de los valores de la variable están representadas por la altura de la barra, todas las cuales tienen el mismo ancho y son equidistantes entre sí. En los histogramas, la frecuencia de cada intervalo está representada por el área de la barra (estos intervalos pueden tener diferentes longitudes), y las barras quedan pegadas dando cuenta de la continuidad de los valores de la variable.

Si se consideran intervalos de clase iguales entre sí, el área de la barra solamente depende de su altura, por lo que el histograma sería similar a un diagrama de barra. Sin embargo, si los intervalos de clase son diferentes, el área de la barra depende tanto de su altura como de su ancho, y en este caso se requieren cálculos especiales para hallar la altura de cada barra de manera que el área de dicha barra sea proporcional a la frecuencia de dicho intervalo.

Ejemplo:



En el gráfico se muestra la variable edad y la frecuencia absoluta de cada intervalo. Las edades oscilan entre 18 y 52 años, por lo que el rango de variación de las edades es 34 años, cada intervalo es de 2 años. Hay tres intervalos que tienen frecuencia cero [44-46], [46-48] y el siguiente. Existen 4 personas en el intervalo [18-20] y 10 personas del grupo cuyas edades oscilan entre 20 y 22 años. En el eje de las abscisas se han señalado las marcas de cada clase.

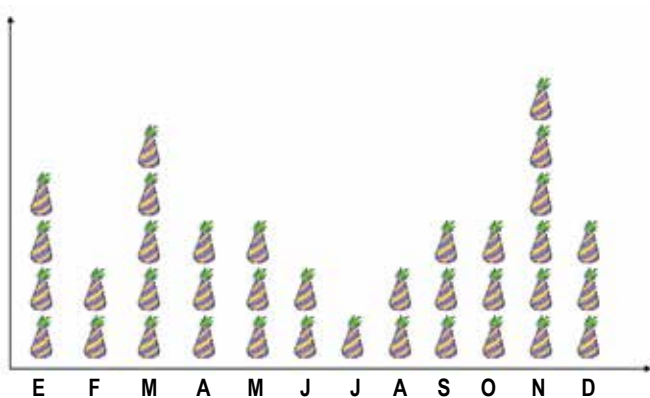
El recuento de datos estableció que 1 persona tenía 18 años y 3 tenían 19 años, este detalle se ha perdido al establecer los intervalos, pues aquí leemos que hay 4 personas cuyas edades están incluidas en [18-20]. Sin ver la tabla y solo “leyendo” el gráfico no podemos asegurar que en este grupo haya alguien que tenga 19 años.

Fuente: <http://www.fisterra.com/mbe/investiga/graficos/graficos.asp#Figura3> (consultada 10/5/12)


Pictogramas

Son **gráficos con dibujos** alusivos a la variable que se está estudiando y cuyo **tamaño es proporcional** a las **frecuencias** que representan. La mayor frecuencia se identifica por la mayor acumulación de símbolos, o también se puede variar el tamaño del dibujo de referencia. Los pictogramas se emplean, sobre todo, para hacer **más amigables y entendibles los informes estadísticos. Es similar al diagrama de barras.**

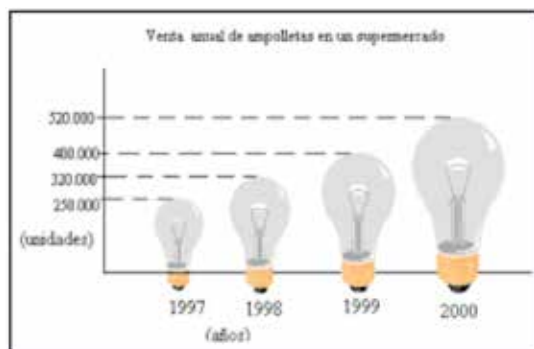
Por ejemplo, para niños de nivel inicial, si se quiere representar la cantidad de niños que cumplen años según el mes:



Fuente: <http://www.icarito.cl/enciclopedia/articulo/primer-ciclo-basico/matematica/datos-y-azar/2009/12/56-8551-9-3-datos.shtml> (consultada 23/3/12)

El símbolo  se usa para representar un niño que cumple años, podemos leer que en diciembre cumplen tres niños de la clase.

En el siguiente caso se ha agrandado la figura de referencia, de tal forma que su altura represente la frecuencia absoluta.



Al realizar un gráfico estadístico se han identificado los errores más comunes.

- ▶ Elección incorrecta del tipo de gráfico de acuerdo al tipo de variable considerada (por ejemplo, usar polígonos de frecuencias para representar variables cualitativas).
- ▶ La elección inadecuada de las escalas de representación o su omisión en alguno de los ejes.
- ▶ No especificar el origen de los ejes de coordenadas.
- ▶ Subdivisión inadecuada de los ejes.
- ▶ No respetar las características de cada tipo de gráfico, por ejemplo, obtener un diagrama de sectores en los que estos no son proporcionales a las frecuencias de las categorías.

*La elección de la **representación gráfica** se realiza teniendo en cuenta el tipo de información que se quiere comunicar, los objetivos que se persiguen al presentar la información y el público al cual va dirigida.*

3. Presentación numérica

Definiremos algunas medidas numéricas que se emplean habitualmente para describir un conjunto de datos. Asumiremos que existen dos medidas básicas para la descripción de cualquier conjunto de datos: la identificación de su centro y su variabilidad.

Llamaremos "*tendencia central*" (Canavos, 1988:12) de un conjunto de datos a la disposición de estos para agruparse, ya sea alrededor de un centro o de ciertos valores numéricos y consideraremos como su "*variabilidad*" a la dispersión de datos en el conjunto.

Consideraremos las tres medidas básicas de tendencia central: el *promedio* o media aritmética, la *mediana* y la *moda*.

Promedio es el número que se obtiene sumando todos los valores de la variable estadística (x_i) multiplicados por su correspondiente frecuencia (f_i) y dividiendo esta suma por el número de datos de la distribución (n).

Promedio = $(x_1+x_2+x_3+\dots+x_n)/n$, siendo x_i cada dato obtenido de los n considerados

Teniendo en cuenta que todos los datos son empleados para su cálculo, el valor del promedio puede afectarse desproporcionadamente debido a la existencia de valores extremos en la distribución.

Ejemplo:

Datos: 6-10-12-20, su promedio es $(6+10+12+20)/4 = 12$, si cambio solamente un valor por otro mucho mayor, el promedio se afecta mucho.

Datos nuevos: 6-10-12-100, su promedio es $(6+10+12+100)/4 = 32$

Batanero (2000) cita la investigación de Strauss y Bichler (1988) donde se plantea que un porcentaje importante de niños de entre 8 y 12 años eran capaces de comprender y aplicar adecuadamente las siguientes propiedades del promedio o media:

- ▶ es un valor comprendido entre los extremos de la distribución;
- ▶ es influenciado por los valores de cada uno de los datos;
- ▶ no tiene por qué ser igual a uno de los valores de los datos.

Sin embargo otras propiedades del promedio les resultaban muy abstractas:

- ▶ la suma de las desviaciones de cada valor al promedio es igual a cero;
- ▶ el valor obtenido del promedio de números enteros puede ser una fracción, que no tenga sentido en el contexto de los datos;
- ▶ hay que tener en cuenta los valores nulos en el cálculo del promedio;
- ▶ el promedio es representativo de los valores promediados.

Batanero destaca que *«la idea de promedio no se puede comprender hasta tanto se visualice el conjunto de datos como un todo y también que la forma de presentación de los datos (tabla, gráfico, datos sin tabular) incide en la dificultad de las tareas»*.

Mediana de un conjunto de datos es un número para el cual, si se ordenan todos los datos de la distribución de manera creciente, la mitad de estos es menor que ese número y la otra mitad es mayor. Si la cantidad de datos de la distribución es impar, la mediana es el dato que se encuentra en la mitad del conjunto ordenado. Si el número es par, se considera la mediana como el promedio de los dos datos que se encuentran a la mitad del conjunto ordenado.

Puesto que la mediana es un número que se basa en el orden del conjunto de datos, la existencia de valores extremos no afectará su valor; por lo tanto, si una distribución tiene muchos datos y contiene algunos valores extremos, la mediana puede ser la medida de tendencia central más pertinente para describir dicho conjunto. Se recomienda su uso en distribuciones asimétricas.

Continuemos ahora considerando las series, visto el promedio:

La serie 6-10-12-20 tiene promedio 12, y la serie 6-10-12-100 tiene promedio 32, pero en ambos casos la mediana es 11. Como ambas series tienen una cantidad par, la mediana es, en ambos casos, el promedio de los dos datos centrales de la distribución *ordenada*, es decir, el promedio entre 10 y 12.

Para este último caso sería un mejor representante la mediana que el promedio; en cambio para la serie 6-10-12-20 corresponderían tanto el promedio como la mediana, véase también que son valores muy similares.

Moda de un conjunto de datos es el valor de la variable que tiene mayor frecuencia.

Existen casos en que la moda no aporta como medida descriptiva, por ejemplo, en una distribución con pocos datos puede que no se repita ninguno de ellos o que la frecuencia máxima se encuentre compartida por dos o más valores de la variable. Sin embargo es la única medida de tendencia central que se puede identificar para variables cualitativas, pues solo tomamos en cuenta la frecuencia de aparición de los datos y no los datos en sí mismos. Las restantes medidas de tendencia central realizan cálculos numéricos con los datos.

Otros errores en el cálculo de *promedio*, *mediana* y *moda* fueron descritos por Carvalho (1998), citado por Batanero (2000), a partir del análisis de las producciones escritas de los alumnos al resolver tareas estadísticas:

- ▶ **Moda:** tomar la mayor frecuencia absoluta.
- ▶ **Mediana:** considerar la distribución de datos sin ordenar identificando como mediana el dato central de la distribución, identificar la mediana en la serie de frecuencias absolutas ordenada en forma creciente, tomando el valor central de dichas frecuencias absolutas, error de cálculo al obtener el promedio.
- ▶ **Promedio:** hallar el promedio de los valores de las frecuencias, no tener en cuenta la frecuencia absoluta de cada valor en el cálculo del promedio.

Variabilidad: medidas de dispersión dan una noción del grado de apartamiento de los datos respecto al valor central de la distribución. Los valores que toman estas medidas son positivos o cero, indicando el cero la máxima concentración de los datos respecto de la medida de tendencia central considerada que, en general, es el promedio.

Una medida de variabilidad está relacionada al intervalo en el que se encuentran los datos de una distribución.

Llamaremos *recorrido* o *rango* R de los datos de una distribución a la diferencia entre el dato mayor y el menor de la misma.

En general debe evitarse su uso cuando la distribución tiene valores relativamente extremos o cuando el conjunto de datos es muy grande.

Nota: Las medidas más útiles de dispersión son la *varianza* (el promedio del cuadrado de las distancias entre cada dato y el promedio de todos los datos) y la *desviación típica* o *estándar* (es el cuadrado de la *varianza*), que en este trabajo no se desarrollarán.

Cuadro N° 3

Resumen

Tipo de variable	Ejemplo indicando valores que puede tomar la variable	Tipo de gráfico	Medida de tendencia central/dispersión
Nominal	Sexo: femenino, masculino	Circular/gráfico de barras/pictogramas	Moda
Ordinal	Estado de mantenimiento una vivienda: malo, regular, bueno, muy bueno, excelente	Gráfico de barras	Moda
Discreta	Cantidad de hijos: 0,1,2,...,15	Gráfico de barras/pictogramas	Moda, mediana, promedio, rango, varianza, desviación típica
Continua	Estatura: 1,12 m, 1,3 m, 1,29 m	Histogramas	Moda, mediana, promedio, rango, varianza, desviación típica

4. Presentación textual

En relación al gráfico (Cuadro N° 1), el Instituto Nacional de Estadística destaca que «*Montevideo es el departamento que más viviendas particulares ocupadas tiene con el 90,5%. Por su parte Maldonado posee el mayor porcentaje nacional de viviendas desocupadas (47,0%), debido a la presencia significativa de viviendas de uso temporal, característica de las zonas balnearias*».

Fuente: http://www.ine.gub.uy/censos2011/adelantos_breves.html#viviendas (consultada 22/3/12)

Este tipo de presentación aparece habitualmente en diarios y revistas.

Como ventajas podemos destacar que:

- ▶ generalmente se usan cuando los datos son pocos;
- ▶ se pueden resaltar ciertas cifras individuales;
- ▶ se pueden explicar mejor ciertos aspectos;
- ▶ resultan muy apropiados cuando se quiere presentar una cantidad limitada de información en forma simple y atractiva.

Debemos considerar que no se puede incluir mucha información cuantitativa, porque puede generar confusión en los lectores.

Es necesario leer todo el texto para entender qué es lo que se quiere explicar.

Veamos otro ejemplo usando información extraída de Censos 2011.

DISTRIBUCIÓN DE LA POBLACIÓN POR GRANDES GRUPOS DE EDAD, SEGÚN DEPARTAMENTO (%)

Departamento	0-14	15-34	35-64	65 y más
Total país	21,9	29,3	34,9	13,9
Montevideo	19,1	30,3	35,3	15,3
Artigas	26,0	28,6	34,1	11,4
Canelones	23,0	28,9	35,5	12,6
Cerro Largo	24,7	27,4	34,9	13,0
Colonia	21,4	27,2	35,9	15,5
Durazno	24,5	28,3	33,6	13,5
Flores	22,5	26,6	35,9	15,1
Florida	22,6	27,8	34,5	15,2
Lavalleja	21,5	26,0	36,0	16,4
Maldonado	23,1	30,1	35,5	11,4
Paysandú	25,0	29,1	32,8	13,1
Río Negro	26,1	29,0	33,1	11,8
Rivera	25,4	29,0	33,8	11,9
Rocha	22,5	26,8	35,6	15,1
Salto	26,0	30,5	31,8	11,7
San José	22,9	29,2	34,8	13,1
Soriano	23,9	28,3	34,0	13,7
Tacuarembó	24,5	27,9	34,8	12,7
Treinta y Tres	23,3	28,0	35,0	13,6


El cuadro anterior se acompaña de los siguientes textos:

«Al observar la distribución de la población por grandes grupos de edades, se aprecia que el 21,9% se encuentra entre 0 y 14 años, el 29,3% entre 15 y 34 años, el 34,9% entre 35 y 64 años y el 13,9% tienen 65 o más años.»

En este primer párrafo se destaca la distribución global de la población uruguaya por intervalos de edades, los que como se puede observar no son iguales, sino que parecen estar diseñados para dar cuenta de las diferentes etapas evolutivas de una persona.

El siguiente párrafo destaca los intervalos extremos de edades, los más jóvenes y los de mayor edad de la población, identificando los 4 departamentos que cuentan en proporción a su población con más concentración en dichos intervalos extremos.

«El departamento con mayor porcentaje de población de 65 o más años es Lavalleja (16,4%). Al 27/12, lo siguen Colonia (15,5%), Montevideo (15,3%) y Florida (15,2%). Por otra parte, los departamentos con mayor porcentaje de población menor de 15 años son Río Negro (26,1%), Artigas y Salto (26%) y Rivera (25,4%).»

Tengamos presente que las diferentes formas de presentación de la información se complementan entre sí para ayudar a una mejor comprensión. Realizar dicha presentación siguiendo principios éticos, de respeto por las fuentes, por la naturaleza y vinculación de los datos, por los lectores, aporta a la toma consciente y fundamentada de las decisiones personales y profesionales. 

Bibliografía consultada

- ARTEAGA, Pedro; BATANERO, Carmen; CAÑADAS, Gustavo; CONTRERAS, J. Miguel (2011): "Las Tablas y Gráficos Estadísticos como Objetos Culturales" en *Números. Revista de Didáctica de las Matemáticas*, Vol. 76 (marzo), pp. 55-67.
- BATANERO, Carmen (2000): "Significado y comprensión de las medidas de posición central" en *Uno, Revista de didáctica de las matemáticas*, 25, pp. 41-58. En línea: <http://www.ugr.es/~batanero/ARTICULOS/isboa.pdf>
- BATANERO, Carmen (2001): *Didáctica de la Estadística*. Departamento de Didáctica de la Matemática. Universidad de Granada. En línea: <http://www.uruguayeduca.edu.uy/Userfiles/P0001%5CFile%5C118didacticaestadistica.pdf>
- CANAVOS, George C. (1988): *Probabilidad y Estadística. Aplicaciones y Métodos*. México: Mc Graw Hill.
- FREIRE, Paulo (1994): "Primera Carta. Enseñar - aprender. Lectura del mundo - lectura de la palabra" en *Cartas a quien pretende enseñar*. México: Siglo XXI editores.
- FRIEL, Susan N.; CURCIO, Frances R.; BRIGHT, George W. (2001): "Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications" en *Journal for Research in Mathematics Education*, Vol. 32, N° 2, pp. 124-158. En línea: <http://filebox.vt.edu/users/sboyce/Curriculum%20Summer%202003/Stat%20Articles/JRME2001-03-124a.pdf>
- GÓMEZ BARRANTES, Miguel (2011): *Elementos de Estadística Descriptiva*. Costa Rica: Editorial Universidad Estatal a Distancia.

N. de la A.: Agradezco a las compañeras de la Sala de Matemática, C. Damisa, A. López e I. Piedra Cueva, su lectura atenta y los comentarios fundamentales para mejorar este texto.