

EVIDENCIAS EN PEDIATRÍA

Toma de decisiones clínicas basadas en las mejores pruebas científicas

www.evidenciasenpediatria.es

Fundamentos de medicina basada en la evidencia

Representación gráfica de variables

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

Palabras clave en español: estadística; descriptiva; gráficos.

Palabras clave en inglés: statistics; descriptive; graphics.

Fecha de recepción: 11 de marzo de 2019 • **Fecha de aceptación:** 15 de marzo de 2019

Fecha de publicación del artículo: 20 de marzo de 2019

Evid Pediatr. 2019;15:13.

CÓMO CITAR ESTE ARTÍCULO

Ortega Páez E, Ochoa Sangrador C, Molina Arias M. Representación gráfica de variables. Evid Pediatr. 2019;15:13.

Para recibir Evidencias en Pediatría en su correo electrónico debe darse de alta en nuestro boletín de novedades en <http://www.evidenciasenpediatria.es>

Este artículo está disponible en: <http://www.evidenciasenpediatria.es/EnlaceArticulo?ref=2018;15:13>.

©2005-19 • ISSN: 1885-7388

Representación gráfica de variables

Ortega Páez E¹, Ochoa Sangrador C², Molina Arias M³

¹UGC de Maracena. Distrito Granada-Metropolitano. Granada. España.

²Servicio de Pediatría. Hospital Virgen de la Concha. Zamora. España.

³Servicio de Gastroenterología. Hospital Universitario La Paz. Madrid. España.

Correspondencia: Eduardo Ortega Páez, edortegap@gmail.com

En este capítulo, tercero y último de la serie de estadística descriptiva, hablaremos de la representación gráfica de las variables. En Medicina y Epidemiología cada vez es más frecuente el uso de métodos gráficos para exponer los datos. Son fáciles de comprender y en un solo vistazo dan al investigador una primera información para comprender la distribución de los datos: una imagen vale más que mil palabras. Pero el abuso de las gráficas y el mal uso puede tergiversar la percepción del lector, haciéndole cometer errores de interpretación. Este capítulo tiene la intención de dar pautas para realizar de forma correcta la representación gráfica de las variables para evitar errores. Para ello, hemos utilizado el *software* R (<https://www.r-project.org/>), por dos razones. La primera es que es de libre disponibilidad en varias plataformas Windows, Mac y Linux, y en segundo lugar por su relativa sencillez de manejo y gran versatilidad. Los lectores de este trabajo pueden, si quieren, reproducir de aquí en adelante los ejemplos tratados descargando la aplicación desde The Comprehensive R Archive Network (<https://cran.r-project.org/>) o la versión para Windows en español (<http://knuth.uca.es/R/R-UCA>), así como la base de datos de este capítulo (http://evidenciasenpediatria.es/files/Fundamentos_13_Base_datos.zip).

VARIABLES CUANTITATIVAS

Los gráficos adecuados son aquellos que sean capaces de plasmar sus características. Cuantitativas continuas, si muestran un número infinito de valores, y cuantitativas discretas si muestran un número finito.

Cuantitativas continuas

Los gráficos más usuales son el histograma, el diagrama de tallo y hoja y el diagrama de cajas.

Histograma

Es el más usado, porque es sencillo de interpretar. Se construye representando en el eje de abscisas (X) los valores agrupados de la variable en intervalos, se fijan sus límites y se construyen tantos rectángulos como intervalos haya, cuya área debe ser proporcional a la frecuencia. En el eje de ordenadas (Y) los valores de las frecuencias de la variable en términos

absolutos o relativos. La amplitud de los intervalos no tiene que ser constante, pero la mayoría de las veces lo es: en este caso, las alturas de los rectángulos serán proporcionales a su frecuencia.

Para ilustrar este capítulo se ha construido una base de datos ficticia que incluye distintos tipos de variables (figura 1). Hemos utilizado para realizar los gráficos, como se ha indicado antes, el *software* R, con la librería Rcommander (http://knuth.uca.es/R/doku.php?id=instalacion_de_r_y_rcmdr:r-uca), que convierte al programa en una interfaz mediante menús fácilmente manejable (figura 2).

En nuestro caso hemos utilizado la versión para Mac (OX) (<https://www.rcommander.com/>). En el anexo I se describen los pasos utilizados en R para cada gráfico.

Veamos el primer ejemplo. En la figura 3 podemos ver el histograma para la variable "Talla". Podemos ver que se ha dividido en ocho intervalos de clase de 10 cm, el que tiene mayor frecuencia es el comprendido entre 120-130 cm, con seis sujetos.

- **Ventajas:** es fácil de hacer e interpretar, ya que con una simple ojeada nos podemos hacer una idea de la distribución de los datos. Se puede utilizar como gráfico para una publicación. Se pueden comparar dos o más variables. Se pueden estratificar por una variable cualitativa, como en la figura 4, donde podemos observar la distribución de la talla según el sexo (realizada con el plugin EZR de Rcommander).
- **Inconvenientes:** no informa de los datos individuales, solo agrupados y es fácilmente manipulable, por lo que puede dar lugar a interpretaciones erróneas o interesadas. Esto hace que no sea el gráfico ideal.

Diagrama de tallo y hoja (stem and leaf)

En la figura 5 podemos ver la representación de la variable "Talla". En la primera columna se representan los tallos (que corresponden en nuestro caso al primer dígito) y en la segunda las hojas (en este caso el segundo dígito). Es un diagrama híbrido entre una tabla (información ordenada) y una gráfica (parecida al histograma). Tiene la ventaja de no perder información individual, identifica la distribución de los datos

FIGURA 1. BASE DE DATOS FICTICIA, CON VARIABLES

Base de datos								
	Edad (años)	Peso	Talla	IMC	Sexo	Asma	Ingresos	Tabaquismo
1	5	20,30	110,00	16,78	Masculino	Leve	1	No
2	2	10,20	81,00	15,55	Femenino	Grave	4	Si
3	3	19,00	112,00	15,15	Femenino	Leve	0	No
4	2	12,50	82,50	18,37	Masculino	Leve	0	No
5	12	32,00	145,00	15,22	Femenino	Moderado	2	Si
6	13	65,00	155,00	27,06	Femenino	Leve	1	No
7	5	20,00	105,00	18,14	Masculino	Moderado	3	No
8	7	16,70	111,50	13,43	Femenino	Grave	6	Si
9	12	35,00	150,00	15,56	Femenino	Moderado	2	Si
10	14	50,00	160,00	19,53	Masculino	Leve	0	No
11	8	16,00	116,50	11,79	Femenino	Grave	8	Si
12	4	20,00	105,00	18,14	Femenino	Leve	1	No
13	10	45,00	135,00	24,69	Masculino	Leve	0	Si
14	6	30,00	104,00	27,74	Femenino	Moderado	3	Si
15	8	25,00	125,00	16,00	Masculino	Leve	1	No
16	7	30,00	130,00	17,75	Femenino	Grave	5	Si
17	7	25,00	126,00	15,75	Femenino	Leve	1	No
18	4	20,00	110,00	16,53	Masculino	Leve	0	No
19	3	16,00	97,00	17,00	Femenino	Moderado	2	Si
20	5	14,00	97,00	14,88	Femenino	Grave	9	Si
21	6	23,00	121,00	15,71	Masculino	Leve	1	No
22	9	43,00	130,00	25,44	Femenino	Leve	0	No
23	9	35,00	130,00	20,71	Masculino	Grave	6	Si
24	11	30,00	140,00	15,31	Femenino	Moderado	2	No
25	11	32,00	138,00	16,80	Masculino	Leve	1	Si
Total	25	25	25	25	25	25	25	25
	Variables cuantitativas continuas							
	Variables cualitativas nominal y categórica							
	Variable cualitativa ordinal							
	Variable cuantitativa discreta							

(posible media y mediana) y si existen clases faltantes. Esto hace que para muchos autores sea la representación gráfica de elección.

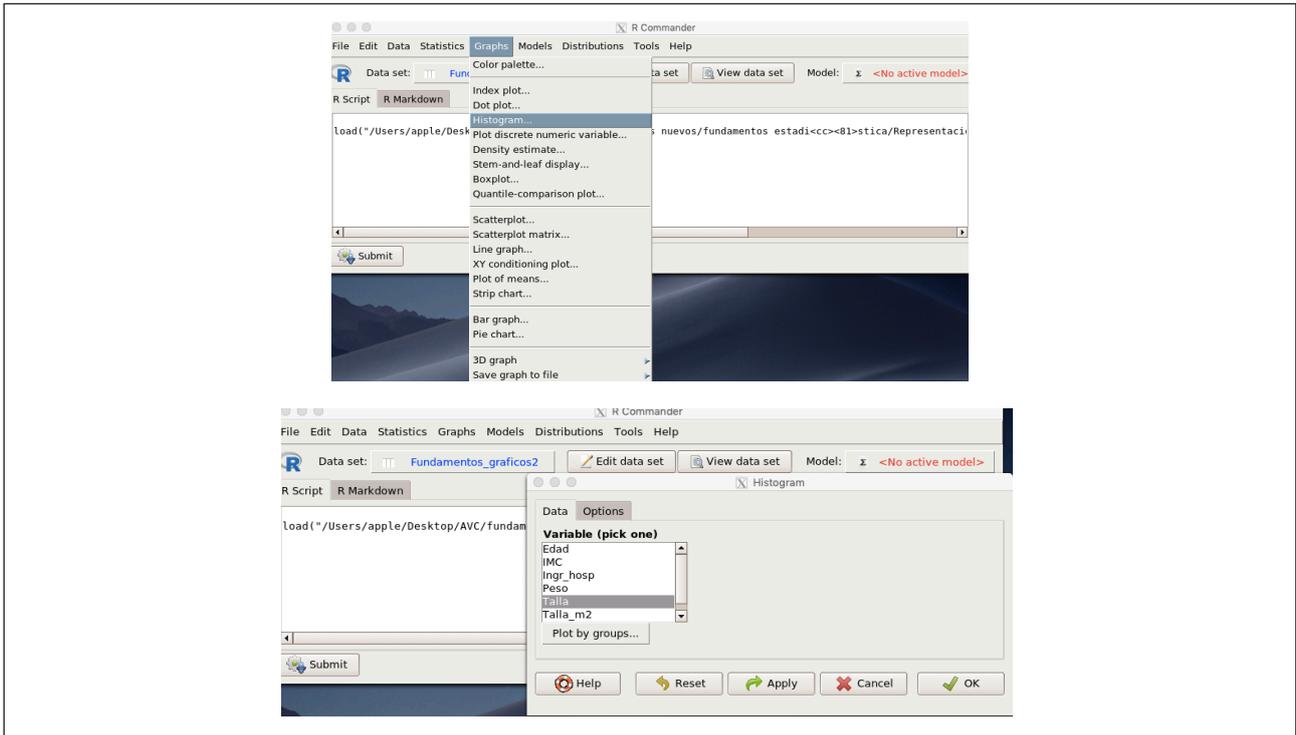
Diagrama de cajas (boxplot)

Es un tipo de gráfico que se usa frecuentemente para representar los estadísticos descriptivos. Consiste en representar la distribución de los datos mediante una caja cuyo límite superior es el percentil 75 o tercer cuartil (Q3), el inferior es el percentil 25 o primer cuartil (Q1) y el centro es la mediana

(percentil 50). Desde los extremos de la caja se prolongan unos “bigotes”, que son los límites superior e inferior de la distribución (valores adyacentes), cuyo valor no puede ser más de 1,5 veces el rango intercuartílico (RIC, distancia entre Q1-Q3). Los valores más allá de los adyacentes se denominan valores alejados (*outliers*) y muestran los valores máximos o mínimos reales de la distribución. En la figura 6 representamos el *boxplot* de la variable “Peso”.

Esta representación gráfica tiene varias ventajas: nos informa de la asimetría de la distribución, de los valores máximo y mínimo, detecta valores alejados, puede comparar dos o varias

FIGURA 2. CAPTURAS DE RCOMMANDER



distribuciones, como se muestra en la figura 7, donde podemos comprobar que la distribución de los datos en el peso del sexo masculino, aunque de mediana muy parecida al femenino es más asimétrico (caja más ancha e irregular).

que se corresponde con siete pacientes. Alternativamente también se puede representar por **grafico de barras** (figura 9), como vemos posteriormente, donde podemos observar los mismos resultados.

Cuantitativas discretas

R dispone de un tipo de gráficos especial (**discrete plot**) para variables discretas, de líneas verticales, de tal forma que en abscisas representamos el valor de la variable y en ordenadas la frecuencia absoluta. En nuestro ejemplo, vemos el resultado de la variable “Número de ingresos” de la figura 8, donde observamos que el mayor número de ingresos son cuatro y

VARIABLES CUALITATIVAS

Existen dos tipos de representaciones gráficas bien diferenciadas, según sean variables cualitativas categóricas nominales u ordinales. La primera se representa por el **gráfico de sectores** (*pie chart*) y la segunda por el **gráfico de barras** (*bar chart*).

FIGURA 3. HISTOGRAMA DE LA VARIABLE “TALLA”

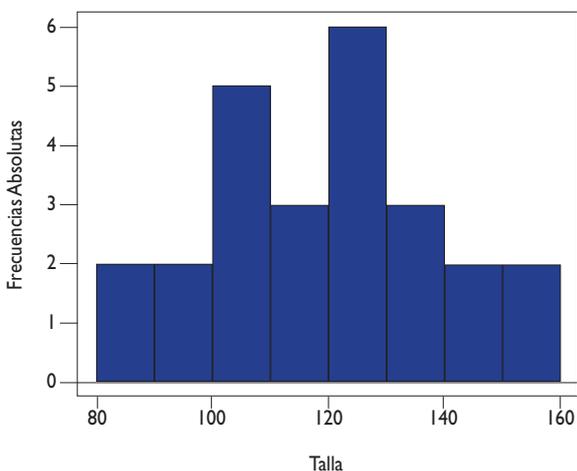


FIGURA 4. DISTRIBUCIÓN DE LA TALLA SEGÚN EL SEXO

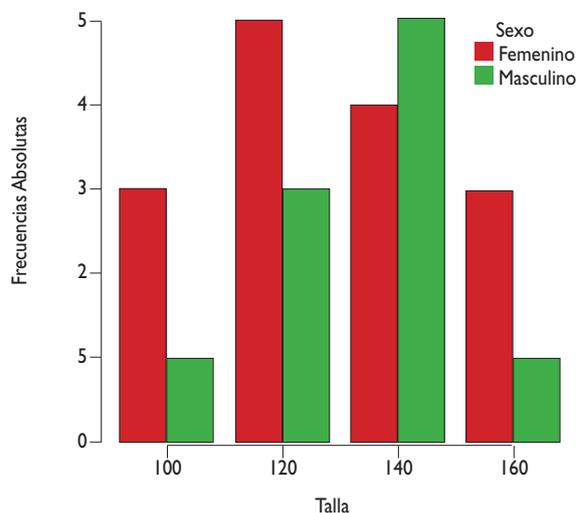


FIGURA 5. GRÁFICA DE TALLO Y HOJA DE LA VARIABLE “TALLA”

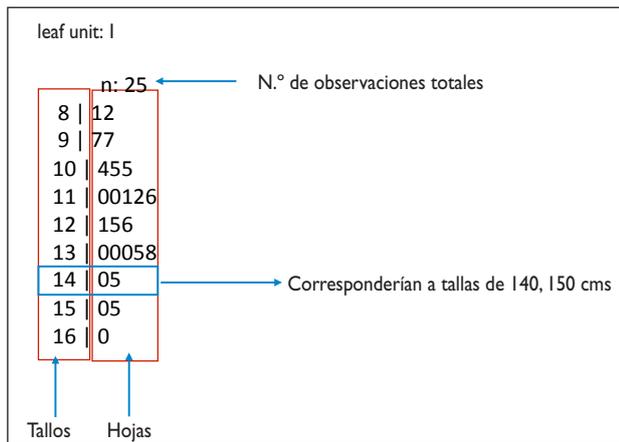


FIGURA 6. DISTRIBUCIÓN DE LA VARIABLE “PESO”

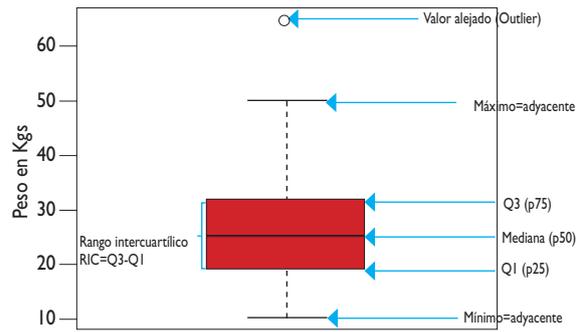


FIGURA 7. DISTRIBUCIÓN DE PESO POR SEXO

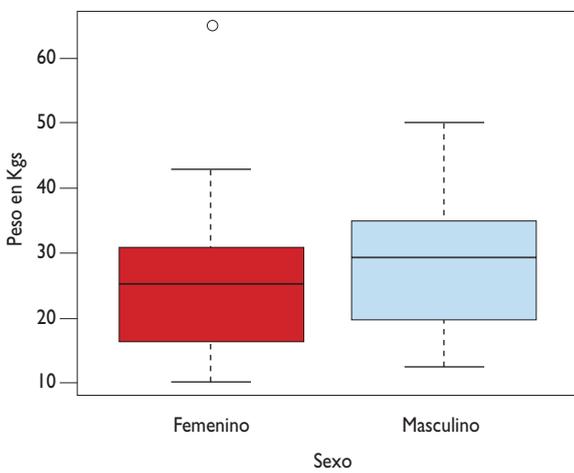


FIGURA 8. DISTRIBUCIÓN DE LA VARIABLE “NÚMERO DE INGRESOS POR ASMA”

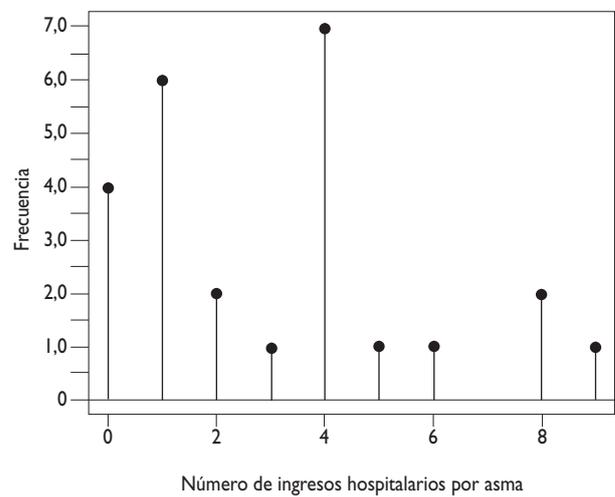


FIGURA 9. DISTRIBUCIÓN DE LA VARIABLE “NÚMERO DE INGRESOS HOSPITALARIOS POR ASMA”

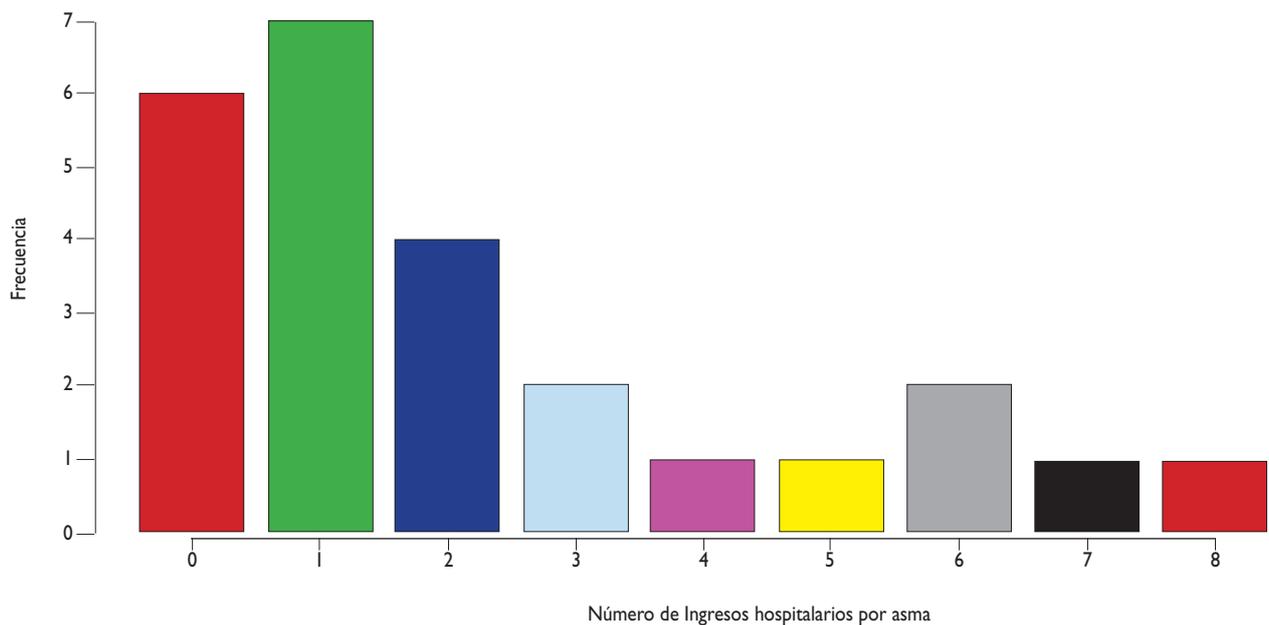
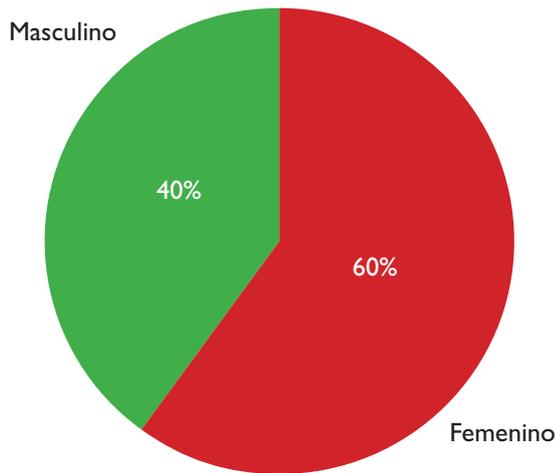


FIGURA 10. DISTRIBUCIÓN DE LA VARIABLE “SEXO”

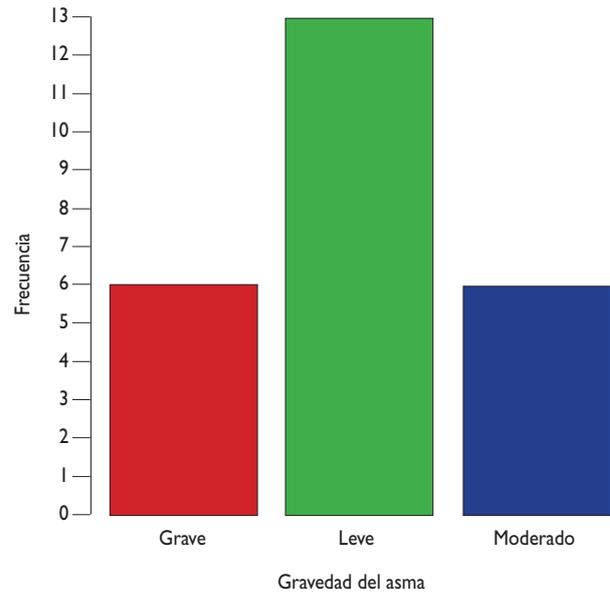


Cualitativas nominales

Ya se ha adelantado que la representación adecuada es por medio del **gráfico de sectores** o circular. Consiste en representar mediante un círculo la distribución de las categorías, donde cada sector representa una categoría y cuya área es directamente proporcional a su frecuencia, con lo que se consigue que los sectores más grandes correspondan con las categorías que presenten mayor frecuencia, esto hace fácilmente reconocible la distribución de los datos mediante un vistazo. Su construcción es muy sencilla, si el total de los datos equivale a 360 grados de la circunferencia, multiplicando la frecuencia relativa de cada variable por 360 grados obtenemos los grados de cada sector.

En nuestro ejemplo, la variable “Sexo” tiene un total de 25 observaciones, de las que 15 son del femenino y 10 del masculino, que corresponden a las frecuencias relativas de 0,6 (15/25) y 0,4 (10/25), respectivamente, y esto equivale a 216 grados (0,6 × 360) para el sector femenino y 144 grados para el masculino

FIGURA 11. DISTRIBUCIÓN DE LA VARIABLE “GRAVEDAD DEL ASMA”



(0,4 × 360). En la figura 10 se representa el gráfico de sectores de la variable “Sexo”. Podemos ver claramente cómo la categoría del sexo femenino tiene mayor frecuencia que la del masculino. Existen gráficos sectoriales en 3D, separados por categorías, pero, a nuestro modo de ver, cuando existen varias categorías (más de siete) pueden ser difíciles de interpretar.

Cualitativas ordinales

Este tipo de variables debe representarse mediante el **gráfico de barras** (*bar chart*), para que los atributos de orden o jerarquía de las categorías no se pierdan. En nuestro ejemplo, la variable

FIGURA 12. DISTRIBUCIÓN DE LA VARIABLE “GRAVEDAD DEL ASMA SEGÚN TABAQUISMO”

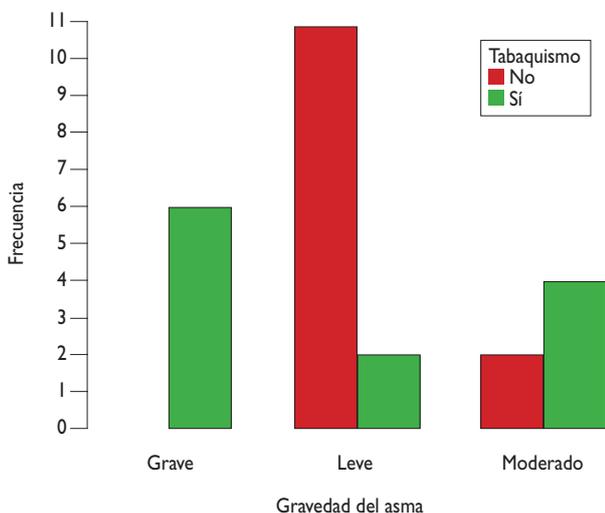


FIGURA 13. DISTRIBUCIÓN DE LA VARIABLE “GRAVEDAD DEL ASMA”

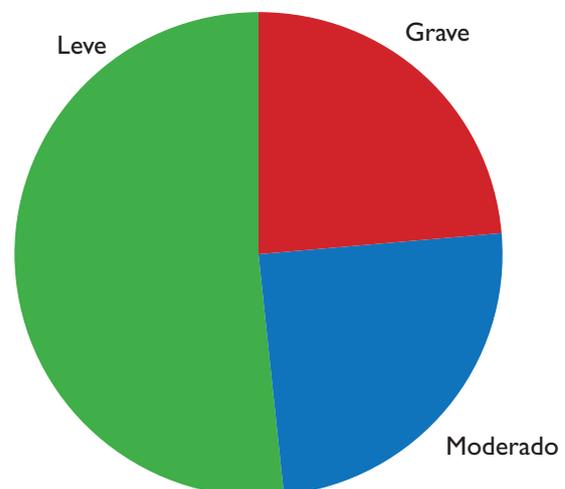
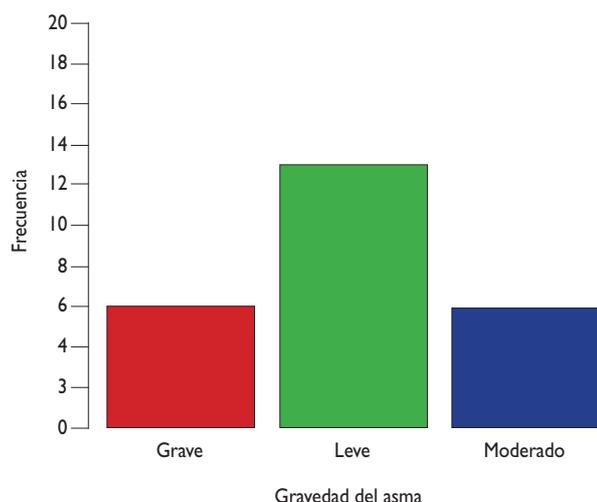


FIGURA 14. DISTRIBUCIÓN DE LA VARIABLE “GRAVEDAD DEL ASMA”

“Asma” se corresponde con la gravedad del asma y consta de tres categorías ordinales: leve, moderada y grave, representado en la figura 11. Observamos cómo la categoría de mayor frecuencia es la de asmáticos leves. Este tipo de gráficos permite estratificar una variable ordinal por otra categórica, en nuestro caso la gravedad del asma según el tabaquismo familiar (figura 12), donde podemos observar que el asma leve es más frecuente en los que no presentan tabaquismo familiar. A nuestro juicio, es un error representar las variables ordinales mediante el gráfico de sectores, ya que, aunque podamos saber qué categoría tiene la mayor frecuencia, se pierde su jerarquía (figura 13). Este tipo de gráficos es fácilmente manipulable para aumentar o disminuir espuriamente las diferencias; en nuestro caso hemos “trucado” la variable cambiando simplemente la escala del eje de ordenadas, consiguiendo disminuir el efecto de las diferencias entre categorías (figura 14).

ERRORES FRECUENTES EN EL USO DE GRÁFICOS

Para terminar, queremos resumir los errores más frecuentes en el uso de los gráficos:

- Histogramas. Construcción cuidadosa para que las áreas de los rectángulos reflejen las frecuencias, no la altura. Utilizarlo únicamente para las variables continuas.
- Gráfico de sectores. Utilizarlo solo para variables nominales, no más de siete categorías y cuanto más sencillo mejor.
- Diagrama de barras. Utilizarlo para variables ordinales y para continuas discretas, el inicio del eje de ordenadas debe ser 0 y utilizar una escala adecuada.
- Común para todos los gráficos. Rotulación de los gráficos, explicar las unidades de medida y no abusar de las escalas cromáticas.

BIBLIOGRAFÍA

- Arriaza Gómez AJ, Fernández Palacín F, López Sánchez M.A, Muñoz Márquez M, Pérez Plaza S, Sánchez Navas S. Estadística básica con R y R-Commander. Cádiz: Universidad de Cádiz; 2008.
- Guisande González C, Vaamonde Liste A. Gráficos estadísticos y mapas con R. Madrid: Díaz de Santos; 2012.
- Martínez González MA, Sánchez-Villegas A, Toledo Atucha E, Faulin Fajardo J. Bioestadística amigable. 3.ª edición. Barcelona: Elsevier; 2014.
- Rendón-Macías ME, Villasís-Keever MÁ, Miranda-Novales MG. Estadística descriptiva. Rev Alerg Mex. 2016;63:397-407.
- Sentís J, Pardell H, Cobo E, Canela J. Manual de Bioestadística. 2.ª edición. Barcelona: Masson; 1995.
- Spriestersbach A, Röhrig B, du Prel JB, Gerhold-Ay A, Blettner M. Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs. Part 7 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009;106:578-83.

ANEXO 1

Para cargar base de datos, cargar **Rcommander** → menú **Data** → **Load data set** → buscar **Base datos gráficos** descargada del enlace y cargarla:

1. Variables cuantitativas continuas:
 - a) Histograma: “Graphics → Histogram → Talla → Options → Frequency counts → Y-axis label: Frecuencias absolutas → X-axis label: Talla → Graph title: Histograma de la variable Talla”.
 - b) Histograma estratificado. Cargar *plugin* EZR. “Graphs and tables → Histogram → Talla → Sexo → Draw in color → Frequency counts”.
 - c) Gráfico de tallo y hoja. “Graphics → Stem → Leaf display → Talla → Modo automático”.
 - d) *Boxplot*: “Graphs → Boxplot → Peso → Options → Automatically → Y-axis label: Peso en kg → Graph title: Histograma de la variable peso”.
2. Variables cuantitativas discretas:
 - a) *Discrete plot*: “Graphs → Plot discrete numeric variables → Número de ingresos, X-axis label = número de ingresos hospitalarios por asma Y-axis label = Frecuencia”.
 - b) Gráfico de barras. Cargar *plugin* EZR, “Graphs and Tables → Bar graph → Frecuencias”.
3. Variables cualitativas nominales
 - a) Gráfico de sectores. “Graphs → Pie chart → Sexo → Graph title: Distribución variable Sexo”.
4. Variables cualitativas ordinales
 - a) Gráfico de barras. “Graphs → Bar graphs → Asma → Frequency → X-axis label = Gravedad del asma → Y-axis label = Frecuencia → Graph title: Distribución de la variable gravedad del asma”.