

AÑO 2014

REPRESENTACIÓN GRÁFICA DE DATOS

Material Complementario

CÁTEDRA: MÉTODOS ESTADÍSTICOS I
CARRERA: LICENCIATURA EN ESTADÍSTICA

Auxiliar 1° Lic. Marcos Prunello

Revisado por:

Lic. Marta Ruggieri

Mgs. Nora Arnesi

Universidad Nacional de Rosario



REPRESENTACIÓN GRÁFICA DE DATOS¹

1. GRÁFICOS PARA VARIABLES CUALITATIVAS

Los dos tipos de gráficos más utilizados para resumir variables cualitativas son el *diagrama de sectores circulares* (también conocido como “gráfico de torta”) y el *gráfico de barras*. Para analizarlos, vamos a trabajar con el siguiente ejemplo.

Ejemplo: Anorexia y Bulimia. Según la Asociación de Lucha contra la Bulimia y la Anorexia de Argentina, las pautas culturales han determinado que la delgadez sea sinónimo de éxito social. Muchos jóvenes luchan para conseguir el “físico ideal” motivados por modelos, artistas o por la publicidad comercial. Durante el mes de octubre del año 2012, dicha asociación recolectó información acerca de 59 personas con síntomas de anorexia que se habían acercado a la institución en busca de ayuda durante los primeros nueve meses de ese año. Una de las variables registradas fue el principal signo visible en la persona relacionado con la enfermedad de anorexia. La siguiente tabla resume esta información.

**PRINCIPAL SIGNO VISIBLE EN PACIENTES CON ANOREXIA
ARGENTINA, OCTUBRE 2012**

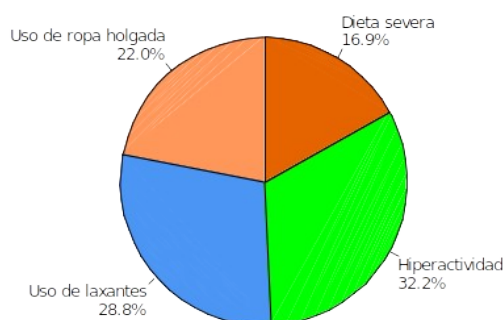
Principal signo visible	Mujeres	Hombres	Total
Dieta severa	6	4	10
Hiperactividad	9	10	19
Uso de laxantes	14	3	17
Uso de ropa holgada	8	5	13
Total	37	22	59

Fuente: Asociación de Lucha Contra la Bulimia y la Anorexia

a. Diagrama de Sectores Circulares

Es un gráfico que presenta la distribución de una variable cualitativa dividiendo un círculo en sectores, cada uno correspondiente a una categoría de la variable, de tamaño proporcional al porcentaje de unidades en ella.

**PRINCIPAL SIGNO VISIBLE EN PACIENTES CON ANOREXIA
ARGENTINA, OCTUBRE 2012**



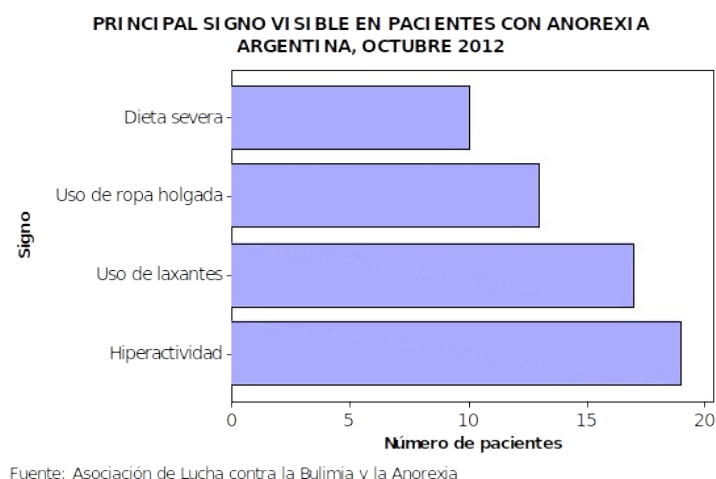
Fuente: Asociación de Lucha contra la Bulimia y la Anorexia

¹El presente material es de carácter complementario a lo desarrollado en clase, no abarca el tema “Estadística Descriptiva” en su totalidad y sólo se enfoca en distintas representaciones gráficas.

Es importante incluir los porcentajes de cada categoría, porque no siempre es claro visualmente si un sector es mayor, menor o igual a otro. Se recomienda evitar el uso de referencias, siendo preferible escribir el nombre de las categorías al lado de cada sector correspondiente.

b. Gráfico de Barras

Este gráfico presenta la distribución de una variable cualitativa listando las categorías de la variable a lo largo del eje vertical y colocando una barra para cada una de ellas, con longitud proporcional a la frecuencia absoluta o relativa de la categoría correspondiente.

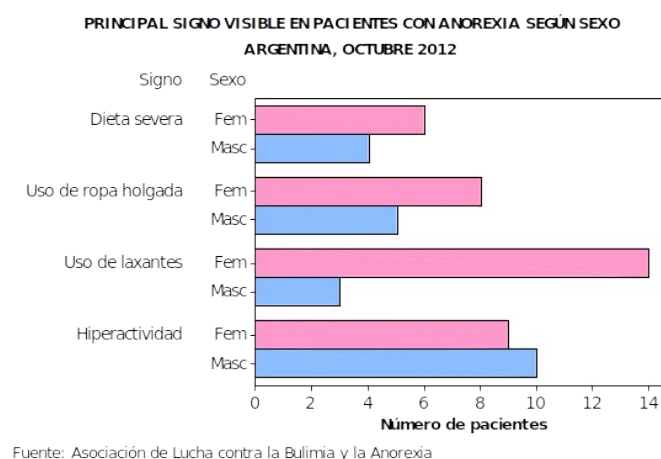


El orden de las categorías en el gráfico no tiene ningún sentido en particular, aunque suelen colocarse en orden creciente o decreciente según sus frecuencias o porcentajes, para lograr una mejor percepción visual. Otra alternativa es ordenarlas alfabéticamente. Si existe un orden natural en las categorías, el mismo se respeta en el gráfico (por ejemplo, para una variable con categorías “Alto”, “Medio” y “Bajo”).

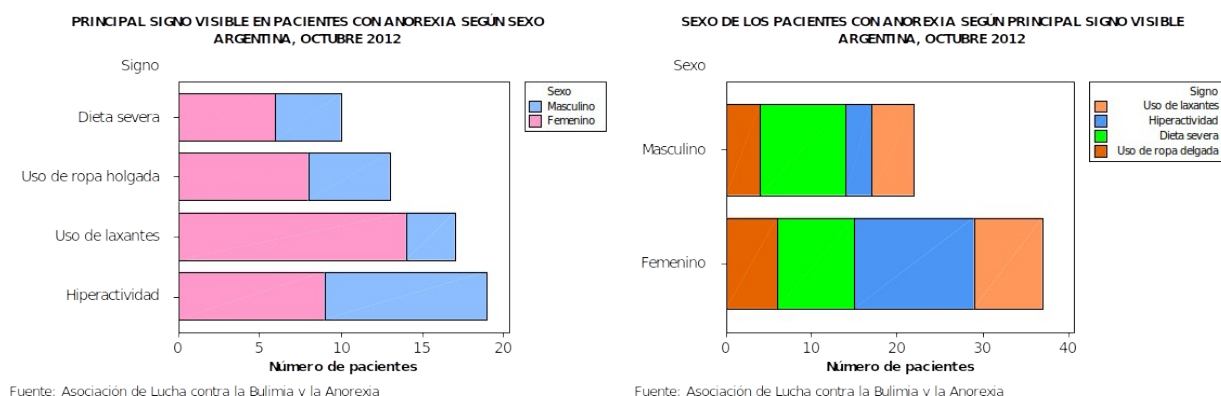
Notas:

- ❖ Un gráfico de barras nos permite comparar varias categorías al contrastar la longitud de las barras. De todos modos, nuestros ojos tienden a enfocarse no sólo en la longitud sino también en el área de las mismas, por lo cual es importante que su ancho sea el mismo para todas.
- ❖ La longitud de la separación entre las barras debe ser menor al ancho de las mismas.
- ❖ La ventaja del gráfico de barras sobre el gráfico de sectores reside en la mejor percepción visual de las diferencias entre las distintas categorías.

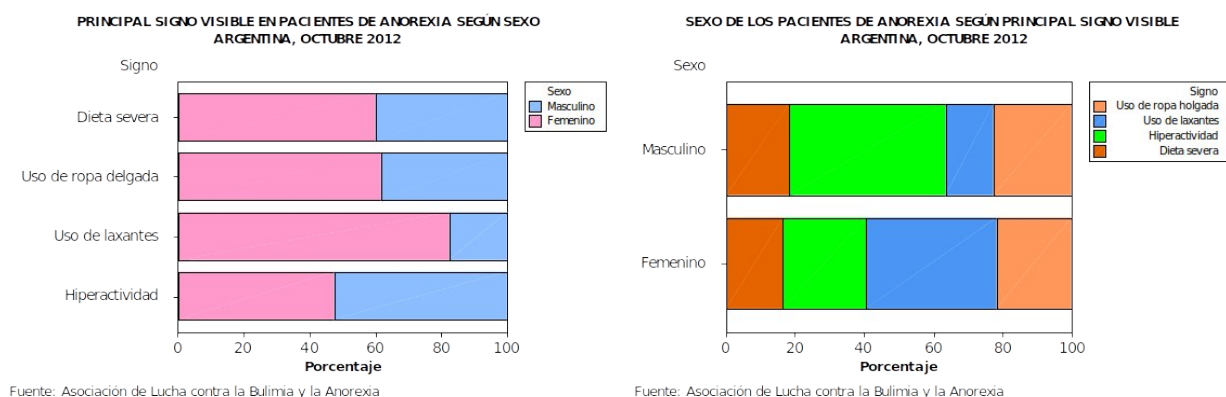
El *gráfico de barras compuesto* permite comparar resultados para diferentes grupos:



En los casos en los que se desea estudiar un total que está compuesto por dos o más partes, se puede utilizar el *gráfico de barras de componentes o subdivididas* como vemos a continuación.



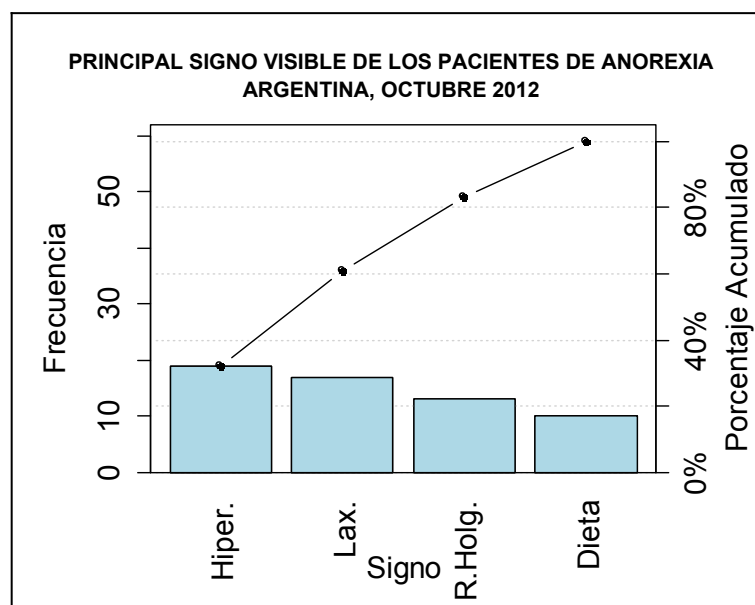
El *gráfico de barras porcentuales* es semejante al de barras subdivididas, pero presenta la frecuencia relativa porcentual. Las barras que representan las categorías de la primera variable de clasificación alcanzan el 100% y dentro de ellas se subdividen las categorías de la segunda variable de clasificación.



El gráfico de la izquierda representa la distribución de la variable *Sexo* en cada nivel de la variable *Principal signo visible* (es decir, la distribución condicional de la variable *Sexo*, dado *Principal signo visible*). El de la derecha representa la distribución de la variable *Principal signo visible* para cada categoría de la variable *Sexo* (es decir, la distribución de la variable *Principal signo visible* en cada nivel de la variable *Sexo*).

c. Diagrama de Pareto

Es un tipo particular de gráfico de barras verticales, con las categorías ordenadas por su frecuencia absoluta o relativa, desde la mayor hasta la menor, frecuentemente acompañado por un polígono acumulativo. Este gráfico representa el “Principio de Pareto”, que afirma que un subconjunto pequeño de categorías suele contener la mayoría de las observaciones. Se utiliza principalmente en control de calidad. En el ejemplo, el siguiente gráfico muestra que hiperactividad y uso de laxantes son los principales signos visibles en el 60% de los pacientes anoréxicos.



2. GRÁFICOS PARA VARIABLES CUANTITATIVAS

Ahora nos centraremos en la presentación gráfica de datos de variables cuantitativas. Pondremos en consideración distintos tipos de gráficos y analizaremos para qué fines son convenientes cada uno.

La siguiente tabla presenta otros datos registrados sobre las personas que acudieron a la Asociación de Lucha contra la Bulimia y la Anorexia.

**SEXO, EDAD Y NÚMERO DE VISITAS DURANTE LOS PRIMEROS 9 MESES DE 2012
PACIENTES CON ANOREXIA
ARGENTINA, OCTUBRE 2012**

Ind	Sexo	Edad	Nº de visitas a la Asociación	Ind	Sexo	Edad	Nº de visitas a la Asociación	Ind	Sexo	Edad	Nº de visitas a la Asociación	Ind	Sexo	Edad	Nº de visitas a la Asociación
1	F	22	2	16	F	13	2	31	F	18	2	46	M	15	1
2	F	15	2	17	F	12	1	32	F	16	2	47	M	15	1
3	F	28	1	18	F	18	3	33	F	18	1	48	M	31	3
4	F	15	3	19	F	13	2	34	F	16	3	49	M	16	3
5	F	17	1	20	F	17	1	35	F	20	2	50	M	15	1
6	F	24	2	21	F	14	1	36	F	18	3	51	M	16	1
7	F	20	1	22	F	12	3	37	F	16	4	52	M	33	3
8	F	17	2	23	F	15	1	38	M	17	2	53	M	16	2
9	F	18	1	24	F	19	2	39	M	18	3	54	M	19	2
10	F	17	2	25	F	26	3	40	M	15	3	55	M	17	5
11	F	13	1	26	F	21	2	41	M	23	2	56	M	22	3
12	F	21	2	27	F	19	2	42	M	18	3	57	M	14	2
13	F	20	1	28	F	14	3	43	M	26	2	58	M	16	2
14	F	16	4	29	F	18	2	44	M	17	3	59	M	16	2
15	F	15	1	30	F	12	1	45	M	29	2				

Fuente: Asociación de Lucha Contra la Bulimia y la Anorexia

Antes de presentar distintos gráficos, es útil resumir los datos en tablas de frecuencias como las que se ejemplifican a continuación.

NÚMERO DE VISITAS DURANTE LOS PRIMEROS 9 MESES DE 2012

ARGENTINA

Nº de Visitas	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
1	17	0.2881	17	0.2881
2	24	0.4068	41	0.6949
3	15	0.2542	56	0.9492
4	2	0.0339	58	0.9831
5	1	0.0169	59	1.0000
Total	59	1		

Fuente: Asociación de Lucha Contra la Bulimia y la Anorexia

EDAD DE LOS PACIENTES CON ANOREXIA

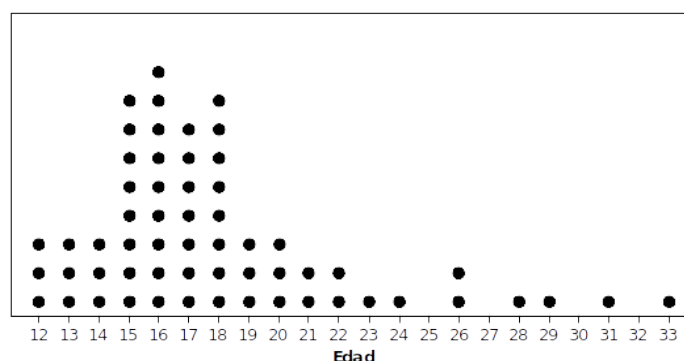
ARGENTINA, OCTUBRE 2012

Intervalo (edad)	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulada	Frecuencia Relativa Acumulada
[11;14)	6	0,1017	6	0,1017
[14;17)	20	0,3390	26	0,4407
[17;20)	18	0,3051	44	0,7458
[20;23)	7	0,1186	51	0,8644
[23;26)	2	0,0339	53	0,8983
[26;29)	3	0,0508	56	0,9492
[29;32)	2	0,0339	58	0,9831
[32;35)	1	0,0169	59	1,0000
Total	59	1		

Fuente: Asociación de Lucha Contra la Bulimia y la Anorexia

a. Gráfico de Frecuencias

Es un modo sencillo de representar la distribución de un conjunto de datos no demasiado grande. Se traza una línea numérica horizontal donde se representan valores de la variable, desde el mínimo hasta el máximo. Para cada observación se coloca un punto o cruz sobre su valor en dicha recta numérica. Para el ejemplo de las edades, el siguiente gráfico de frecuencias nos indica que las edades observadas varían entre 12 y 33 años, con una mayor concentración entre 15 y 18 años. La distribución de esta variable presenta forma asimétrica hacia la derecha.

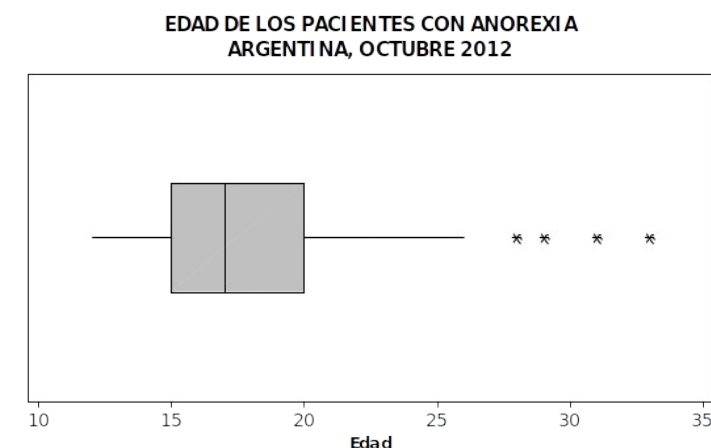
EDAD DE LOS PACIENTES CON ANOREXIA
ARGENTINA, OCTUBRE 2012

Fuente: Asociación de Lucha contra la Bulimia y la Anorexia

b. Diagrama de Tallo y Hojas

Cada dato se representa con un tallo y una hoja. Los tallos suelen ser todos los dígitos excepto el último, (la hoja) y se listan a la izquierda de una recta vertical y las hojas a la derecha. Una hoja tiene sólo un dígito,

distribución, desde Q_1 hasta Q_3 . Las líneas que se extienden desde la caja se llaman *bigotes*; el izquierdo se extiende desde el límite inferior de la caja hasta la menor observación que no sea un potencial outlier y el derecho desde el límite superior de la caja hasta la mayor observación que no sea un potencial outlier. Dichos potenciales outliers u observaciones extremas (se encuentran a más de $1.5 \times IQR^2$ debajo de Q_1 o encima de Q_3) se representan con puntos separados de los bigotes.



Observación: Los cinco números resumen que se muestran en este boxplot son: mín.=12 años; Q_1 =15 años; Q_2 =17 años; Q_3 =20 años y máx.=33 años. El 50% central de las edades de los pacientes con anorexia varía entre 15 y 20 años, que son los extremos de la caja. La mediana de 17 años se indica con la línea vertical dentro de la caja.

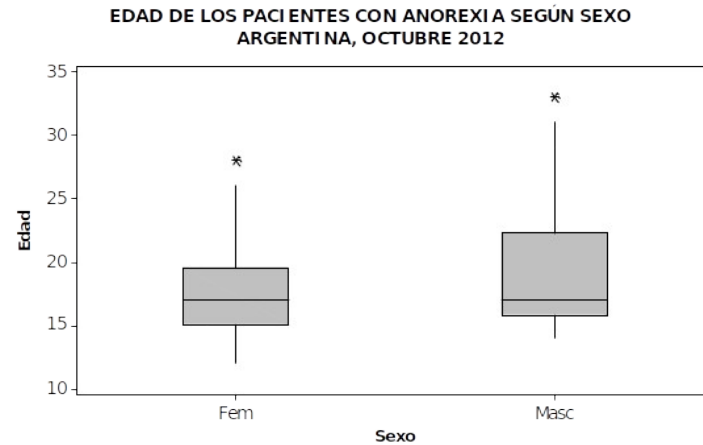
Además: $IQR = Q_3 - Q_1 = 5$; $1.5 \times IQR = 7$; $Q_1 - 7 = 8$; $Q_3 + 7 = 27$. Entonces, de acuerdo con el criterio de $1.5 \times IQR$, hay cuatro personas cuyas edades son valores atípicos en relación al resto de los datos.

Notas:

- ❖ Las observaciones extremas se señalan separadamente para identificarlas: ¿ese valor fue ingresado incorrectamente?, ¿la unidad era especialmente diferente al resto en algún sentido? Otro motivo es que estas observaciones no proveen demasiada información sobre la forma de la distribución; por el contrario, pueden dar una falsa idea de asimetría. Sin embargo, forman parte de los datos recolectados y no deben ser descartados (a menos que se tenga una firme justificación).
- ❖ Un boxplot no refleja ciertas características de una distribución tan claramente como lo hace un histograma. Por ejemplo, no podríamos notar si estamos ante una distribución bimodal. Sin embargo, un boxplot suele ser útil para analizar la asimetría de una distribución a través de la longitud relativa de los bigotes y de las dos partes de la caja.

Los *boxplots comparativos* son muy útiles para la comparación gráfica de distribuciones en grupos distintos:

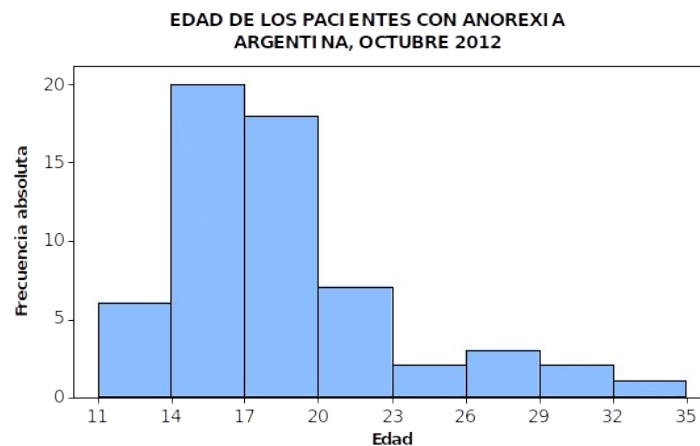
² IQR: Rango intercuartílico, se define como $Q_3 - Q_1$.



Fuente: Asociación de Lucha contra la Bulimia y la Anorexia

d. Histograma

Un histograma divide el rango de variación de la variable en intervalos o clases mutuamente excluyentes y exhaustivas y representa la frecuencia absoluta (o relativa) correspondiente con un rectángulo de área proporcional. Se utiliza con variables continuas o discretas que asumen muchos valores distintos.

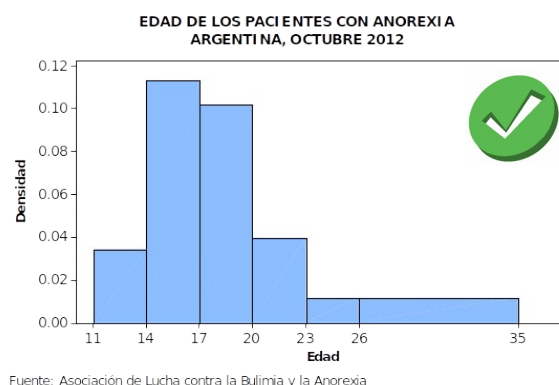
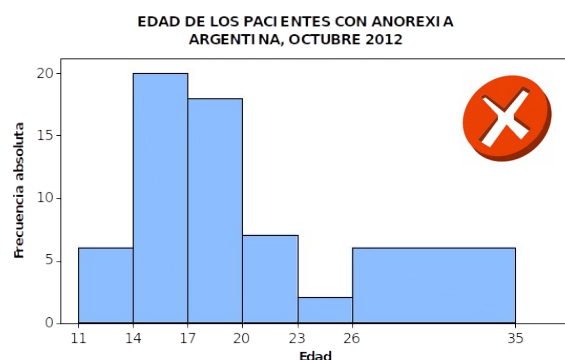


Fuente: Asociación de Lucha contra la Bulimia y la Anorexia

Notas:

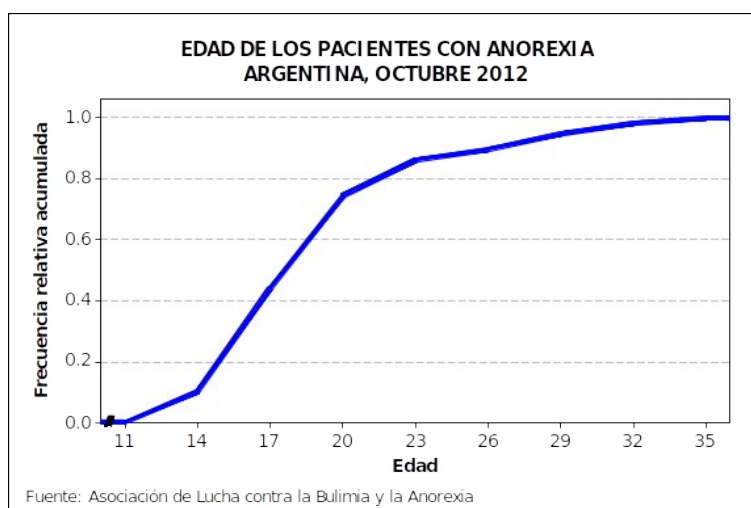
- ❖ El histograma no muestra los valores numéricos reales. Por ejemplo, sabemos que hay 1 persona con edad mayor o igual a 32 y menor a 35, pero no sabemos exactamente qué valor es.
- ❖ El área de las barras, no la altura, es proporcional a la frecuencia.
- ❖ El histograma del ejemplo tiene en su eje vertical las frecuencias absolutas: siendo todos los intervalos de igual amplitud y considerando a la misma como una unidad, el área total es igual al número de observaciones. Si, en cambio, se representara la frecuencia relativa, el área total es igual a 1 y el gráfico tendría la misma apariencia.
- ❖ En cuanto a la determinación del número de intervalos, se debe tener en cuenta que si se usan pocos o demasiados, se podría perder información sobre la forma de la distribución. Aunque no hay un único criterio, un método es establecer como cantidad aproximada de intervalos a la raíz cuadrada del número de observaciones, con amplitud igual al rango dividido por la cantidad de intervalos.

- ❖ El gráfico de frecuencias y el diagrama de tallo y hojas son útiles para conjuntos de datos pequeños, dado que representan cada observación individual. Con un conjunto de datos grandes, los histogramas son más convenientes, al poder presentar toda la información de un modo más compacto.
- ❖ Si se tienen intervalos de distinta amplitud, para que el área de las barras sea proporcional a la frecuencia, en el eje vertical se debe representar la densidad (frecuencia relativa el intervalo dividida por la amplitud del mismo), de modo que el área total sea igual a 1. No se debe graficar la frecuencia absoluta o relativa, ya que conducen a interpretaciones erróneas. Por ejemplo:



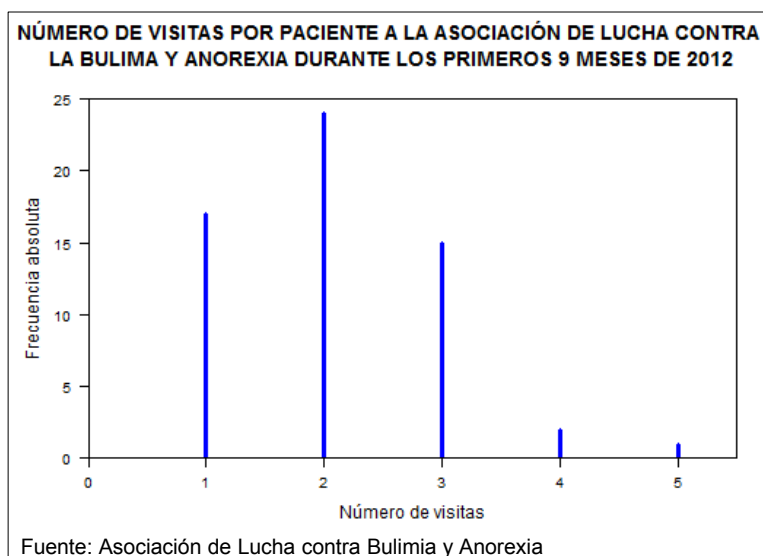
e. Polígono acumulativo u ojiva

El histograma suele estar acompañado por un polígono acumulativo que representa las frecuencias absolutas o relativas acumuladas. Sobre el límite superior de cada intervalo (en el eje horizontal) se marca un punto cuya posición con respecto al eje de ordenadas es la frecuencia acumulada correspondiente. Estos puntos se unen con segmentos. Con este gráfico es posible observar qué porcentajes de pacientes tienen edades menores o iguales a un valor determinado.



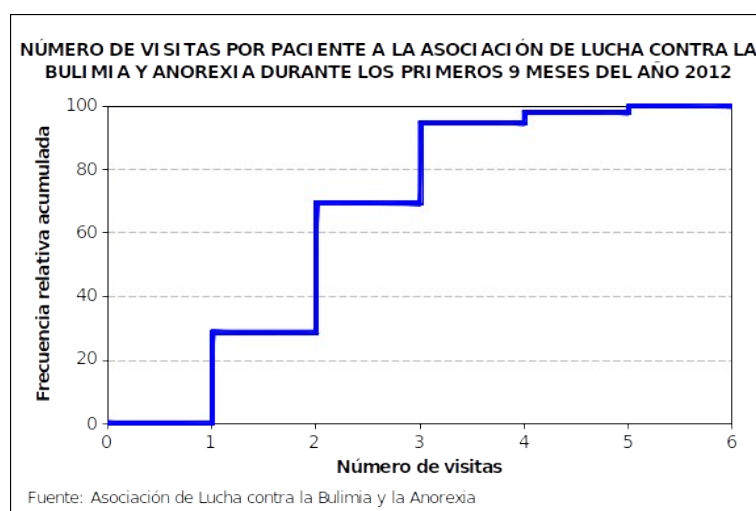
f. Gráfico de bastones

Se emplea para representar gráficamente datos de variables discretas (siempre que estas no tomen demasiados valores distintos). En el eje horizontal se representan los valores observados y en el vertical las frecuencias absolutas o relativas. Sobre cada valor observado (eje horizontal) se grafica una línea vertical, el "bastón", cuya altura es igual a la frecuencia absoluta o relativa correspondiente (eje vertical). Este gráfico permite distinguir de forma rápida algunas características importantes, como la forma de la distribución, el mínimo, el máximo y el modo, entre otros.



g. Gráfico escalonado

Se utiliza para mostrar la distribución acumulada de una variable discreta.



Observación: muchos autores sugieren que los segmentos verticales del gráfico anterior estén marcados con líneas de puntos o directamente no marcados.

3. SUGERENCIAS PARA LA CONSTRUCCIÓN DE GRÁFICOS

- ❖ Proveer siempre un título, que deje en claro lo que el gráfico quiere representar, así como también la fuente de los datos.
- ❖ Nombrar ambos ejes, con las unidades de medida correspondientes.
- ❖ En general, comenzar el eje vertical en 0, para lograr una comparación visual de los tamaños relativos adecuadamente.
- ❖ Ser cuidadosos en el uso de dibujos en lugar de las usuales barras o puntos. Pueden hacer más atractivo a un gráfico, pero también pueden llevar a malas interpretaciones.
- ❖ Chequear que los ejes mantengan una escala constante.
- ❖ No presentar más de un grupo por gráfico cuando los valores de la variable difieren mucho entre ellos. Se debe considerar la utilización de gráficos separados o de medidas relativas, como porcentajes.

BIBLIOGRAFÍA

Agresti, A., & Franklin, C. (2009). *Statistics. The art and science of learning from data*. Pearson Prentice Hall.

Aliaga, M., & Gunderson, B. (2006). *Interactive Statistics*. New Jersey: Pearson Prentice Hall.