

Organización y Representación de Datos

Gráfica

Dependiendo del tipo de variable, categórica o numérica, se pueden hacer diversas representaciones gráficas de los datos.

Barras

La representación gráfica más común para datos categóricos es el diagrama de barras. Se construye ya sea mediante la frecuencia (cantidad) de cada una de las categorías o bien mediante la proporción o frecuencia relativa al total de casos.

Por ejemplo, en el estacionamiento de un centro comercial, se contó la cantidad de coches estacionados que había por color. El cuadro [1.1](#) muestra color, frecuencia y proporción de cada coche y la figura [1.1](#) el diagrama de barras con la proporción por color.

Cuadro 1.1: Coches por color

COLOR	FREC	PROP
AMARILLO	13	0.09
AZUL	15	0.10
BLANCO	24	0.16
GRIS	25	0.17
NEGRO	42	0.28
ROJO	23	0.15
VERDE	8	0.05

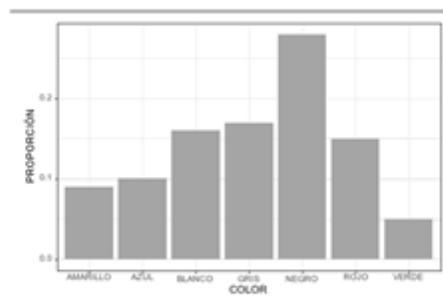


Figura 1.1: Coches por color

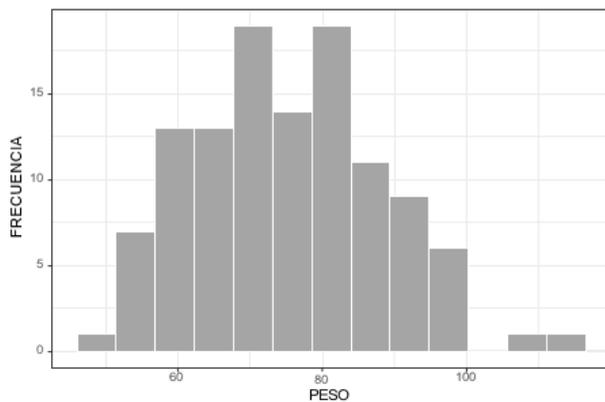
Dependiendo del uso de una gráfica, se puede aplicar algún estilo, por ejemplo, que el color de la barra corresponda al color de la categoría correspondiente. Sin embargo, para uso técnico se recomienda un diseño sobrio donde se haga énfasis en la distribución de frecuencias más en el diseño gráfico.

Histograma

En datos continuos, la gráfica más común es el histograma, el cual consiste en dividir todo el rango de valores de la muestra en intervalos usualmente equidistantes. Se recomienda su uso para una cantidad grande de datos para que sea informativa de la variable que representa.

La cantidad de intervalos o clases en un histograma puede variar; una cantidad razonable varía entre 9 y 13 aproximadamente. Los programas de estadística tienen algoritmos integrados para determinar la cantidad de clases.

Se toma una muestra aleatoria de estudiantes de distintas facultades de la UAdeC campus Arteaga y se les pregunta su peso. Se selecciona a hombres y se construye el histograma como se muestra en la siguiente imagen.



Caja

Una gráfica que aporta información complementaria al histograma es la de caja. Esta incorpora cuartiles y puntos atípicos.

Se toman los datos del ejemplo [6](#), pero se incorpora el peso de mujeres para comparar las gráficas de caja entre sexo. El diagrama de caja o boxplot se muestra en la figura [1.3](#).

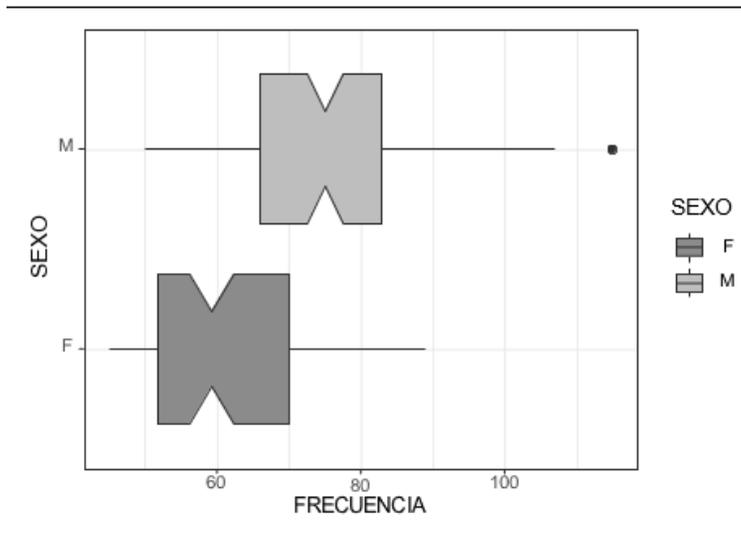


Figura 1.3: Peso de estudiantes UAdC

Los límites de la caja corresponden al primer y tercer cuartil respectivamente; la línea intermedia a estos es la mediana o segundo cuartil. Se observa también un punto a la derecha de la gráfica de hombres, que corresponden a los puntos atípicos identificados. Los límites de la línea transversal a los cuartiles son el mínimo y máximo. Para el caso de puntos atípicos, esos son los límites.

En un diagrama de caja se pueden visualizar otras características, por ejemplo, la asimetría. Otro elemento opcional son las muescas que presentan un intervalo para la mediana y sirve para comparar grupos en términos de su comportamiento promedio.

En el ejemplo se muestra que el peso promedio de las mujeres es significativamente menor que el de los hombres. Evidentemente esto se debe también a la diferencia de estaturas, por lo que no se puede emitir algún juicio utilizando solo la variable peso ya que los grupos no son comparables; para

esto se requiere de indicadores más complejos que garanticen la comparabilidad.

Dispersión

En el estudio de poblaciones, la relación entre variables es uno de los aspectos más relevantes. Por ejemplo, identificar la relación peso - estatura entre personas, o bien establecer la relación entre el consumo de carbohidratos y el sobrepeso.

Una de las gráficas recomendadas y más utilizadas para visualizar la relación entre dos variables numéricas es el diagrama de dispersión. Este tipo de representaciones consiste en ubicar a los datos como puntos en un plano cartesiano donde cada una de las variables corresponde a uno de los ejes coordenados.

Se toman los datos del ejemplo [6](#) para explorar la relación entre peso y estatura de hombres y mujeres. En la figura [1.4](#) se muestra el resultado.

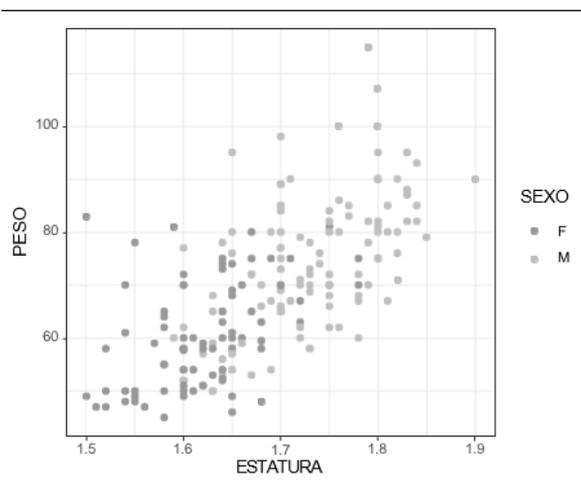


Figura 1.4: Peso y estatura de estudiantes UAdC

Para que la gráfica aporte más información se difieren los sexos por color.

La primera relación evidente es la tendencia de que, a mayor estatura, mayor peso. Este comportamiento es similar tanto en hombres como mujeres, con la diferencia que el conjunto de puntos tiende a concentrarse en distinta posición del sistema coordenado.

Se pueden analizar otros aspectos, como identificar posibles errores en la captura de datos. Notemos el punto correspondiente a una mujer ubicado en la parte izquierda y en la parte media del eje vertical. Dicho dato correspondería a una estatura aproximada de 1.50m y un peso de 83kg surge la pregunta de si es factible esa combinación de las dos variables.

Series de tiempo

Diversos fenómenos o procesos son monitoreados a través del tiempo por intervalos lo que se denomina serie de tiempo. La temperatura, pérdida de peso, ingresos, etc. Su representación más común a través de puntos conectados por segmentos de recta.

El producto interno bruto (PIB) per cápita, ingreso per cápita o renta per cápita es un indicador económico que mide la relación entre el nivel de renta de un país y su población.

En la figura [1.5](#) se muestra el PIB per cápita del año 1960 al 2010. Se incluyen los países de Canadá (CA), China (CN), Japón (JP), México (MX), y Estados Unidos (US).

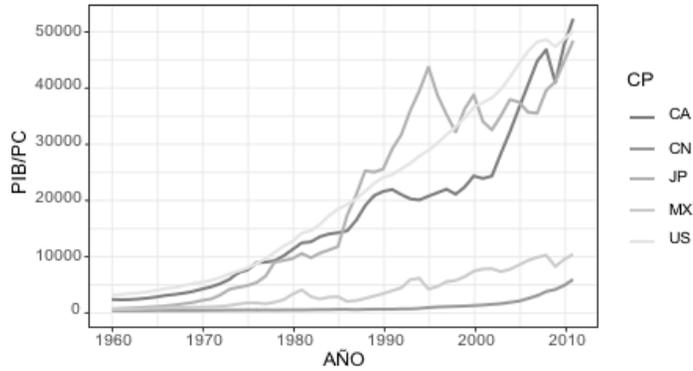


Figura 1.5 Producto Interno Bruto Per cápita

Entre los países seleccionados los PIB per cápita más bajos corresponden a China y México. Al 2010, Canadá, Japón y Estados Unidos tienen valores similares.

Una gráfica de serie de tiempo permite visualizar tendencias y son el principal esquema (representación gráfica) de monitoreo de la economía en un país. Existen una variedad de modelos principalmente para pronóstico.

Referencias:

Hoel, P. G. (1984). Elementary Statistics. John Wiley & Sons.