

PRESENTACIÓN DE DATOS

¿QUÉ ES LA PRESENTACIÓN DE DATOS?



La presentación de datos estadísticos constituye en sus diferentes modalidades uno de los aspectos de más uso en la estadística descriptiva. A partir podemos visualizar a través de los diferentes medios escritos y televisivos de comunicación masiva la presentación de los datos estadísticos sobre el comportamiento de las principales variables económicas y sociales, nacionales e internacionales.

Ejemplo 1.5. La siguiente tabla muestra los resultados de las enfermedades más comunes que tienen los pacientes en un hospital.

TABLA DE NUMEROS DE PERSONAS CON DICHAS ENFERMEDADES.

VARIABLE CUALITATIVA: ENFERMEDADES	FRECUENCIA	GRADOS	PORCENTAJE
HEPATITIS	25	59.21°	16%
SARAMPION	40	94.73°	26%
GRUPE	65	151.94°	43%
VARICELA	18	42.63°	12%
SIDA	4	9.47°	3%
TOTAL	152	360°	100%

Tabla 1.1: número de personas con ciertas enfermedades

TIPOS DE PRESENTACIÓN:



Presentación escrita: Esta forma de presentación de informaciones se usa cuando una serie de datos incluye pocos valores, por lo cual resulta mas apropiada la palabra escrita como forma de escribir el comportamiento de los datos; mediante la forma escrita, se resalta la importancia de las informaciones principales.



Presentación tabular: Cuando los datos estadísticos se presentan a través de un conjunto de filas y de columnas que responden a un ordenamiento lógico; es de gran uso e importancia para el usuario ya que constituye la forma más exacta de presentar las informaciones. Una tabla consta de varias partes:

ELEMENTOS DE ESTADÍSTICA

DIPLOMA EN INGENIERÍA DE FORMACIÓN

REPRESENTACIÓN TABULAR

CLASE	FRECUENCIA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA ACUMULADA
Bachiller	40	0,33	40	0,33
Técnico	27	0,23	67	0,56
Tecnólogo	27	0,23	94	0,78
Profesional	20	0,17	114	0,95
Posgrado	6	0,05	120	1,00

En un conjunto de datos, se define *moda* como el valor de mayor frecuencia. En nuestro ejemplo, Bachiller es la clase modal.



Cuerpo: El cuerpo contiene todas las informaciones numéricas que aparecen en la tabla.

Título: Es la parte más importante del cuadro y sirve para describir todo el contenido de este.

Fuente: La fuente de los datos contenidos en la tabla indica la procedencia de estos.

Encabezados: Son los diferentes subtítulos que se colocan en la parte superior de cada columna.

Notas al pie: Son usadas para hacer algunas aclaraciones sobre aspectos que aparecen en la tabla o cuadro y que no han sido explicados en otras partes.

Columna matriz: Es la columna principal del cuadro.

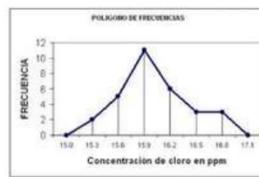




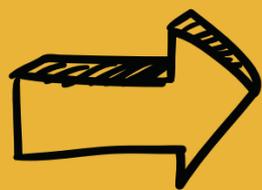
Presentación gráfica: Proporciona al lector o usuario mayor rapidez en la comprensión de los datos, una gráfica es una expresión artística usada para representar un conjunto de datos.

REPRESENTACIÓN GRÁFICA DE LOS DATOS

El **histograma** es la herramienta más común de representar gráficamente la distribución de frecuencia de los datos. Se lo construye dibujando rectángulos cuya base corresponde a cada intervalo de clase, y su altura según el valor de la frecuencia. Puede ser la frecuencia absoluta o la frecuencia relativa.



El **polígono de frecuencias** es una manera de representar el perfil de la distribución de los datos. Se obtiene uniendo mediante segmentos de recta los puntos (marca de clase, frecuencia). Para cerrar el polígono se puede agregar un punto a cada lado con frecuencia 0.



ESTADÍSTICA DESCRIPTIVA

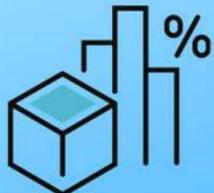
¿QUÉ ES LA ESTADÍSTICA DESCRIPTIVA?

El término estadística descriptiva se refiere al análisis, el resumen y la presentación de los resultados relacionados con un conjunto de datos derivados de una muestra o de toda la población.

La estadística descriptiva comprende tres categorías principales: distribución de frecuencias, medidas de tendencia central y medidas de variabilidad.

Por estadística descriptiva entendemos, por ejemplo, el cálculo de la media y la mediana, dos indicadores muy importantes y sobre todo diferentes. La mediana es un indicador que «no tiene en cuenta los valores extremos, a veces poco frecuentes», a diferencia de la media, que está muy influida por estos valores extremos

Estadística descriptiva



OBJETIVO DE LA ESTADÍSTICA DESCRIPTIVA

El objetivo de la estadística descriptiva es describir los datos observados de forma sintética y significativa para poder analizarlos mejor. Es recoger observaciones sobre sujetos con una determinada propiedad y traducir estas observaciones en números que proporcionen información sobre dicha propiedad.

IMPORTANCIA DE LA ESTADÍSTICA DESCRIPTIVA

La estadística descriptiva facilita la visualización de los datos. Permiten presentarlos de forma significativa y comprensible, lo que a su vez da pie a una interpretación simplificada del conjunto de datos en cuestión.

Los datos brutos serían difíciles de analizar, y la determinación de tendencias y patrones puede ser un reto. Además, los datos en bruto dificultan la visualización de lo que muestran los datos.

Además, el uso de la estadística descriptiva permite resumir y presentar un conjunto de datos mediante una combinación de descripciones tabuladas y gráficas. La estadística descriptiva se utiliza para resumir datos cuantitativos complejos

ESTADÍSTICA DESCRIPTIVA



Se encarga de recoger, almacenar, ordenar, realizar tablas o gráficos y calcular parámetros básicos sobre el conjunto de datos.



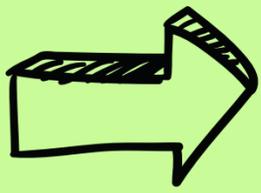
Es junto a la inferencia estadística una de las dos grandes ramas de la estadística.



Su propio nombre lo indica, trata de describir algo. Pero no describirlo de cualquiera forma, sino de manera cuantitativa.



Responde a preguntas como por ejemplo: ¿cuál es la altura promedio en centímetros de una población?



TIPOS DE ESTADISTICA

DISTRIBUCIÓN DE FRECUENCIAS

La distribución de frecuencias se presenta normalmente en una tabla o un gráfico. Cada entrada de la tabla o el gráfico va acompañada del recuento o la frecuencia de aparición de los valores, en un intervalo, rango o grupo específico.

La distribución de frecuencias es básicamente una presentación o un resumen de datos agrupados que se han clasificado en función de clases mutuamente excluyentes y del número de ocurrencias en cada clase respectiva. Permite una forma más estructurada y organizada de presentar los datos en bruto.

Los cuadros y gráficos más comunes utilizados en la presentación y visualización de la distribución de frecuencias incluyen gráficos de barras, histogramas, gráficos circulares y gráficos de líneas.

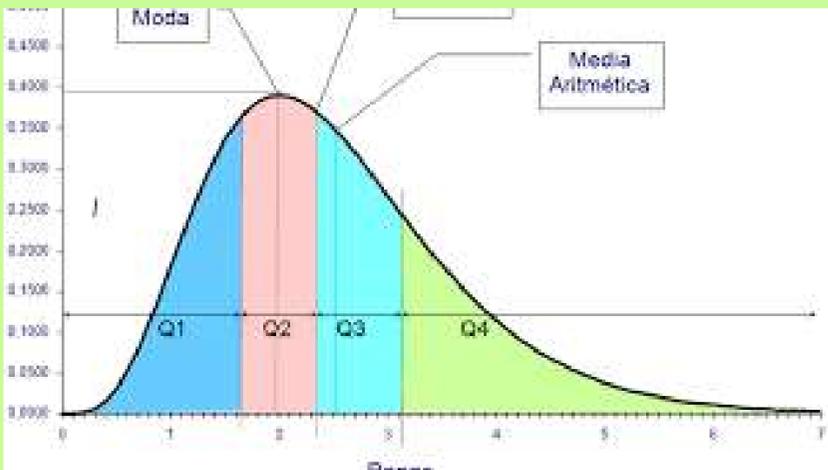


TENDENCIA CENTRAL

La tendencia central es otro de los tipos de estadística descriptiva, y se refiere al resumen descriptivo de un conjunto de datos utilizando un único valor que refleja el centro de la distribución de los datos.

Las medidas de tendencia central también se conocen como medidas de localización central. La media, la mediana y la moda son consideradas las medidas de tendencia central.

La media, que se considera la medida de tendencia central más popular, es el valor medio o más común en un conjunto de datos. La mediana se refiere a la puntuación media de un conjunto de datos en orden ascendente. La moda se refiere a la puntuación o valor más frecuente en un conjunto de datos.

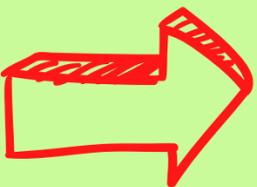


VARIABILIDAD

Una medida de variabilidad es una estadística de resumen que refleja el grado de dispersión de una muestra. Las medidas de variabilidad determinan la distancia que los puntos de datos parecen tener con respecto al centro.

La dispersión y la variabilidad se refieren y denotan el rango y la amplitud de la distribución de los valores en un conjunto de datos. El rango, la desviación estándar y la varianza se utilizan, respectivamente, para representar diferentes componentes y aspectos de la dispersión.

El rango representa el grado de dispersión o un ideal de la distancia entre los valores más altos y más bajos dentro de un conjunto de datos. La desviación estándar se utiliza para determinar la varianza media de un conjunto de datos y proporciona una idea de la distancia o la diferencia entre un valor de un conjunto de datos y el valor medio del mismo conjunto de datos.



TIPOS DE GRAFICAS

Gráfico de barras

El más conocido y utilizado de todos los tipos de gráficos es el gráfico o diagrama de barras. En éste, se presentan los datos en forma de barras contenidas en dos ejes cartesianos (coordenada y abscisa) que indican los diferentes valores. El aspecto visual que nos indica los datos es la longitud de dichas barras, no siendo importante su grosor.

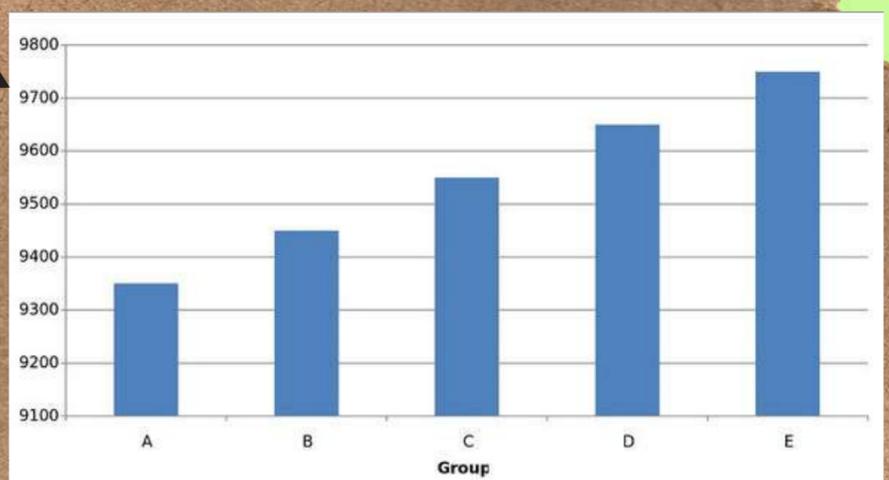
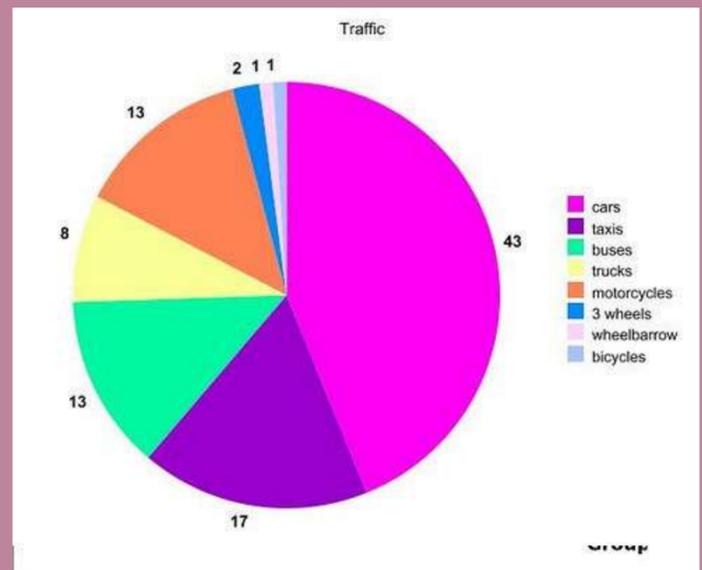


Gráfico circular o por sectores

El también muy habitual gráfico en forma de “quesito”, en este caso la representación de los datos se lleva a cabo mediante la división de un círculo en tantas partes como valores de la variable investigada y teniendo cada parte un tamaño proporcional a su frecuencia dentro del total de los datos.



Histograma

Aunque a simple vista muy semejante al gráfico de barras, el histograma es uno de los tipos de gráfica que a nivel estadístico resulta más importante y fiable. En esta ocasión, también se utilizan barras para indicar a través de ejes cartesianos la frecuencia de determinados valores, pero en vez de limitarse a establecer la frecuencia de un valor concreto de la variable evaluada refleja todo un intervalo.

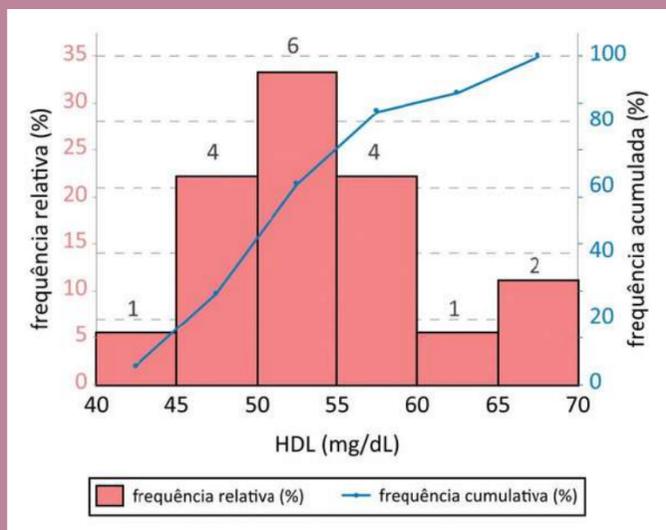


Gráfico de líneas

En este tipo de gráfico se emplean líneas para delimitar el valor de una variable dependiente respecto a otra independiente. También puede usarse para comparar los valores de una misma variable o de diferentes investigaciones utilizando el mismo gráfico (usando diferentes líneas). Es usual que se emplee para observar la evolución de una variable a través del tiempo.

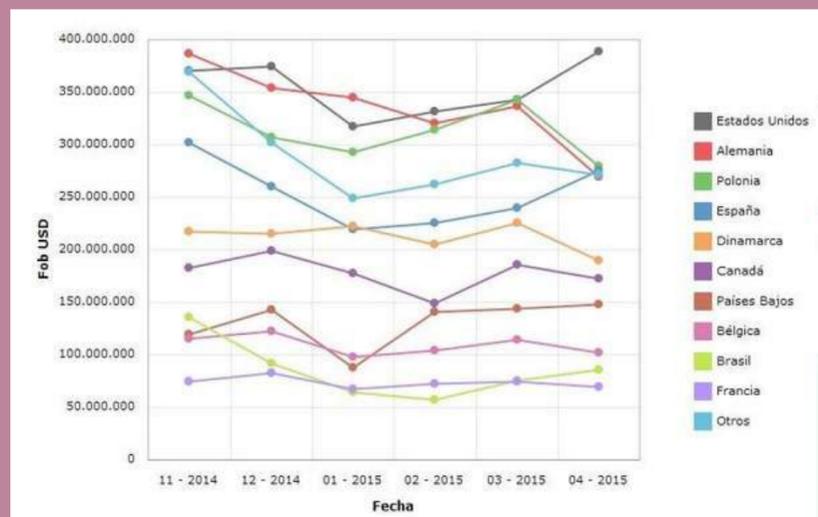


Gráfico de dispersión

El gráfico de dispersión o gráfico xy es un tipo de gráfico en el cuál mediante los ejes cartesianos se representa en forma de puntos todos los datos obtenidos mediante la observación. Los ejes x e y muestran cada uno los valores de una variable dependiente y otra independiente o dos variables de la que se esté observando si presentan algún tipo de relación.

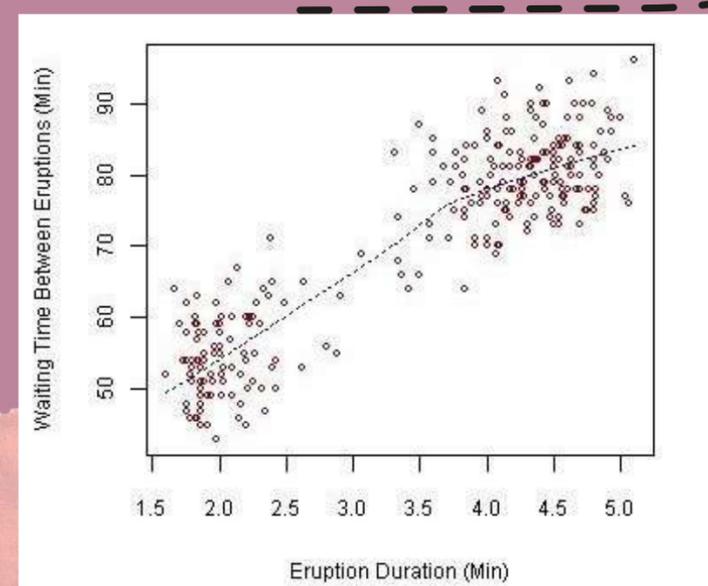


Gráfico de caja y bigotes

Los gráficos de caja son uno de los tipos de gráficas que tienden a utilizarse de cara a observar la dispersión de los datos y cómo éstos agrupan sus valores. Se parte del cálculo de los cuartiles, los cuales son los valores que permiten dividir los datos en cuatro partes iguales.

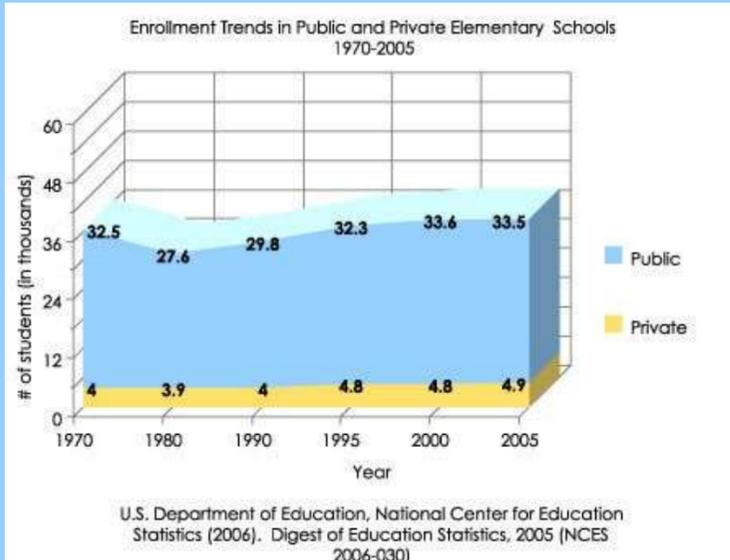
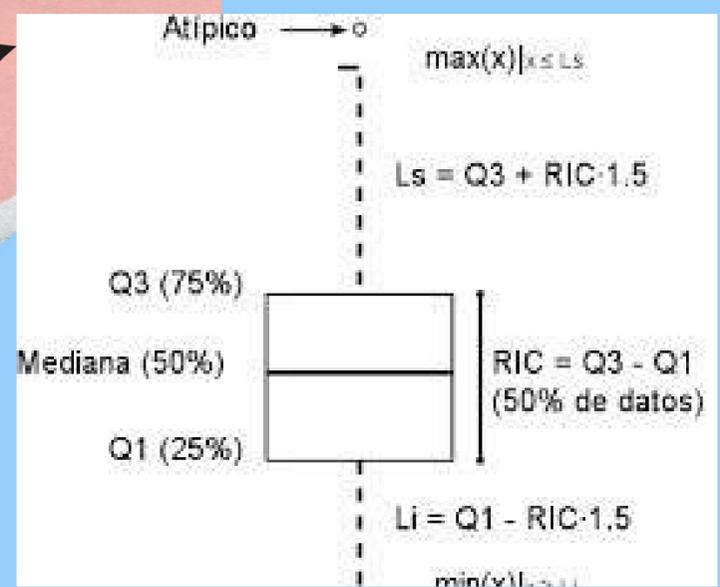
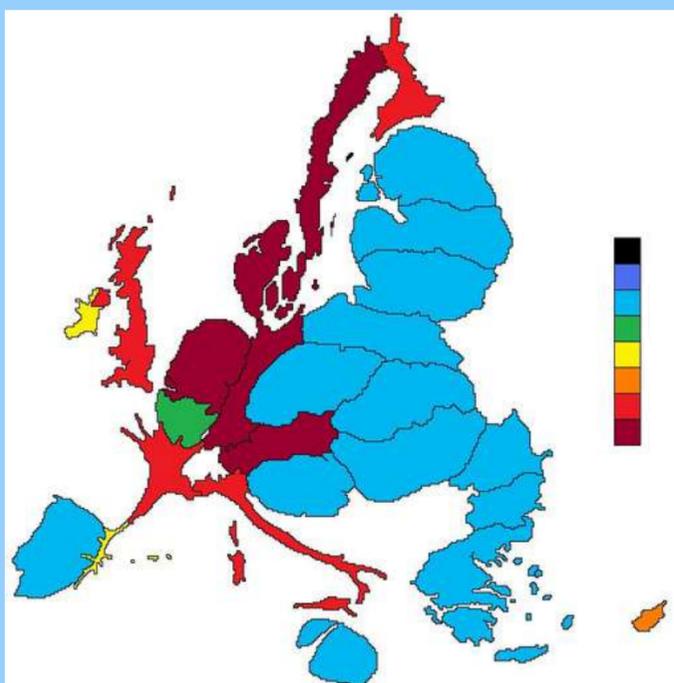
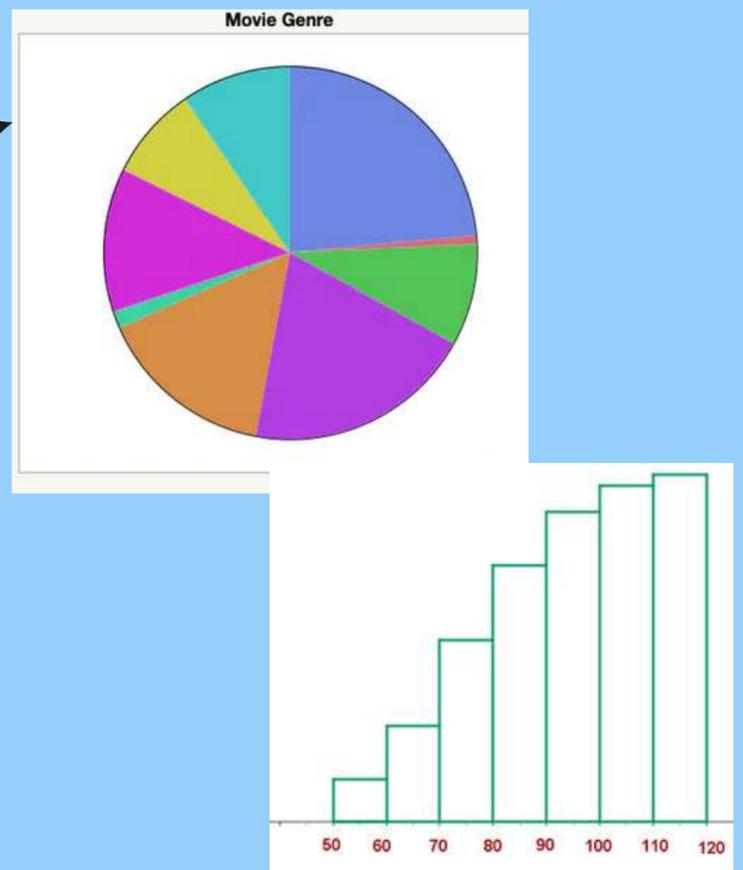


Gráfico de áreas

En este tipo de gráfico se observa, de manera semejante lo que ocurre con los gráficos de líneas, la relación entre variable dependiente e independiente. Inicialmente se hace una línea que une los puntos que marcan los diferentes valores de la variable medida, pero también se incluye todo lo situado por debajo: este tipo de gráfica nos permite ver la acumulación (un punto determinado incluye a los situados por debajo).

Pictograma

Se entiende por pictograma a un gráfico en el que, en vez de representar los datos a partir de elementos abstractos como barras o círculos, se emplean elementos propios del tema que se está investigando. De este modo se hace más visual. Sin embargo, su funcionamiento es semejante al del gráfico de barras, representando frecuencias de la misma manera



Cartograma

Este gráfico resulta de utilidad en el terreno de la epidemiología, indicando las zonas o áreas geográficas en las que aparece con mayor o menor frecuencia un determinado valor de una variable. Las frecuencias o rangos de frecuencias se indican mediante el uso del color (requiriéndose una leyenda para comprenderse) o el tamaño.



¿QUÉ ES UN DIAGRAMA DE CAJA?

un diagrama de caja es un gráfico que representa una variable cuantitativa o cualitativa a través de los cuartiles
En estadística, es una herramienta útil para representar conjuntos de datos tanto discretos como continuos.

Es importante tener en cuenta que las variables cualitativas o que pretenden representar un orden o una categoría siempre tienen que ir ligadas a un índice numérico mayor que 0 para que puedan aparecer en el gráfico y se puedan calcular los estadísticos correspondientes.

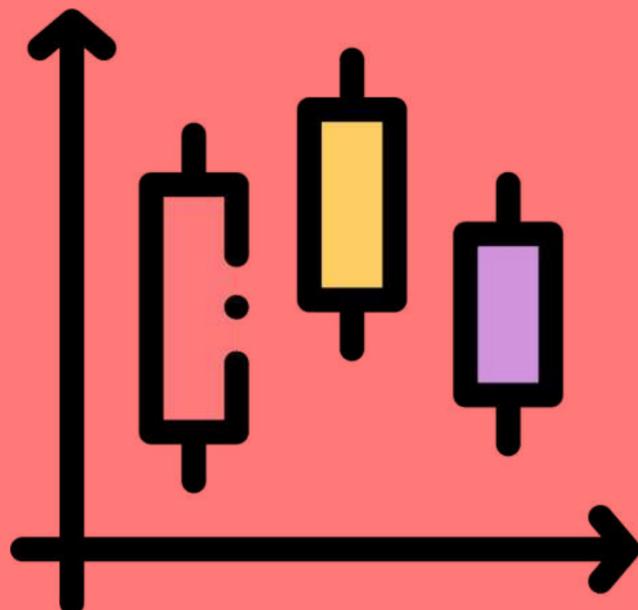
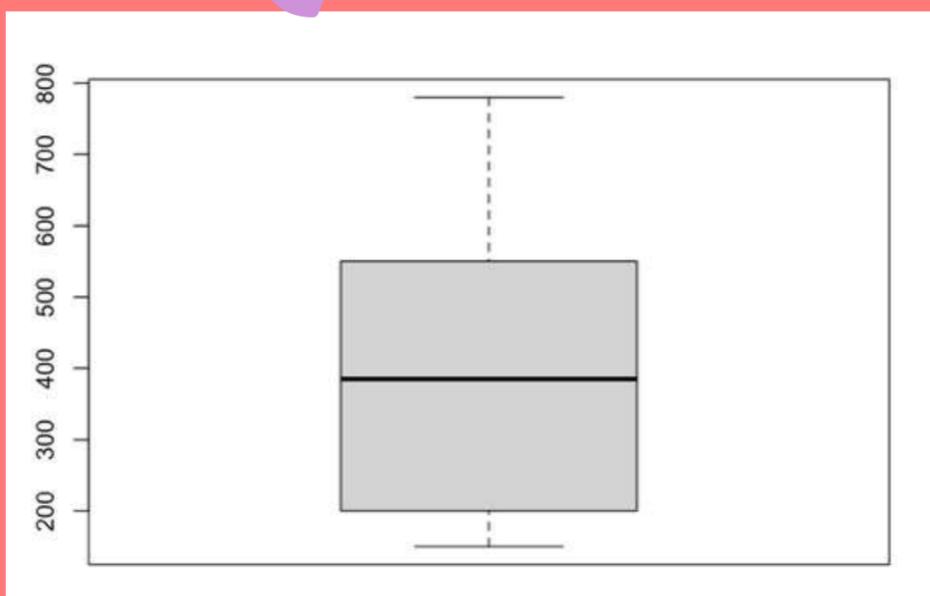


DIAGRAMA DE CAJA



1. La parte inferior de la caja es el primer cuartil (Q1).
2. La barra del medio de la caja es la mediana o segundo cuartil (Q2).
3. La parte superior de la caja es el tercer cuartil (Q3).
4. El rango intercuartílico o RIC sería la altura de la caja, es decir, la diferencia entre Q3 y Q1.

VENTAJAS

- Visión globalizada de la frecuencia absoluta de una variable cuantitativa o cualitativa.
- Conocimiento del mínimo, máximo y cuartiles del conjunto de datos sin tener que calcularlos.
- Reconocimiento de los valores extremos más rápido que con un histograma debido a que los valores extremos están indicados con circunferencias fuera del máximo y mínimo.

DESVENTAJAS

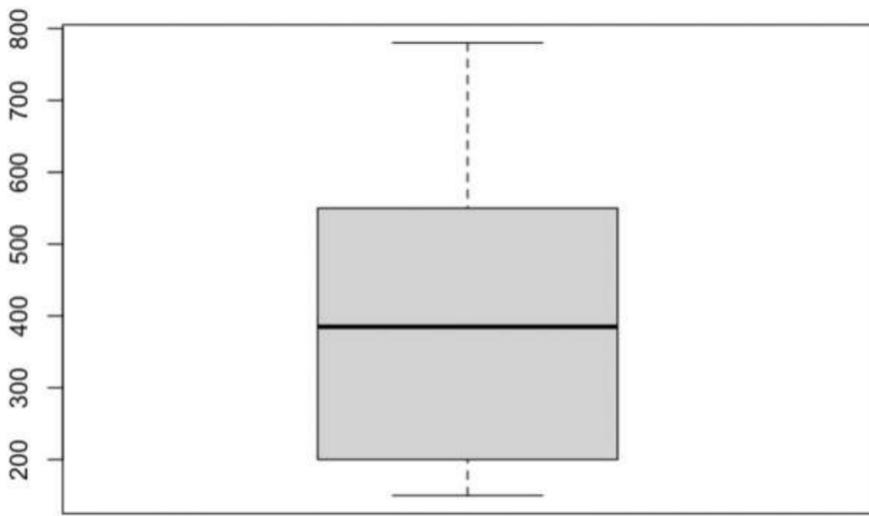
- Si el conjunto de datos es pequeño, hacer un diagrama de caja para expresar esos datos hará más difícil su comprensión.



Suponemos que queremos representar el número de ciclistas que pasan por delante de nuestra casa a lo largo de un año. Primero, contamos los ciclistas y recogemos la información en una tabla.

Número de ciclistas	
Mes	Ciclistas
Enero	200
Febrero	150
Marzo	200
Abril	300
Mayo	370
Junio	400
Julio	600
Agosto	700
Septiembre	780
Octubre	500
Noviembre	400
Diciembre	200

Diagrama de caja del número de ciclistas



Gracias a representar el número de ciclistas a través del diagrama de caja podemos saber rápidamente donde está la mediana, los otros cuartiles, el máximo y el mínimo. En este caso no tenemos datos atípicos dado que más allá del máximo y del mínimo no hay nada.

Es muy común la aplicación de este tipo de gráfico dada su simplicidad y utilidad en muchos ámbitos más allá de las finanzas y la economía.

DIAGRAMA DE CAJA E HISTOGRAMA

La primera característica que comparten ambos gráficos es el mínimo y el máximo del conjunto de datos. Las barras del extremo después de los guiones indican el mínimo (barra inferior) y el máximo (barra superior). Los datos atípicos estarían representados fuera del intervalo del máximo y el mínimo. Es decir, más allá de las barras después de los guiones.

También podemos encontrar diagramas de caja horizontales.

A simple vista se puede ver que la mayoría de datos se centra en las observaciones cerca del mínimo en ambos gráficos. Esto está indicado por el diagrama de caja en que la caja está más cerca del mínimo que del máximo. Si miramos al histograma, veremos que las observaciones con más frecuencia son las que se sitúan cerca del mínimo y lejos del máximo.

