



Estadística descriptiva

VARIABLES, MEDIDAS DE RESUMEN, REPRESENTACIONES GRÁFICAS



Variable

Es en principio un concepto que determina una cualidad de un elemento, es un atributo que puede variar de una o más maneras y que sintetiza conceptualmente lo que se quiere conocer acerca del objeto de estudio

Temperatura corporal, glicemia, estadio de una enfermedad, peso, IMC, número de un dado, cara de una moneda...

Tipos de variables

Cualitativas:

- Refiere a atributos descritos por cualidades, no toman un valor numérico medido, solo separan “categorías”

Cuantitativas:

- El atributo medido toma un valor numérico

Variables cuantitativas

Pueden dividirse a su vez en:

- Discretas:
 - Toman valores solo en el grupo de los números enteros, generalmente ocurren cuando se trata de conteo de elementos.
- Continuas:
 - Aquellas que podrían tener infinitos valores posibles (números reales) y parten de las magnitudes básicas (longitud, masa, tiempo, intensidad de corriente, temperatura, cantidad de sustancia, intensidad luminosa)

Escalas de medida

- Escala nominal:
 - Se limita a nombrar los atributos que puede tomar una variable
 - No se puede operar sobre estos valores porque no tienen magnitud, solo podemos contarlos
 - Se usa para variables cualitativas
- Escala ordinal:
 - Al igual que la nominal, nombra los atributos pero agregando un orden lógico entre ellos. Este orden no es numérico.
 - Además de contarlos, podemos establecer relaciones comparativas
 - Se usa para las variables cualitativas que permiten establecer este orden

Escalas de medida

- Escala de intervalo:
 - Los valores se posicionan sobre la recta numérica (enteros o reales)
 - No tiene cero absoluto
 - Agrega la posibilidad de sumar y restar valores (cálculo de media)
 - Se usa para variables cuantitativas
- Escala de razón:
 - Agrega a la de intervalo un cero absoluto
 - Se pueden realizar todas las operaciones



Clasificar

- Temperatura corporal (C, F, K y R)
- Glicemia
- Estadio de una enfermedad
- Peso
- Estado civil
- IMC
- Número de un dado
- Puesto en un concurso
- Cara de una moneda

Representaciones gráficas

Variables Cualitativas – Nominales:

- Gráfico de barras, diagrama circular

Variables Cualitativas – Ordinales:

- Diagrama de barras

Variables Cuantitativas Discretas:

- Diagrama de barras

Variables Cuantitativas Continuas:

- Histograma, polígono de densidad u ojiva de frecuencia acumulada
- Box plot (Cuantitativa en función de cualitativa)
- Diagrama de dispersión (Cuantitativa en función de cuantitativa)

Gráficos de barras simples y agrupadas

Los gráficos más frecuentes para representar **variables cualitativas** son los diagramas de barras.

Se representa en el eje de ordenadas las modalidades y en abscisas las frecuencias absolutas o las frecuencias relativas.

Si se intentan comparar varias poblaciones entre sí, usando el diagrama, existen otras modalidades como las **barras agrupadas**

Gráficos de barras simples y agrupadas

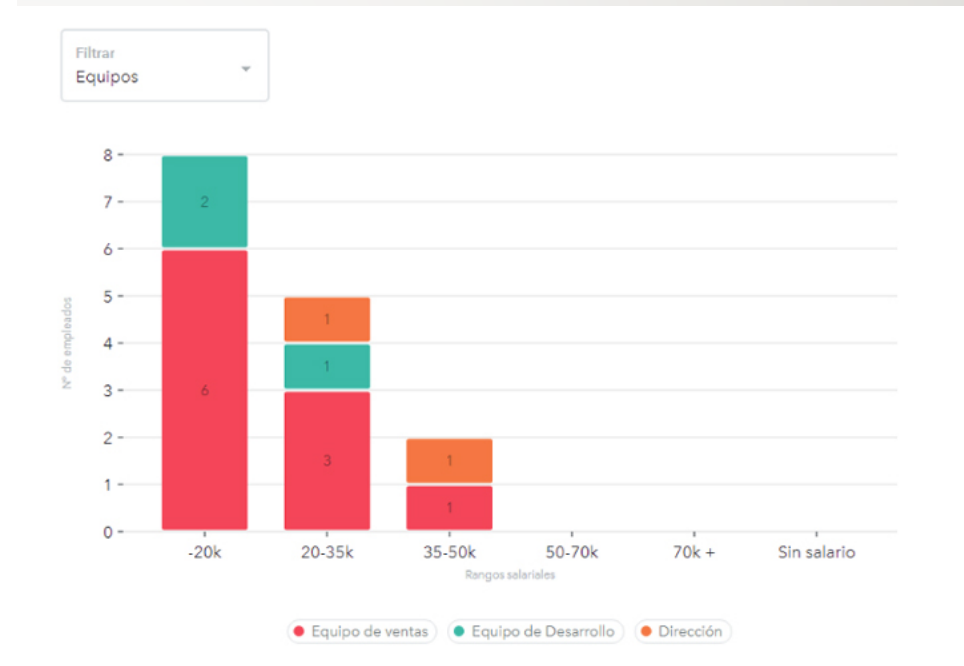
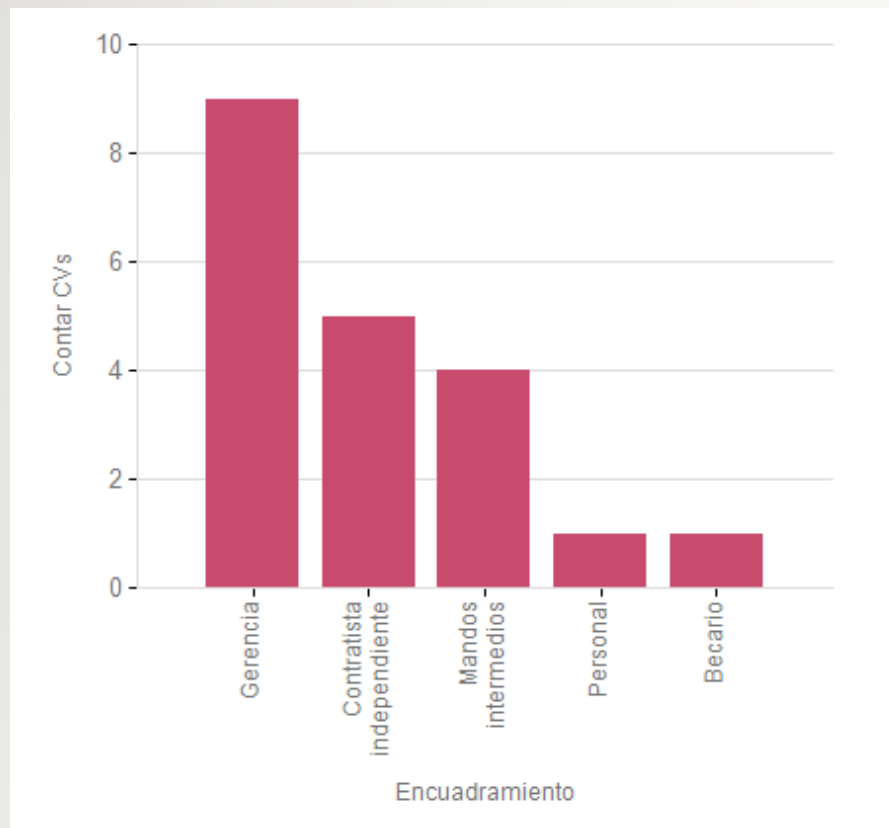


Gráfico de sectores

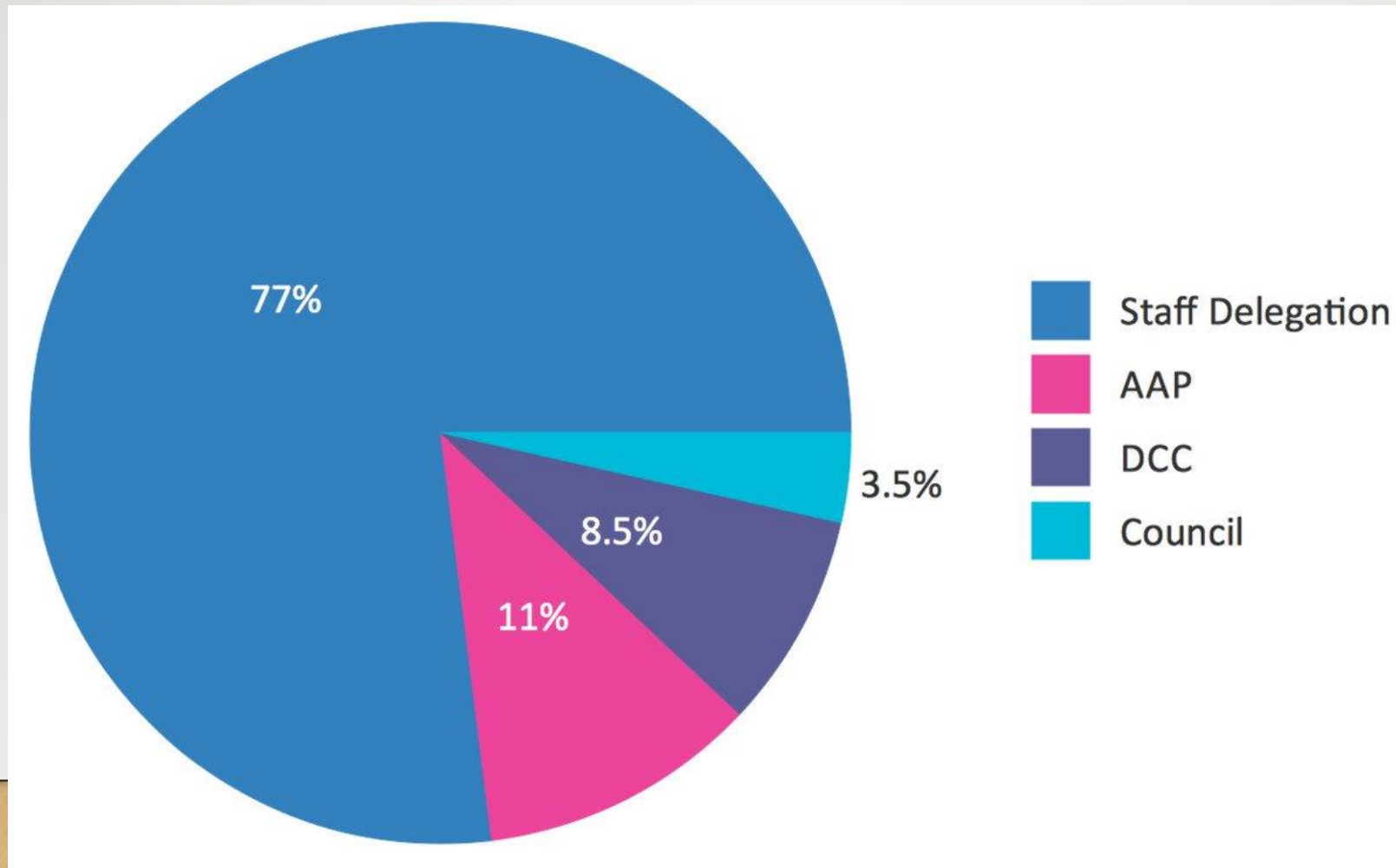
Es el otro tipo de gráfico para representar **variables cualitativas**.

Se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa.

El arco de cada porción se calcula con una regla de tres entre 360 grados, tamaño de la muestra y frecuencia de la clase.

Se pueden comparar dos poblaciones con círculos concéntricos.

Gráfico de sectores



Gráficos para variables cuantitativas - Generalidades

Para las variables cuantitativas, consideraremos dos tipos de gráficos, en función de que para realizarlos se usen las frecuencias (absolutas o relativas) o las frecuencias acumuladas.

Se le llaman **diagramas diferenciales** a los primeros (F y f) y **diagramas integrales** a los últimos (FA)

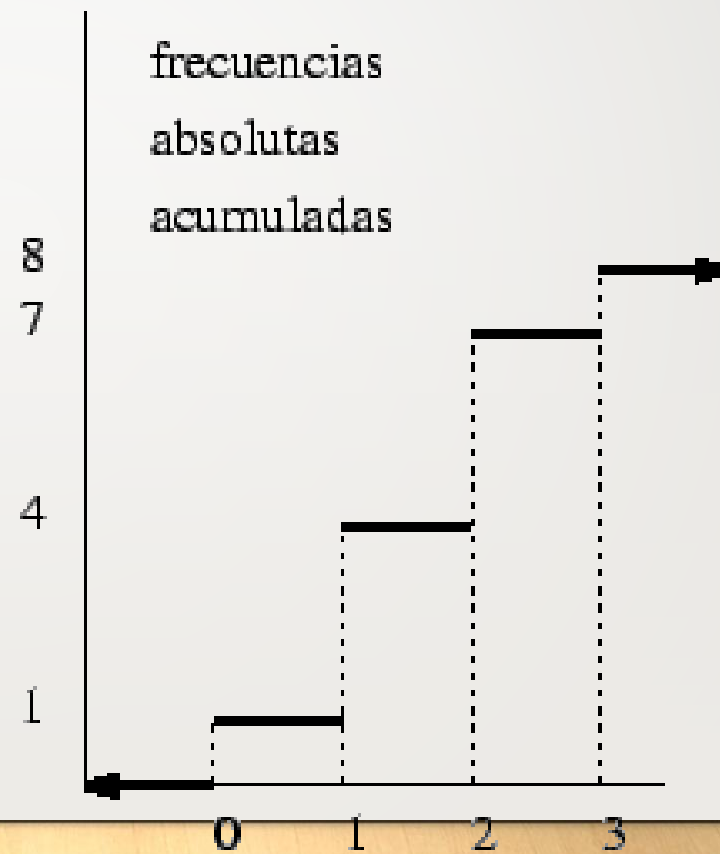
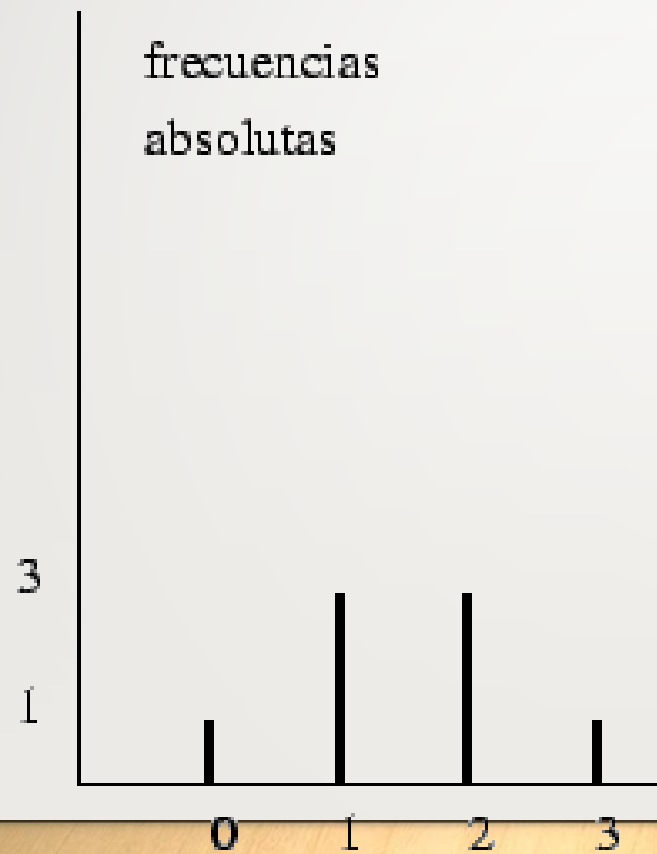
Variables discretas

Usamos el *diagrama de barras* cuando pretendemos hacer una **gráfica diferencial**.

Las barras deben ser estrechas para representar que los valores que toma la variable son discretos.

El **diagrama integral** o *acumulado* tiene, por la naturaleza de la variable, forma de escalera.

Gráficos de barras y escaleras



Variables continuas

Cuando las variables son **continuas**, utilizamos como diagramas diferenciales los *histogramas* y los *polígonos de frecuencias*.

Histograma

Un histograma se construye a partir de la tabla estadística, representando sobre cada intervalo, un rectángulo que tiene a este segmento como base. El criterio para calcular la altura de cada rectángulo es el de *mantener la proporcionalidad entre las frecuencias absolutas (o relativas) de cada intervalo y el área de los mismos*

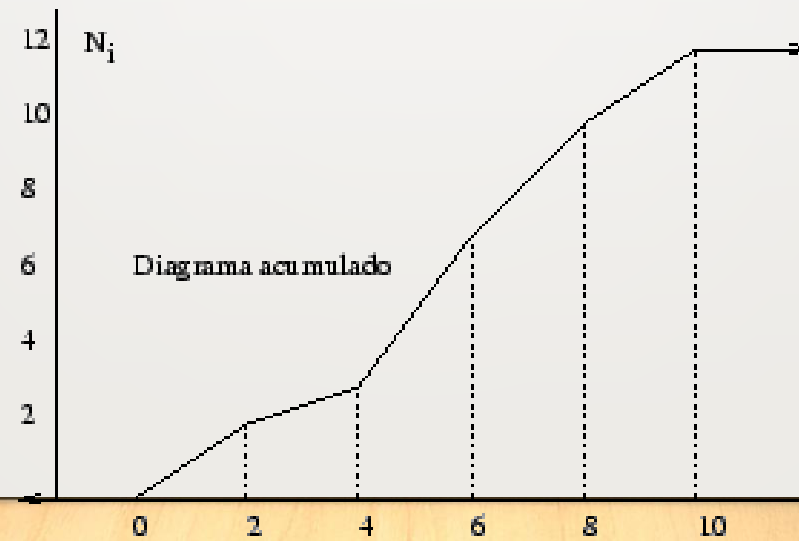
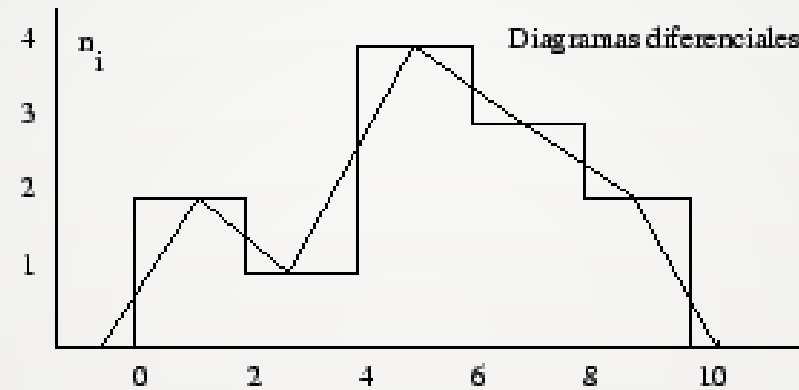
Polígono de frecuencia

El polígono de frecuencias se construye fácilmente si tenemos representado previamente el histograma, ya que consiste en unir mediante líneas rectas los puntos del histograma que corresponden a las marcas de clase. Para representar el polígono de frecuencias en el primer y último intervalo, suponemos que adyacentes a ellos existen otros intervalos de la misma amplitud y frecuencia nula, y se unen por una línea recta los puntos del histograma que corresponden a sus marcas de clase.

Polígono de frecuencias acumuladas (Ojiva)

El diagrama integral para una variable **continua** se denomina también *polígono de frecuencias acumulado*, y se obtiene como la poligonal definida en abscisas a partir de los extremos de los intervalos en los que hemos organizado la tabla de la variable, y en ordenadas por alturas que son proporcionales a las frecuencias acumuladas.

Diagramas de variables continuas



Box and whisker plot

AKA diagrama de caja y bigote o *boxplot*.

Son una forma de representar de manera gráfica una serie de datos utilizando los cuartiles (caja) y otra serie de datos (bigotes).

Construcción de un *box plot*

Los límites de la caja son el primer y tercer cuartil, la banda central es la mediana y se puede representar la media con un punto dentro de la caja.

Los bigotes pueden representar una serie de posibles valores alternativos, como por ejemplo:

- Mínimo y máximo (dando idea del rango de la distribución)
- Media \pm Desviación estándar
- Percentil 9 y 91
- Percentil 1 y 99
- Cuartil 1 y 3 \pm 1,5 recorridos intercuartílicos

Todos los datos que no entren en los bigotes se tienen que representar con un punto o estrella.

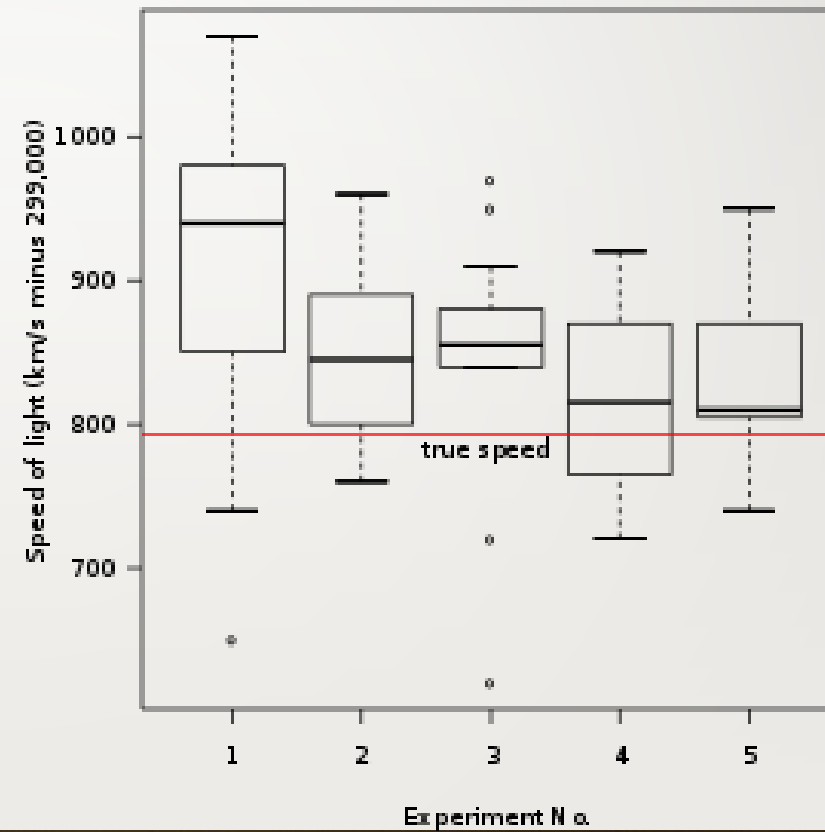
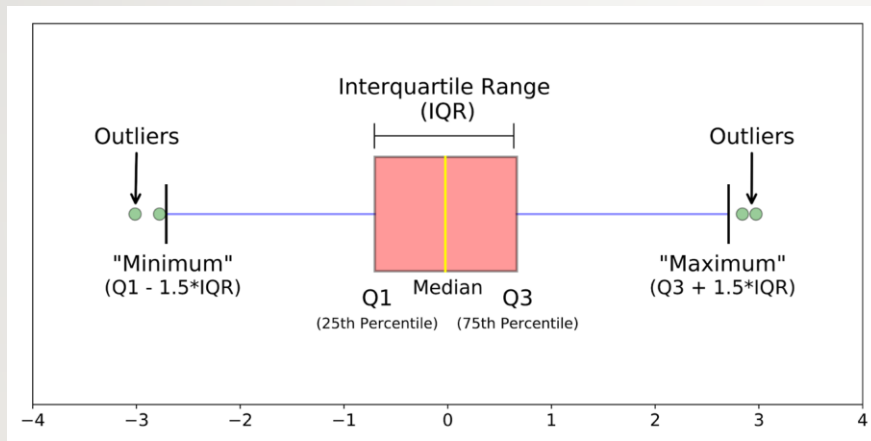
Usos

Los box plot permiten mostrar rápidamente un gran grupo de medidas de resumen (Cuartiles, mediana, media, desviación y varianza o mínimo y máximo) y muchas otras que se desprenden del análisis rápido del gráfico (CV, Q, rango, asimetría, curtosis).

Como ocupan poco espacio permiten comparar muchos grupos juntos.

La desventaja que presentan es que omiten valores específicos de la distribución, mostrando solo medidas de resumen.

Ejemplos



Medidas de resumen

Permite caracterizar poblaciones o muestras en base a valores que podemos utilizar para compararlas sin recurrir a todo el conjunto de datos

Se clasifican en:

- Medidas de Tendencia Central
- Medidas de Dispersión
- Medidas de Posición



Medidas de Tendencia Central



Media

Es la media aritmética o promedio del conjunto de datos

Se calcula como la sumatoria de los valores dividido el total de elementos

En datos agrupados se utiliza la sumatoria de las marcas de clase por su frecuencia absoluta

Tiene la misma unidad que la observación

Practicar uso de calculadora

Es útil conocer como utilizar el modo estadístico para simplificar estos cálculos

Se ingresa al modo estadístico presionando **MODE, 2 (SD)**

En S-VAR se encuentran las diferentes funciones estadísticas como la media y el desvío estándar



Tecla para seleccionar el modo estadístico

En M+ se cargan cada uno de los datos

Borrar la memoria de datos

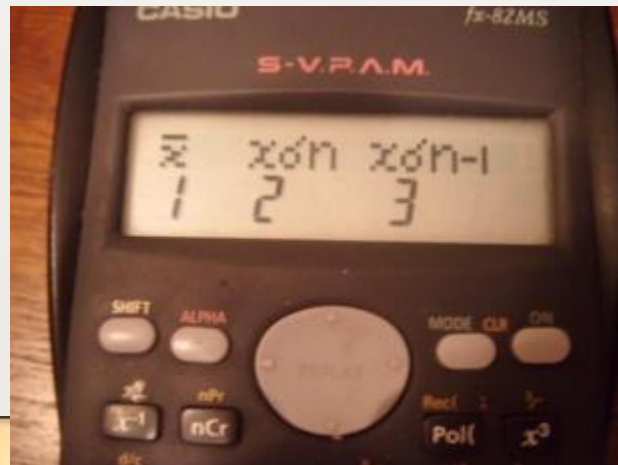
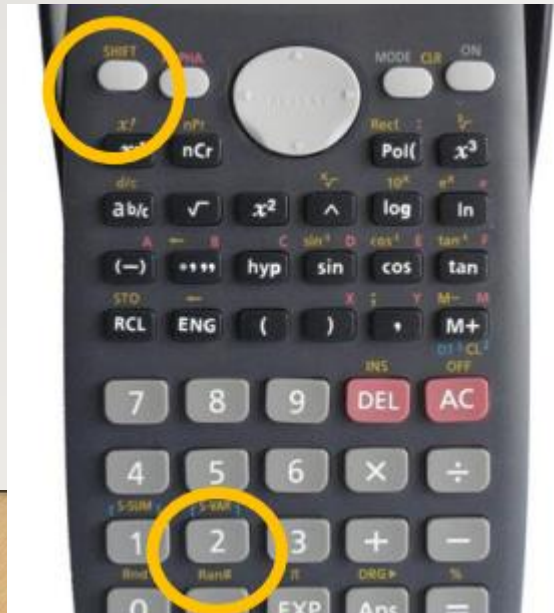
Al apagar la calculadora no se borran los últimos datos ingresados, para asegurarnos que comenzamos sin datos previos se presiona

SHIFT, MODE (CLR), 1



Cálculo de media de una serie de datos simple

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
5	7	2	3	1	0	5	4	3	2	2	0	4	9	2



Cálculo de media de una serie de datos agrupados

X_i	$f a (X_i)$
0	2
1	1
2	4
3	2
4	2
5	2
7	1
9	1
Totales	$n = 15$



Mediana

Valor central de una serie de datos ordenados

Separa al grupo en dos mitades

Si la distribución es simétrica permite estimar la media poblacional, aunque no es precisa

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
5	7	2	3	1	0	5	4	3	2	2	0	4	9	2

Moda

Valor mas frecuente del conjunto de datos

X_i	$f a (X_i)$
0	2
1	1
2	4
3	2
4	2
5	2
7	1
9	1
Totales	n = 15

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
5	7	2	3	1	0	5	4	3	2	2	0	4	9	2



Medidas de Dispersión

Rango

Diferencia entre el valor máximo y mínimo

Solo se puede calcular en escalas de razón o de intervalo

X_i	$f a (X_i)$
0	2
1	1
2	4
3	2
4	2
5	2
7	1
9	1
Totales	n = 15

Varianza y desvío estándar

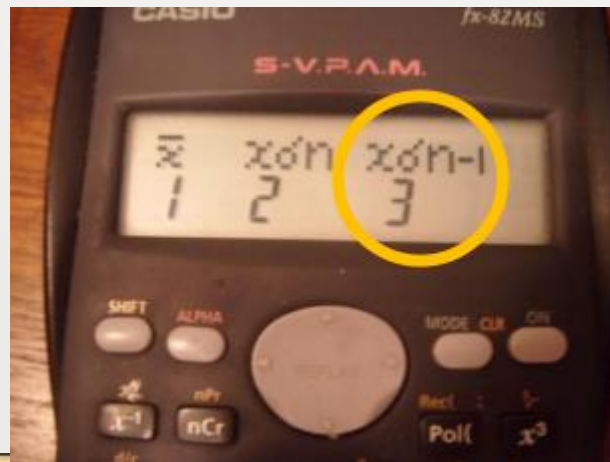
Es útil como medida del nivel de alejamiento de la media en variables con escala de intervalo o de razón

La varianza se evalúa como la sumatoria del cuadrado de las diferencias entre cada valor y la media, dividido el número de elementos

A la raíz cuadrada de la varianza le llamamos desvío estándar, y es útil ya que tiene la misma unidad que el parámetro

Cálculo de desvío y varianza

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
5	7	2	3	1	0	5	4	3	2	2	0	4	9	2



Coeficiente de variación (CV)

Solo se puede aplicar a variables medidas en escala de razón

$$CV = \frac{S}{\bar{X}} \cdot 100\%$$


Donde:

CV = coeficiente de variación

S = desviación típica

\bar{X} = media aritmética

X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
5	7	2	3	1	0	5	4	3	2	2	0	4	9	2



(Semi)Rango intercuartílico

El rango IC es la diferencia entre el cuartil 3 y el cuartil 1

El semi rango IC es la mitad del rango

Se utiliza para distribuciones asimétricas



Medidas de Posición



Cuantiles

Son valores que separan al grupo de datos en subconjuntos de distintos tamaños

Se pueden pensar como un concepto análogo al de la mediana, solo que en lugar de dividir al conjunto en dos subconjuntos iguales, los dividen en subconjuntos de distintos tamaños



Percentiles (P)

Parte de la noción de porcentaje

Son valores que dejan por debajo un determinado porcentaje del grupo

Ej: El percentil 50 es el valor que tiene por debajo al 50% de los elementos

Deciles (D)

Parte de la noción de décimos (10 partes iguales)

Son valores que dejan por debajo un determinada cantidad de décimos del grupo

Se corresponden a los percentiles múltiplos de 10

Ej: El decil 5 es el valor que tiene por debajo al 50% de los elementos

Cuartiles (Q)

Parte de la noción de cuartos (4 partes iguales)

Son valores que dejan por debajo un determinada cantidad de cuartos del grupo

Se corresponden a los percentiles 25, 50, 75 y 100

Ej: El cuartil 2 es el valor que tiene por debajo al 50% de los elementos

Equivalencias de cuantiles

Percentiles	Deciles	Cuartiles	Mediana
P10	D1		
P20	D2		
P25		Q1	
P30	D3		
P40	D4		
P50	D5	Q2	Mn
P60	D6		
P70	D7		
P75		Q3	
P80	D8		
P90	D9		
P100	D10	Q4	