

CONCEPTOS CLAVE

CONCEPTO	TEORÍA
Tipos de variables. Cualitativas vs Cuantitativas	Módulo 1 (Sesión 1. 1)
Representación gráfica. Escoger el gráfico adecuado	Módulo 1 (Sesión 1. 2, 3, 4)
Correlación lineal	Ficha 006 - Correlación lineal
Medidas de centro y dispersión.	Módulo 1 (Sesiones 2 y 3)
Cardinal de un suceso	Módulo 3 (Sesión 1. 3)
Combinatoria	Módulo 3 (Sesión 2)
Cálculo de probabilidad	Módulo 3 (Sesión 3)
Esperanza, varianza y desviación	Módulo 4 (Sesión 2)
Distribuciones discretas. Distribución binomial	Módulo 4 (Sesión 3. 2)
Distribuciones discretas. Distribución de Poisson	Módulo 4 (Sesión 3. 4)
Teorema central del límite	Módulo 5
Distribución normal	Módulo 4 (Sesión 5)
Intervalos de confianza para la media aritmética	Módulo 6 (Sesión 1)
Intervalos de confianza para una proporción	Módulo 6 (Sesión 2)
Obtener el tamaño de la muestra	Módulo 6 (Sesión 1. 2.4 y Sesión 2. 2)
Contrastes de hipótesis	Módulo 7 (Sesión 2)
Contrastes sobre la diferencia de medias	Módulo 8 (Sesión 1)
Contrastes sobre la diferencia de proporciones	Módulo 8 (Sesión 2)
Contraste de la varianza	Módulo 9 (Sesión 1)
Comparación de varianzas	Módulo 9 (Sesión 2)

RESUMEN

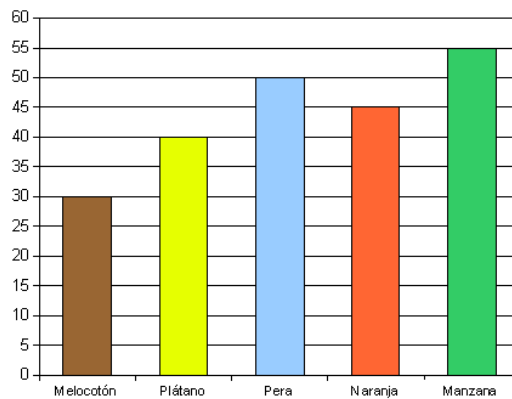
Tipos de variables. Cualitativas vs Cuantitativas

- **Variables cualitativas (o categóricas)** → Expresadas en términos de categorías. Su valor no tiene un sentido numérico o de proporción, sino únicamente clasificatorio. (Ejemplo: Una variable que toma los valores "1" si el año es bisiesto o bien "0" si no lo es)
- **Variables cuantitativas** → Expresadas numéricamente
 - **Cuantitativas discretas** → Toman valores enteros. (Ejemplo: El número de trabajadores de una empresa)
 - **Cuantitativas continuas** → Toman cualquier valor. Fruto de la experimentación o de la observación física. (Ejemplo: El peso de una persona)

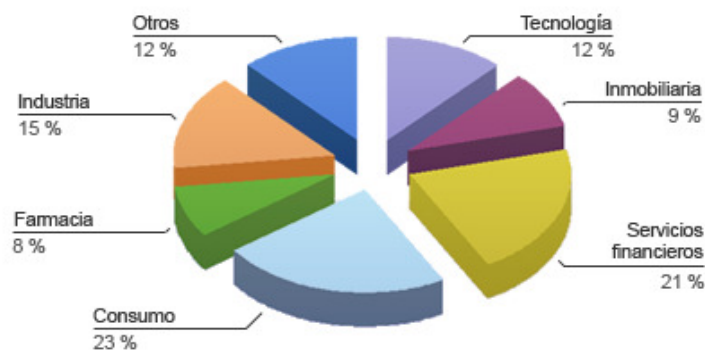
Representación gráfica. Escoger el gráfico adecuado

- Representación de **variables cualitativas o cuantitativas discretas con un número reducido de valores distintos:**

- **Diagrama de barras** → Se representa una barra por categoría, y la longitud de la barra consiste en el número de observaciones que correspondan a dicha categoría (frecuencia absoluta)

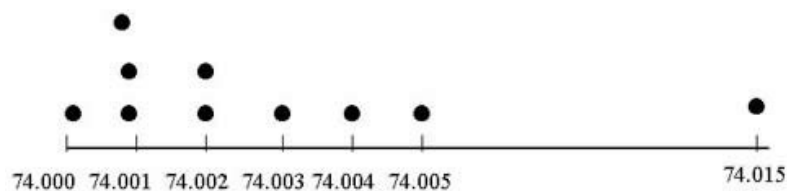


- **Diagrama de sectores** → Cada categoría se representa como un sector dentro del espacio circular. El tamaño de cada sector se determina a partir de la proporción de las observaciones cada categoría sobre el total (frecuencia relativa)

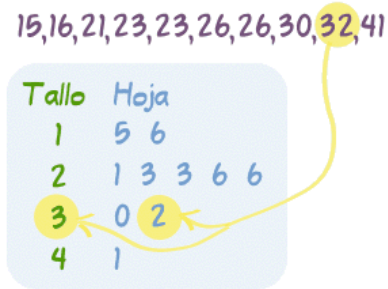


- Representación de **variables cuantitativas con un número moderado de valores distintos:**

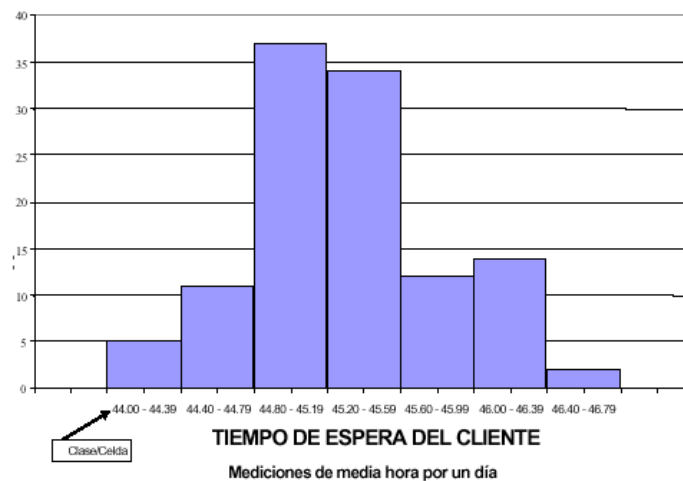
- **Diagrama de puntos** → Se representan mediante puntos las observaciones en un eje horizontal donde aparezcan representados los valores.



- **Diagrama de tallo y hojas** → Se parten los valores numéricos de las observaciones en su dígito de más peso (tallo), y representando el resto de dígitos a su lado (hojas)



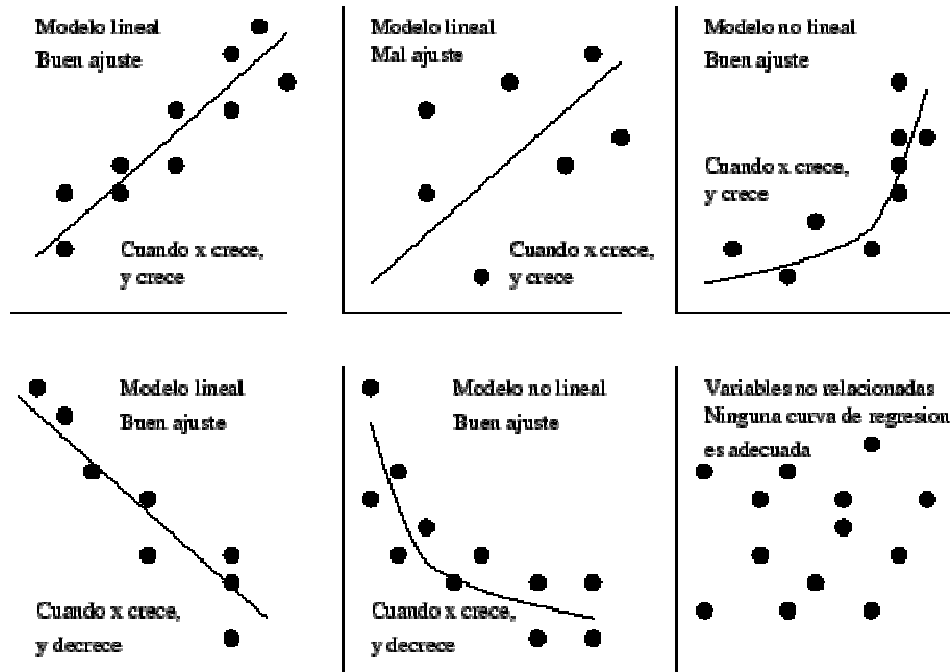
- Representación de **variables cuantitativas con un número elevado de valores distintos o de individuos**:
 - **Histograma** → Se definen intervalos (clases) y dentro de cada clase se dibuja una barra con una longitud igual a la frecuencia de su clase (puede ser de frecuencia absoluta o bien de frecuencia relativa).



Correlación lineal

Correlación lineal → Mide el grado de relación lineal entre dos variables. El valor de la correlación se mueve entre -1 y 1. Cuanto más cerca a -1 mayor es la relación lineal inversa entre ambas variables y cuando más cerca de 1 mayor es la relación lineal positiva entre ambas.

- Análisis gráfico → **Diagrama de dispersión.**



El gráfico de dispersión muestra una nube de puntos expresando la relación entre una variable (eje x) y otra (eje y).

Si los puntos se agrupan formando una forma parecida a una recta → hay una correlación lineal elevada (cercana a 1 o -1).

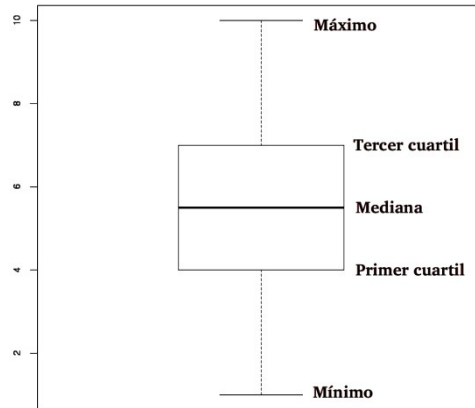
- Si la recta es creciente → correlación cercano a 1
- Si la recta es decreciente → correlación cercano a -1

Si los puntos se distribuyen sin ningún tipo de relación aparente → correlación lineal baja (cercana a 0)

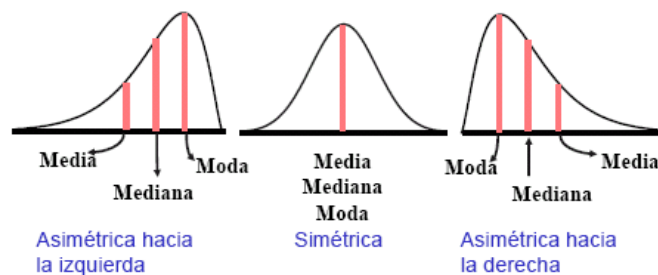
Medidas de centro y dispersión

- **Medidas de centro:** Aportan información del valor central de la variable
 - **Moda** → Es el valor más repetido
 - **Mediana** → Una vez ordenadas las observaciones, consiste en la observación situada en el punto medio
 - **Media** → Es el valor promedio de todas las observaciones. $\mu = \frac{1}{N} \cdot \sum_{i=1}^N x_i$
- **Medidas de dispersión:** Aportan información sobre la dispersión o separación de las observaciones
 - **Cuartiles:**
 - **Primer cuartil (Q1)** → Valor hasta el que se acumula el 25% de las observaciones
 - **Segundo cuartil (Q2) también llamado mediana** → Valor hasta el que se acumula el 50% de las observaciones
 - **Tercer cuartil (Q3)** → Valor hasta el que se acumulan el 75% de las observaciones.
 - **Rango intercuartílico** → $Q3 - Q1$. A mayor rango intercuartílico, mayor es la dispersión.

- **Representación gráfica** → Diagrama de caja. Se representan los cuartiles, así como los valores máximos y mínimos. Nota: A menudo en lugar de marcarse el valor máximo y mínimo se representan el percentil 90% y 10% respectivamente, y considerándose observaciones atípicas las que salen fuera de este límite.



- **Valor máximo** → El máximo valor observado
- **Valor mínimo** → El mínimo valor observado
- **Varianza** → Es la suma de las diferencias entre cada observación y la media, al cuadrado, y dividido entre el número de observaciones. A mayor varianza, mayor es la dispersión de los datos. $\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2$
- **Desviación típica** → Raíz cuadrada de la varianza. $\sigma = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}$
- **Asimetrías. Relación entre media y mediana:**
 - Si media > mediana → asimetría positiva o hacia la derecha. Es decir, la cola de la distribución está a la derecha
 - Si media < mediana → asimetría negativa o hacia la izquierda. Es decir, la cola de la distribución está a la izquierda.



- **Relaciones entre media y desviación típica:**
 - **Datos estandarizados** → Estandarizar un valor comporta trasladarlo a un nuevo dominio de referencia, donde la variable pasa a tomar una media de 0 y una desviación típica de 1. Utilidad → Poder determinar la posición relativa de una observación sobre un conjunto.
Para estandarizar → Obtenemos la variable estandarizada "z" restando a la variable original "x" la media y dividiendo entre la desviación típica: $z = \frac{x - \text{media}}{\text{desviación típica}}$