

## TEMA 3

### ANÁLISIS DESCRIPTIVO DE DATOS CON DOS VARIABLES

- 
- 3.1. Algunas representaciones gráficas y tabulares para relacionar dos variables
  - 3.2. Análisis de regresión lineal simple y correlación
  - 3.2. Uso de herramientas tecnológicas para relacionar dos variables
- 

- 3.1. Algunas representaciones gráficas y tabulares para relacionar dos variables:
    - a) Dos variables cualitativas: tabla de contingencia, diagrama;
    - b) Una cuantitativa y una cualitativa: diagrama de cajas por categorías;
    - c) Dos variables cuantitativas: diagrama de dispersión.
  - 3.2. Análisis de regresión lineal simple y correlación.
    - ecuación de la recta de estimación por el método de mínimos cuadrados,
    - coeficientes de correlación y determinación,
    - error estándar de estimación.
  - 3.3. Uso de herramientas tecnológicas para relacionar dos variables
    - Uso de la calculadora para la determinar
      - la recta de regresión por el método de mínimos cuadrados y
      - el coeficiente de correlación.
    - Uso de software para realizar el análisis de regresión y correlación con el computador.
- 

#### 3.1 a) Dos variables cualitativas: Tabla de contingencia.

Es una Tabla que clasifica las observaciones de acuerdo con dos características identificables. Se utiliza para estudiar la relación entre dos variables cuando una o ambas son de escala nominal u ordinal.

| Tabla de contingencia sobre la relación entre ganancia y concesionaria |              |       |           |          |       |
|--|--------------|-------|-----------|----------|-------|
| Abajo/arriba<br>Ganancia   | mediana Kane | Olean | Sheffield | Tionesta | Total |
| Por abajo  | 25           | 20    | 19        | 26       | 90    |
| Por arriba   | 27           | 20    | 26        | 17       | 90    |
| Total  | 52           | 40    | 45        | 43       | 180   |

#### 3.1 b) Una cuantitativa y una cualitativa: diagrama de cajas por categorías.

Un diagrama de caja es una representación gráfica, basada en cuartiles, que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja, sólo necesita cinco estadísticos:

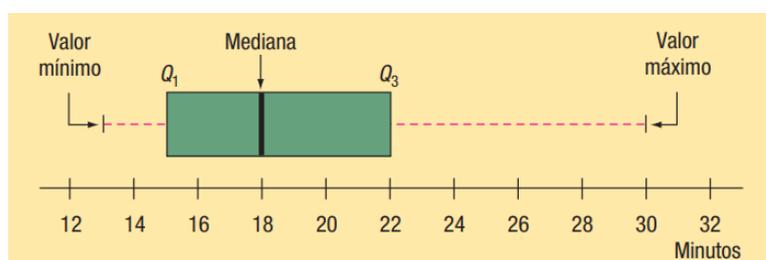
el valor mínimo,

Q1 (primer cuartil),

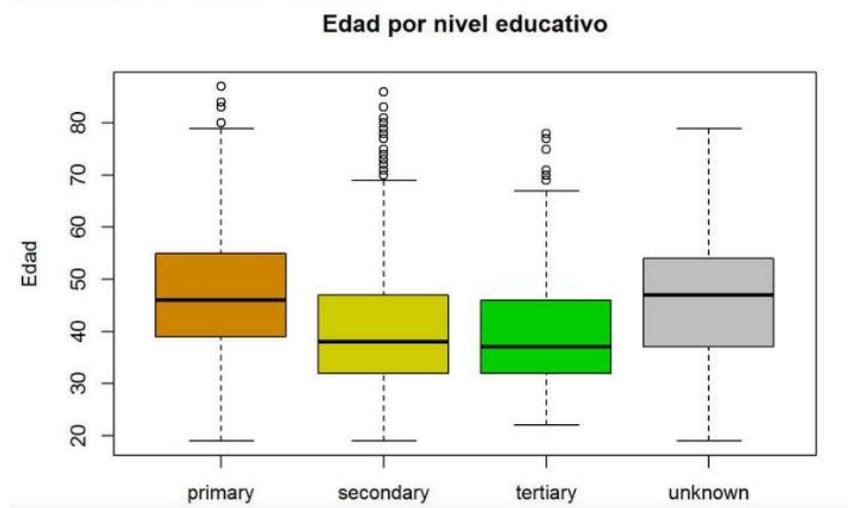
la mediana,

Q3 (tercer cuartil) y

el valor máximo.

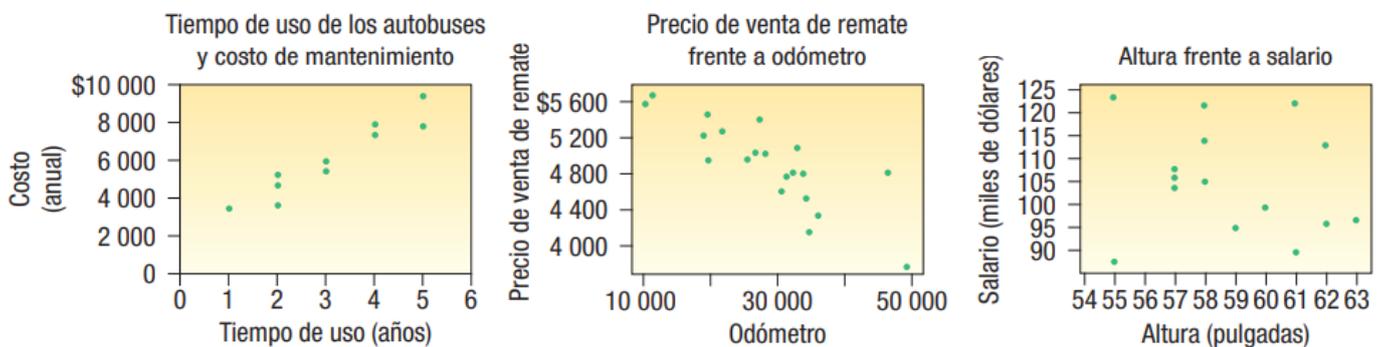


En el diagrama de cajas por categorías, se busca comparar una serie de grupos (variable cualitativa) en relación a un valor medible (variable cuantitativa) con el fin de clasificarlos o tomar decisiones sobre estos.



3.1 c) Dos variables cuantitativas: diagrama de dispersión.

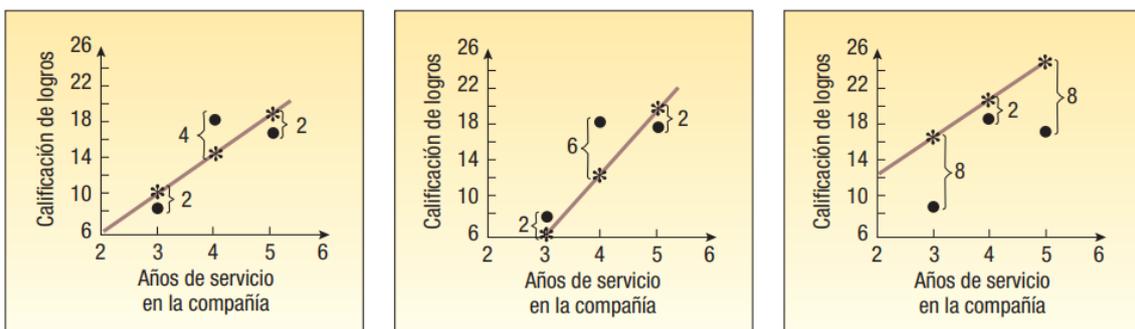
Es una técnica gráfica útil para mostrar la relación entre dos variables cuantitativas. Se escala una de las variables sobre el eje horizontal (eje X) de una gráfica y la otra variable a lo largo del eje vertical (eje Y).



3.2. Análisis de regresión lineal simple y correlación.

Ecuación de la recta de estimación por el método de mínimos cuadrados.

El principio de los mínimos cuadrados determina una ecuación de regresión al minimizar la suma de los cuadrados de las distancias verticales entre los valores reales de Y y los valores pronosticados de Y.



Así tenemos que la forma general de la ecuación de regresión lineal viene dada por:

$$\hat{Y} = a + bX$$

donde:

$\hat{Y}$  = que se lee Y prima, es el valor de la estimación de la variable Y para un valor X seleccionado.

a: es la intersección Y. Es el valor estimado de Y cuando X = 0.

b: es la pendiente de la recta.

X: es cualquier valor de la variable independiente que se seleccione.

El propósito de un análisis de regresión es calcular los valores de a y b para desarrollar una ecuación lineal que se ajuste mejor a los datos.

Las fórmulas de a y b son:

**PENDIENTE DE LA RECTA DE REGRESIÓN**

$$b = r \frac{s_y}{s_x}$$

**(13-4)**

donde:

r es el coeficiente de correlación.

$s_y$  es la desviación estándar de Y (la variable dependiente).

$s_x$  es la desviación estándar de X (la variable independiente).

**INTERSECCIÓN CON EL EJE Y**

$$a = \bar{Y} - b\bar{X}$$

**(13-5)**

Coeficientes de correlación y determinación. Viene dado por la siguiente fórmula:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y}$$

Error estándar de estimación. Se calcula mediante:

$$s = \sqrt{\frac{\sum f(M - \bar{X})^2}{n - 1}}$$

### TRABAJO DE INVESTIGACIÓN

3.3. Uso de herramientas tecnológicas para relacionar dos variables

- Uso de la calculadora para la determinar
  - la recta de regresión por el método de mínimos cuadrados y
  - el coeficiente de correlación.
- Uso de software para realizar el análisis de regresión y correlación con el computador.