

Apuntes Unidad 1

Gráficos de dispersión





**GRÁFICOS DE DISPERSIÓN**

Para analizar la relación entre dos variables se usa otro tipo de gráfico llamado **gráfico de dispersión**. Esta representación gráfica también se conoce como **nube de puntos.**

La diferencia de este gráfico con los anteriores es que en este se representan dos variables cuantitativas distintas, una en cada eje coordenado. En este gráfico, cada punto representa un par ordenado $(x,y)$, donde $x$ e $y$ son variables distintas que representan a un mismo individuo.

Figura 1: Gráfico de dispersión entre la esperanza de vida y el PIB per cápita
invertido en la salud de un país.

Podemos notar que cada punto verde representa a un país, donde la componente en el eje $x$ corresponde a su inversión en salud, mientras que el eje $y$ toma el valor de la esperanza de vida.

En algunas ocasiones podemos toparnos con que gran parte de los datos siguen una clara tendencia, pero a su vez, existen algunos valores que están muy lejos de la línea de tendencia. A continuación, estudiaremos la definición de estos datos y cómo identificarlos.

**DATOS ATÍPICOS EN UN GRÁFICO DE DISPERSIÓN**

Un valor o dato atípico es una observación extrema que no sigue el patrón o la tendencia general del resto de los datos. En un gráfico de dispersión, los datos atípicos son, en general, aquellos que se encuentran alejados del resto de las observaciones.

 Figura 2: Gráfico de dispersión con tendencia lineal con los datos atípicos encerrados en rojo.

**SÍNTESIS**

* Los gráficos de dispersión, también llamados nube de puntos, se utilizan cuando se quiere estudiar si existe relación entre dos variables cuantitativas.
* En los gráficos de dispersión, cada punto representa un par ordenado $(x,y)$, donde $x$ e $y$ corresponden a los valores de las variables que identifican a un mismo individuo
* Un valor o dato atípico en un gráfico de dispersión es una observación que **no sigue** el patrón o la tendencia de la mayoría de los datos.