

Apuntes Unidad 1

Correlaciones

RELACIONES ENTRE VARIABLES

Los gráficos de dispersión nos ayudan a explorar y visualizar si existe algún tipo de relación entre dos variables cuantitativas. En ciertas ocasiones, se puede observar cierto patrón en los datos. A continuación, se pueden identificar algunos de estos casos.

- Puede ocurrir que la variable en el eje y tienda a aumentar a medida que aumenta la variable del eje x .

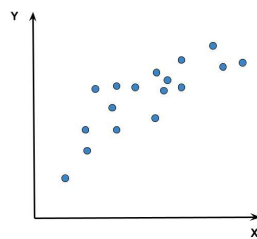


Figura 1: Gráfico de dispersión

- O también, podemos ver un tipo de relación opuesta, donde a medida que aumenta el valor de la variable del eje x , disminuye el de la variable del eje y .

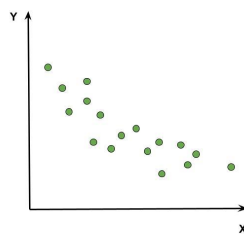


Figura 2: Gráfico de dispersión

- Otro caso posible es que no se aprecie ningún patrón claro en la nube de puntos.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Gráficos de dispersión y correlaciones

Contenido: Correlaciones

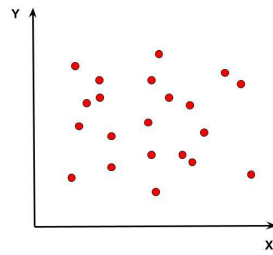


Figura 3: Gráfico de dispersión sin patrón claro

LÍNEAS DE TENDENCIA EN UN GRÁFICO DE DISPERSIÓN

Muchas veces, al estudiar la relación entre dos variables usando un gráfico de dispersión, es posible observar un patrón o regularidad en cómo se ubican los puntos. A la representación de este patrón en el gráfico le llamaremos **línea de tendencia**.

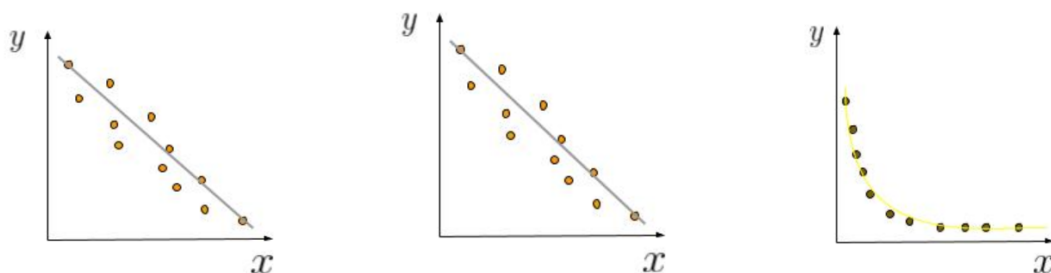


Figura 1: Gráficos de dispersión con sus líneas de tendencia.

Esta tendencia se puede clasificar en dos principales grupos, tendencia lineal y tendencia no lineal. Diremos que la relación entre dos variables sigue una tendencia lineal cuando, al representarlas en un gráfico de dispersión, sus puntos se encuentran cercanos a una recta que marca la línea de tendencia.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Gráficos de dispersión y correlaciones

Contenido: Correlaciones

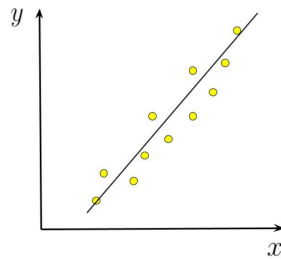


Figura 2: Gráficos de dispersión que siguen tendencias lineales.

Por ejemplo, en el gráfico de dispersión de los pesos y alturas de los estudiantes de un curso, vemos que los puntos no están alineados formando una recta, pero que sí tienden a agruparse en torno a una.

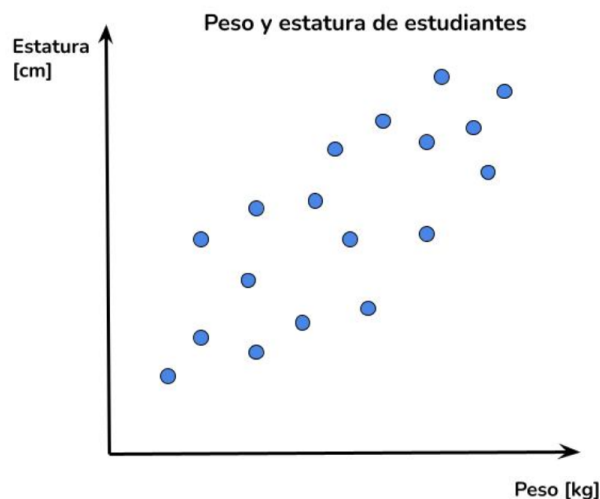


Figura 3: Gráfico de dispersión que sigue una tendencia lineal.

En este caso, podemos hablar de que hay una tendencia lineal en la relación entre peso y estatura.

También existen situaciones en las que, si bien los puntos representados en el gráfico de dispersión siguen un patrón, la línea de tendencia no es una recta. En estos casos decimos que la relación entre las variables no sigue una tendencia lineal.

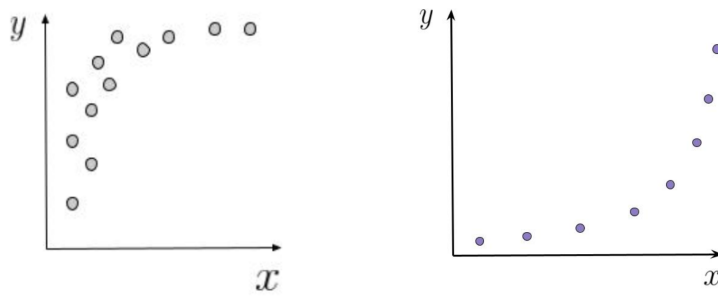


Figura 4: Gráficos de dispersión que siguen tendencias no lineales.

Por último, puede ocurrir que, al representar dos variables en un gráfico de dispersión, los puntos no describen ningún patrón claro, es decir, que no se puede trazar ninguna línea de forma que la mayoría de los puntos queden cercanos a ella. En estos casos diremos que no hay una tendencia en la relación.

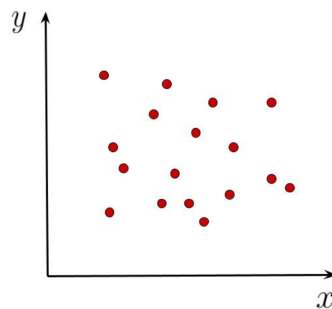
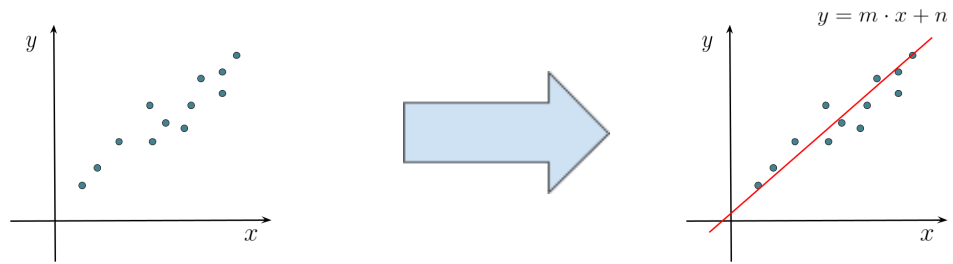


Figura 5: Gráfico de dispersión sin ninguna tendencia

RECTAS QUE MODELAN UNA RELACIÓN UNA TENDENCIA LINEAL

Si en un gráfico de dispersión los puntos siguen una tendencia lineal, podemos representar el patrón mediante una recta que está cerca de los puntos. Esta recta modela la relación entre las variables x e y para los datos y puede resultar útil para estimar el valor de una a partir de la otra. La recta que modela los datos se puede describir con la ecuación:

$$y = mx + n$$



Donde m es la pendiente y n es el coeficiente de posición de la recta. En este caso, existe una correspondencia entre el signo de la pendiente y la relación entre las variables, la cual se puede separar en dos casos:

- Si al aumentar la variable x la variable y también aumenta, la pendiente de la recta es positiva y por tanto decimos que la relación entre las variables sigue una **tendencia lineal positiva**.

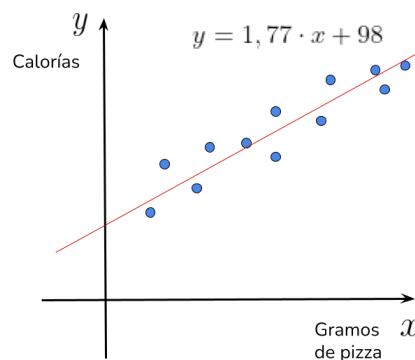


Figura 6: Ejemplo de tendencia lineal positiva.

- Por otro lado, si al aumentar la variable x , la variable y disminuye, la pendiente de la recta es negativa. En tal caso decimos que la relación entre las variables sigue una **tendencia lineal negativa**.

Curso: Probabilidad y estadística descriptiva

Unidad 1 : ¿Qué dicen los gráficos? Análisis crítico de la información

Tema: Gráficos de dispersión y correlaciones

Contenido: Correlaciones

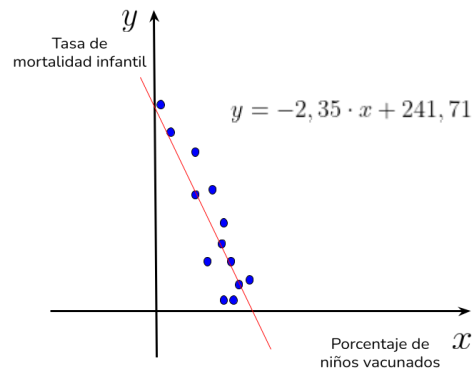


Figura 7: Ejemplo de tendencia lineal negativa.

Conocer la línea de tendencia nos permite entender de mejor manera la relación entre dos variables cuantitativas. Además, nos brinda la posibilidad de estimar una de las variables, conociendo el valor de la otra.

USANDO EL MODELO LINEAL PARA ESTIMAR

La variable que se representa en el eje x se llama **variable independiente**, mientras que a la variable y se le llama **variable dependiente**. La ecuación de la recta $y = mx + n$, que describe la tendencia lineal de los puntos en un gráfico de dispersión, puede usarse para predecir o estimar el valor de la variable dependiente y para un cierto valor de la variable independiente x .

ESTIMANDO VALORES UTILIZANDO LA TENDENCIA LINEAL

Hemos visto que la recta $y = mx + n$ que modela la relación entre dos variables cuantitativas x e y sigue una tendencia lineal, la cual permite describir el comportamiento de los datos en su conjunto. Supongamos que x toma un valor en particular, que denominaremos x_0 . El valor $y_0 = m \cdot x_0 + n$ entrega una aproximación del valor real de la variable y cuando x toma el valor de x_0 , si asumimos que los datos siguen ese patrón o tendencia.

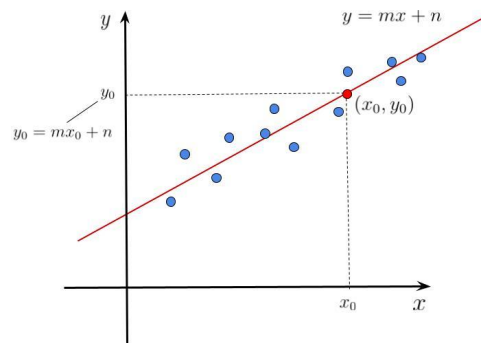


Figura 1: Gráfico de dispersión junto a la línea de tendencia asociada a los datos. Además de los valores usados en la explicación.

Notemos que los puntos (x, y) que están sobre la recta son exactamente aquellos para los que se cumple que $y = m \cdot x + n$. De esta forma, para un punto (x_0, y_0) del gráfico de dispersión, tenemos lo siguiente:

- Si está arriba de la recta, se cumple que $y_0 > m \cdot x_0 + n$ y por tanto, tiene un valor mayor al que se obtiene de forma aproximada por medio de la recta.

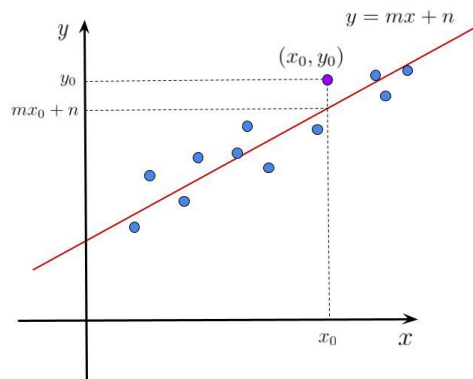


Figura 2: Situación donde el valor real es mayor al aproximado por medio de la recta.

- Si está debajo de la recta, se cumple que $y_0 < m \cdot x_0 + n$ y por tanto, tiene un valor menor que el que obtiene de forma aproximada por medio de la recta.

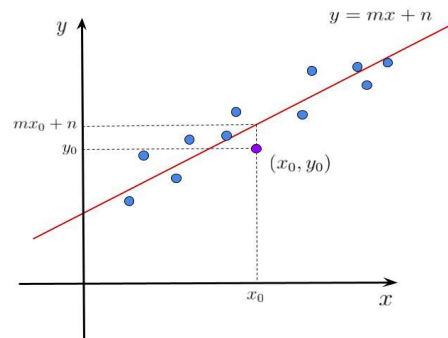


Figura 3: Situación donde el valor real es menor al aproximado por medio de la recta.

- Si está sobre o muy cerca de la recta, se cumple que $y_0 \approx m \cdot x_0 + n$ y por tanto, tiene un valor igual o cercano al que se obtiene de forma aproximada por medio de la recta.

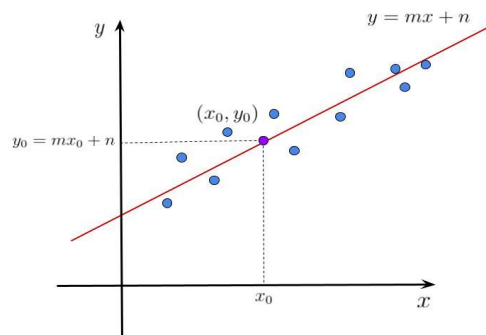


Figura 4: Situación donde el valor real es igual o cercano al aproximado por medio de la recta.

SÍNTESIS

- Si los datos en un gráfico de dispersión que representa la relación entre dos variables x e y siguen una tendencia **lineal**, entonces esta relación se puede modelar mediante la ecuación de una recta de la forma: $y = mx + n$.
- La variable x , que se representa en el eje horizontal, se llama variable **independiente**, mientras que la variable y , que está en el eje vertical, se llama variable **dependiente**.

- Si la recta que modela la relación entre las variables x e y tiene pendiente positiva, decimos que los datos siguen una tendencia lineal **positiva**. Si la pendiente de la recta es negativa, entonces se dice que la relación sigue una tendencia lineal **negativa**.
- La ecuación de la recta $y = mx + n$ que describe la tendencia lineal puede resultar útil para encontrar el valor **aproximado** que debe tener la variable y para un cierto valor de la variable x
- Al estudiar la relación entre dos variables usando un gráfico de dispersión pueden ocurrir distintas situaciones:
 - Si la variable en el eje x aumenta, la variable en el eje y aumenta.
 - Si la variable en el eje x aumenta, la variable en el eje y disminuye.
 - No se observa ningún **patrón** claro en la nube de puntos.
- La línea de tendencia corresponde a un **patrón** que da cuenta de la relación de dos variables en el gráfico.
- Dos variables tienen una tendencia lineal cuando al representarlas en un gráfico de dispersión la mayoría de sus puntos tienden a alinearse en torno a una recta
- Los puntos (x, y) que están sobre la recta que modela la relación entre estas variables son exactamente aquellos para los que se cumple que $y = m \cdot x + n$. De esta forma, para un punto (x_0, y_0) del gráfico de dispersión que se ubica:
 - arriba de la recta, se cumple que $y_0 > m \cdot x_0 + n$ y por tanto, tiene un valor mayor al que se obtiene de forma aproximada por medio de la recta.
 - debajo de la recta, se cumple que $y_0 < m \cdot x_0 + n$, por lo que tiene un valor menor que el que obtiene de forma aproximada por medio de la recta.
 - sobre o muy cerca de la recta, se cumple que $y_0 \approx m \cdot x_0 + n$ y por tanto, tiene un valor igual o cercano al que se obtiene de forma aproximada por medio de la recta.
- Comparar los datos reales con los calculados usando la recta que modela la relación lineal entre dos variables puede ser útil para describir o explicar fenómenos y sacar conclusiones.