

4 TEMA

MÓDULO:
FUNDAMENTOS DE ESTADÍSTICA

REGRESIÓN Y CORRELACIÓN

DOLORES LORENTE

Diplomada en Estadística y Graduada
en Estadística aplicada por la UCM.
Responsable científica de datos en Big
Data Analytics e Innovación.

STARWARS EPISODE II ATTACK OF THE CLONES



Institut de Formació Contínua-IL3
UNIVERSITAT DE BARCELONA

© de esta edición: Fundació IL3-UB, 2020

ÍNDICE

Objetivos Específicos

1. Regresión y correlación

Ideas clave



OBJETIVOS ESPECÍFICOS

- Trabajar los principales métodos de análisis bivariantes, como son la regresión y la correlación.
- Interpretar analíticamente los resultados obtenidos, las variables, sus atributos y la manera en la que influyen unas con otras.

1. REGRESIÓN Y CORRELACIÓN

Su **objetivo** es determinar si existe **relación entre dos variables**. Por ejemplo, si tenemos dos variables como el peso y la altura, o la edad y el salario... se trata de dar respuesta y averiguar si una variable influye en la otra. Por ejemplo:

- ¿Influye la altura en el peso de un individuo?
- ¿Influye la edad en la renta que percibe alguien?
- ¿Influye el consumo de tabaco en el cáncer de pulmón?
- ¿Influye el consumo de alcohol en el cáncer de pulmón?

Para dar respuesta a este tipo de preguntas, contamos con dos teorías:

1. Teoría de Regresión:

- Consiste en **encontrar una función que se ajuste lo mejor posible a un conjunto de puntos observados**. Es decir, consiste en la **búsqueda** de una función que exprese, lo mejor posible, **el tipo de relación entre dos o más variables**.
- Gráficamente, equivale a encontrar una curva que, aunque no pase por todos los puntos, esté lo más próxima posible de ellos.
- Para seleccionar la curva que mejor se ajusta a la nube de puntos, se usa el **método de los Mínimos Cuadrados**.
- Una de las aplicaciones más importantes que tiene la Regresión es la de **predecir**. Es decir, conocido el valor de una de las variables, se querrá **estimar el valor que presentará** la otra variable relacionada con ella.

2. Teoría de Correlación:

- Estudia el grado de dependencia entre las variables. En esta ocasión, el objetivo es medir **el grado de ajuste existente** entre la función teórica (función ajustada) y la nube de puntos.
- Cuando la relación funcional que liga las variables X e Y es una recta, entonces la regresión y correlación reciben el nombre de **Regresión Lineal** y **Correlación Lineal**, respectivamente.
- Una de las medidas más importantes de la **Correlación Lineal es el Coeficiente de Correlación Lineal de Pearson**.

La parte de **estadística inferencial** (de momento no profundizaremos en ella en este módulo) trata de analizar si este tipo de relación es extrapolable a la población. Por ahora, sólo se analizará la relación entre dos variables.

A partir de este punto, vamos a continuar en Colab, herramienta con la que ya sabes trabajar.



IDEAS CLAVE

- Se pueden realizar **inferencias o predicciones a través de las rectas de regresión**.
- **La correlación no implica causalidad**, se debe de analizar para realizar una interpretación correcta de los datos.
- **Siempre hay que analizar la robustez del modelo a través de la bondad de ajuste y de la correlación**. En función del tipo de variables del modelo de estudio, la correlación se estudiará con el coeficiente de correlación lineal de Pearson, de Spearman, Kendall, chi-cuadrado o el de contingencia.
- Para aquellas **regresiones que no sean de tipo lineal se buscará una transformación** para "linealizar" la expresión matemática y simplificar así el cálculo.