



Regresión Lineal

Tutoría

{desafío}
latam_



¿Qué entendemos por
correlación?

¿Cómo podríamos expresar
correlaciones?



Autoaprendizaje

Recursos asincrónicos

- ¿Revisaste los recursos de la semana 6 (Guía y desafío)?
- ¿Tienes dudas sobre alguno de ellos?



Ideas fuerza



La **correlación** nos indica el grado de **interrelación** entre dos o más variables, y nos permite suponer eventual **causalidad** entre ellas aunque no necesariamente será así.



La **regresión lineal** es un modelo que nos permite **relacionar variables** independientes con una dependiente, mediante una recta o un hiperplano



Podemos **evaluar un modelo** de regresión por medio de **métricas e indicadores**, que nos señalan cuánto se ajusta a los datos y su efectividad

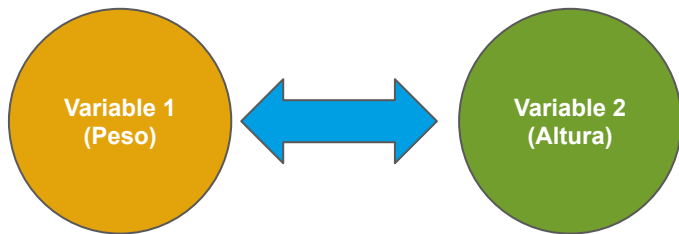
/*Causalidad y correlación*/

Correlación y causalidad

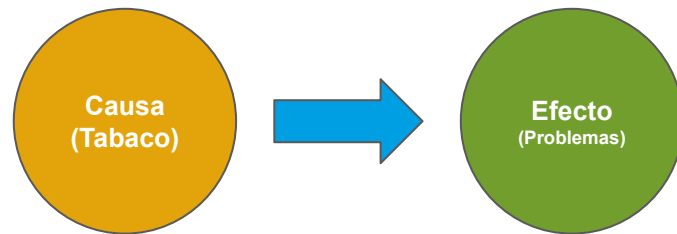
Diferencias

Causas y efectos son entidades ontológicamente diferentes. Las causas producen (o previenen) los efectos, pero los efectos no pueden producir las causas.

A esta asimetría lógica también le sigue una asimetría temporal: las causas siempre preceden en el tiempo a los efectos.



correlación

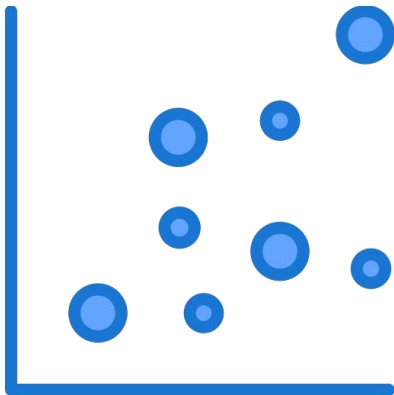


causalidad

Correlación

Definición

La correlación es una medida estadística que evalúa la relación y la fuerza de asociación entre dos o más variables. Se utiliza para determinar si existe una relación entre las variables, y para medir la dirección y la intensidad de esta relación.



/* Correlación lineal */

Correlación lineal

Coeficiente de correlación

La covarianza por sí sola puede ser difícil de interpretar porque su magnitud depende de las unidades en las que se miden las variables.

Podemos eliminar la influencia de las unidades dividiendo por las desviaciones estándar respectivas, y obtenemos el **coeficiente de correlación de Pearson (r)**

$$Cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Correlación lineal

Correlación lineal y coeficiente

Si los puntos parecen seguir una forma de recta hablamos de **correlación lineal**, que es cuantificada mediante el **coeficiente de correlación**. Este toma valores entre -1 y 1, donde:

- **-1** indica una correlación negativa perfecta (una variable aumenta mientras la otra disminuye).
- **1** indica una correlación positiva perfecta (ambas variables aumentan o disminuyen juntas).
- **0** indica una correlación nula (no hay relación lineal entre las variables).

/*Regresión*/

Regresión

Regresión Lineal - dos variables

En la regresión lineal entre dos variables, la relación entre las variables se modela mediante la ecuación de una línea recta, que tiene la forma:

$$y = \beta_0 + \beta_1 * x \pm \varepsilon$$

Donde:

- y es la variable dependiente que estamos tratando de predecir.
- x es la variable independiente.
- β_0 es la intersección de la recta con el eje Y, también llamada **intercepto**.
- β_1 es la pendiente de la recta
- ε es el término de error, que tiene en cuenta las variaciones no explicadas por el modelo.

Regresión

Regresión lineal múltiple

Una extensión de la regresión lineal simple es la regresión múltiple, que permite modelar la relación entre una variable dependiente y dos o más variables independientes.

En lugar de considerar solo una variable independiente, como en la regresión lineal simple, la regresión múltiple tiene en cuenta múltiples factores que podrían influir en la variable dependiente. Se obtiene así un plano (o hiperplano) de la forma

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_n * x_n + \varepsilon$$

/*Analizando el modelo*/

Statsmodels

Resultados de la regresión

	coef	std err	t	P> t	[0.025	0.975]
const	-5872.0927	310.285	-18.925	0.000	-6482.472	-5261.713
flipper_length_mm	50.1533	1.540	32.562	0.000	47.123	53.183

Python aplica un algoritmo iterativo, calculando pendientes entre puntos y promediando. Por esta razón, el intercepto y los coeficientes (pendiente) tienen valores dados pero con intervalos de confianza.

{desafío}
latam_

Analizando el modelo

Significancia de los predictores

La significancia de los predictores se refiere a la pregunta de si cada una de las variables independientes (predictores) individuales tiene un impacto significativo en la variable dependiente. Esto se evalúa mediante las pruebas de hipótesis t para los coeficientes de regresión individuales. Un valor de p bajo ($< \alpha$) para un coeficiente de regresión indica que la variable independiente correspondiente es significativa en el modelo y tiene un impacto significativo en la variable dependiente.

	coef	std err	t	P> t	[0.025	0.975]
const	-5872.0927	310.285	-18.925	0.000	-6482.472	-5261.713
flipper_length_mm	50.1533	1.540	32.562	0.000	47.123	53.183

{desafío}
latam_

Analizando el modelo

Significancia del modelo

La significancia del modelo se refiere a la pregunta de si el modelo de regresión en su conjunto es estadísticamente significativo.

Esto se evalúa mediante el estadístico F y su correspondiente valor p en la prueba de análisis de varianza (ANOVA). Un valor de p bajo (menor que α , donde α es el nivel de significancia elegido) indica que el modelo en su conjunto es significativo, es decir, al menos una de las variables independientes tiene un impacto significativo en la variable dependiente.

Statsmodels

Significancia del modelo

OLS Regression Results

Dep. Variable:	body_mass_g	R-squared:	0.762
Model:	OLS	Adj. R-squared:	0.761
Method:	Least Squares	F-statistic:	1060.
Date:	Mon, 11 Sep 2023	Prob (F-statistic):	3.13e-105
Time:	18:48:48	Log-Likelihood:	-2461.1
No. Observations:	333	AIC:	4926.
Df Residuals:	331	BIC:	4934.
Df Model:	1		
Covariance Type:	nonrobust		

Analizando el modelo

Métricas de evaluación

Las métricas de evaluación del modelo se utilizan para medir la calidad del ajuste del modelo y su capacidad para hacer predicciones precisas. Algunas de las métricas comunes incluyen:

- R-cuadrado (R^2)
- R-cuadrado Ajustado (R^2 ajustado)
- Error Cuadrático Medio (MSE)
- Raíz del Error Cuadrático Medio (RMSE)
- Error Absoluto Medio (MAE)
- Validación Cruzada

Analizando el modelo

Métricas de evaluación

En resumen, la significancia del modelo se evalúa globalmente, la significancia de los predictores se evalúa individualmente y las métricas de evaluación del modelo proporcionan información sobre la calidad del ajuste y la capacidad predictiva del modelo de regresión lineal. Estos conceptos son fundamentales para comprender y evaluar la eficacia de un modelo de regresión lineal.

¡Manos a la obra!

Regresión lineal con Python



Regresión lineal con Python

Algoritmos y evaluación del modelo

Vamos a repasar lo visto en clases, utilizando Python, Además, veremos las herramientas con las que contamos para analizar los modelos creados.



Desafío “Regresión Lineal”



Desafío

"Regresión Lineal"

- ¿Leíste el desafío de esta semana? ¿Comprendes bien lo que se solicita en cada caso?
- ¿Hay contenidos que necesitas repasar antes de comenzar este desafío?
- ¿Necesitas algún ejemplo o indicación para alguna pregunta o requerimiento específico?





Próxima sesión...

Generar modelos predictivos utilizando scikit-learn de acuerdo a requerimientos

{desafío}
latam_

*Academia de
talentos digitales*

