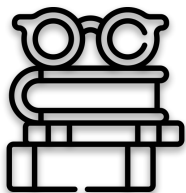


Glosario



Modelos de regresión lineal y series de tiempo



Consulta el Glosario de Científico de Datos según el tema relacionado con las clases.

- **Conoce los supuestos de la regresión de mínimos cuadrados ordinarios**
- **Verifica los supuestos clave en regresión de mínimos cuadrados**
- **Revisa el cumplimiento de supuestos en la regresión de mínimos cuadrados**

Autocorrelación:

Es la correlación de una señal con una versión retrasada de sí misma a través de distintos intervalos de tiempo. En el contexto de la regresión, se refiere a la correlación entre los residuales de un modelo y sus valores en observaciones anteriores. La autocorrelación puede ser un indicador de que el modelo no está capturando alguna estructura temporal en los datos.

Curtosis:

Es una medida estadística que describe la "agudeza" o "achatamiento" de la distribución de una variable aleatoria en comparación con la distribución normal. Una curtosis alta indica una distribución con colas pesadas o picos pronunciados, mientras que una curtosis baja sugiere colas ligeras o una distribución más plana.

Coeficiente de determinación (R cuadrada):

Es una medida estadística que indica la proporción de la varianza para una variable dependiente que es explicada por las variables independientes en un modelo de regresión. El R cuadrado varía entre 0 y 1, donde un valor más alto indica un mejor ajuste del modelo a los datos.

Desviación estándar del error de pronóstico:

Es una medida de la dispersión de los errores de pronóstico en un modelo de regresión, que indica cuánto se espera que varíen los pronósticos reales de los valores pronosticados.

Distribución T de Student:

Es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

Error estándar:

Es una medida de la dispersión de las estimaciones de un parámetro a lo largo de diferentes muestras. En el contexto de la regresión, se refiere a la estimación de la desviación estándar de los errores del modelo.

DurbinWatson:

Es un estadístico utilizado para detectar la presencia de autocorrelación en los residuales de un análisis de regresión. Un valor cercano a 2 sugiere ausencia de autocorrelación, mientras que valores significativamente menores o mayores indican autocorrelación positiva o negativa, respectivamente.

Factor de inflación de varianza (VIF):

Es una medida que cuantifica el aumento de la varianza de los coeficientes de un modelo de regresión causado por la multicolinealidad. Un VIF alto indica una fuerte correlación entre las variables independientes y un posible problema de multicolinealidad.

Heterocedasticidad:

Se refiere a la situación en la que la varianza de los errores de un modelo de regresión no es constante a lo largo de las observaciones. Esto puede ser un problema porque los métodos de estimación estándar, como los mínimos cuadrados ordinarios, asumen homocedasticidad, es decir, que la varianza de los errores es constante.

Intervalo de confianza:

Es un rango de valores, derivado de los datos de la muestra, que se cree, con un cierto nivel de confianza, que contiene el valor verdadero de un parámetro desconocido de la población.

Homocedasticidad:

Es la propiedad de tener una varianza constante de los residuales en un modelo de regresión. La homocedasticidad es un supuesto clave en la regresión lineal, ya que violaciones de este supuesto pueden llevar a estimaciones ineficientes y pruebas de hipótesis no válidas.

JarqueBera:

Es una prueba estadística que mide la desviación de la curtosis y el sesgo de una distribución respecto a una distribución normal. Se utiliza para probar la normalidad de los residuos en un modelo de regresión.

- Mínimos cuadrados ordinarios (Ordinary Least Squares OLS):

Es un método de estimación en el que se minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores ajustados por el modelo.

- Notación matricial:

Es una forma de representar conjuntos de ecuaciones lineales y operaciones sobre estas utilizando matrices y vectores.

- Multicolinealidad:

Es una situación en la que dos o más variables predictoras en un modelo de regresión múltiple están altamente correlacionadas, lo que puede dificultar la estimación de los coeficientes individuales de las variables y la interpretación del modelo.

- Pronóstico puntual (y gorro):

Es una estimación única para el valor de la variable dependiente dado un conjunto de valores de las variables independientes en un modelo de regresión.

Prueba de heterocedasticidad de White:

Es una prueba estadística que se utiliza para detectar la presencia de heterocedasticidad en un modelo de regresión. Si se rechaza la hipótesis nula de homocedasticidad, entonces hay evidencia de heterocedasticidad.

Residuales:

Son las diferencias entre los valores observados y los valores predichos por un modelo de regresión. Los residuales son una parte crucial en el diagnóstico de un modelo de regresión, ya que su distribución y propiedades pueden indicar la presencia de supuestos violados.

Raíz cuadrada de un vector:

En este contexto, se refiere a la operación de tomar la raíz cuadrada de cada elemento de un vector, que es parte del cálculo para obtener el margen de error en un pronóstico de intervalo.

Resíduos:

Son las diferencias entre los valores observados y los valores predichos por un modelo de regresión. Los resíduos son una parte crucial en la verificación de los supuestos de un modelo de regresión.

● Sesgo:

En estadística, el sesgo es una medida de la tendencia de una distribución a ser tirada hacia la izquierda o la derecha de la media. Una distribución con sesgo positivo tiene una cola más larga hacia la derecha, mientras que una con sesgo negativo tiene una cola más larga hacia la izquierda.

● Suma total de cuadrados (TSS):

Es la suma de los cuadrados de las diferencias entre los valores observados y el promedio de esos valores. Es una medida de la variabilidad total de los datos.

● Suma de residuales al cuadrado (RSS):

Es la suma de los cuadrados de las diferencias entre los valores observados y los valores pronosticados por un modelo de regresión. Es una medida de la discrepancia entre los datos y el modelo.

● Valor estadístico de la distribución T de Student:

Es un valor utilizado en pruebas estadísticas que sigue una distribución t de Student. Se utiliza para determinar la significancia estadística de los coeficientes en un modelo de regresión y para construir intervalos de confianza.

● Varianza del error:

Es una medida de la dispersión de los errores en un modelo de regresión, y se calcula como la suma de los cuadrados de los residuales dividida entre el número de grados de libertad.

● VIF (Factor de Inflación de la Varianza):

Es una medida que cuantifica el aumento de la varianza de un coeficiente de regresión debido a la multicolinealidad en un modelo de regresión múltiple. Un VIF alto indica una fuerte correlación entre las variables independientes y la posibilidad de multicolinealidad.

¡Éxito en tus estudios!