

Analítica y Proyecciones para Macroeconomía y Finanzas Ayudantía 04

Profesor: Andrés Sagner
Ayudante: Hiarela Aravena*

Otoño 2024

1. Comentes

1. La multicolinealidad no afecta la capacidad del modelo para predecir la variable dependiente, solo afecta las estimaciones de los coeficientes.

Respuesta:

Verdadero. La multicolinealidad no afecta las predicciones del modelo ni la calidad del ajuste (R-cuadrado), pero sí puede afectar la interpretación de los coeficientes individuales.

2. El VIF proporciona una medida precisa de la magnitud del sesgo inducido por la multicolinealidad en las estimaciones de los coeficientes de regresión.

Respuesta:

Falso. Si bien el VIF es una herramienta útil para detectar multicolinealidad y evaluar su impacto en un modelo de regresión, no proporciona una medida directa del sesgo en las estimaciones de los coeficientes. En cambio, el VIF calcula cuánto aumenta la varianza de los coeficientes debido a la multicolinealidad, lo que puede indicar la presencia de problemas en la estimación, pero no cuantifica el sesgo en sí mismo.

* haravenaro@fen.uchile.cl



3. La multicolinealidad perfecta siempre se manifiesta como una relación lineal exacta entre dos o más variables independientes en un modelo de regresión.

Respuesta:

Falso. Si bien la multicolinealidad perfecta implica una relación lineal exacta entre variables independientes, no siempre se manifiesta de manera evidente en los datos. En algunos casos, la multicolinealidad puede ser sutil y requerir técnicas estadísticas específicas para su detección, como la evaluación de los valores de VIF o la inspección de los coeficientes de correlación entre las variables.

2. Multicolinealidad y VIF utilizando la base de datos mtcars

La base de datos mtcars es un conjunto de datos clásico que contiene información sobre diferentes modelos de automóviles. Cada fila representa un modelo de automóvil, y las columnas contienen diversas características como el rendimiento en millas por galón (mpg), la cilindrada (disp), la potencia en caballos de fuerza (hp), el peso en toneladas (wt), y la relación de eje trasero (drat), entre otros.

Siga los siguientes pasos utilizando el software de su elección.

1. Ajuste un modelo de regresión utilizando la variable de respuesta mpg y las variables predictoras disp, hp, wt y drat. ¿Qué conclusiones puedes sacar del resultado del modelo, considerando el valor de R cuadrado, el estadístico F y los valores p de las variables predictoras?

Respuesta:

Podemos ver en el resultado que el valor de R cuadrado para el modelo es 0,8376. También podemos ver que el estadístico F general es 34.82 y el valor p correspondiente es $2.704e-10$, lo que indica que el modelo de regresión general es significativo. Además, las variables predictoras hp y wt son estadísticamente significativas con un nivel de significancia de 0,05, mientras que disp y drat no lo son.



2. Utilice la función `vif()` de la biblioteca de automóviles para calcular el VIF para cada variable predictora en el modelo. Interprete el significado del VIF y sus implicaciones en el análisis de regresión.

Respuesta:

Podemos ver que el VIF tanto para `disp` como para `wt` es mayor que 5, lo cual es potencialmente preocupante.

3. Para visualizar los valores de VIF para cada variable predictora, cree un gráfico de barras horizontales y agregaremos una línea vertical en 5 para identificar claramente los valores de VIF que exceden este umbral.
4. Para comprender mejor por qué una variable predictora puede tener un valor VIF alto, cree una matriz de correlación para observar los coeficientes de correlación lineal entre cada par de variables. ¿Qué conclusiones puedes sacar sobre la relación entre las variables predictoras? ¿Por qué una variable predictora puede tener un valor VIF alto? ¿Cuál sería tu recomendación sobre cómo manejar esta situación en el modelo de regresión?

Respuesta:

A partir de la matriz de correlación podemos ver que `disp` está fuertemente correlacionado con las otras tres variables predictoras, lo que explica por qué tiene un valor de VIF tan alto.

En este caso, es posible que desee quitar `disp` del modelo porque tiene un valor VIF alto y no fue estadísticamente significativo en el nivel de significación de 0,05.

Una matriz de correlación y un VIF le proporcionarán información similar: ambos le indican cuándo una variable está altamente correlacionada con una o más variables en un modelo de regresión.

3. Explorando los Rincones de Boston

El conjunto de datos de Boston contiene información sobre diversas características de los suburbios de Boston y el valor medio de las viviendas en esos suburbios. Las variables incluidas en el conjunto de datos son características socioeconómicas, como la tasa de criminalidad, la proporción de terreno residencial, la tasa de impuesto a la propiedad, entre otras.

Siga los siguientes pasos utilizando el software de su elección.



-
1. Divida aleatoriamente los datos en un conjunto de entrenamiento (80 % de los datos) para crear un modelo predictivo y un conjunto de prueba (20 % de los datos) para evaluar el modelo. Asegúrese de establecer la semilla para la reproducibilidad.
 2. Cree un modelo de regresión que incluya todas las variables predictoras del conjunto de datos de Boston, con *medv* como variable dependiente. ¿Qué significan las métricas de rendimiento del modelo, RMSE y R2?
 3. Utilice la función `vif()` del paquete *car* para detectar multicolinealidad en el modelo de regresión. En nuestro ejemplo, la puntuación VIF para la variable predictora es muy alta. ¿Qué implicaciones podría tener esto para nuestro modelo de regresión?
 4. Actualice nuestro modelo eliminando las variables predictoras con un valor VIF alto. ¿Qué conclusiones puedes sacar de esto sobre la influencia de la variable eliminada en el modelo original?