

Copilot

Claro! Vamos realizar um teste de **Granger-Causality** usando a função `grangercausalitytests()` da biblioteca **statsmodels** em Python. Essa métrica é usada para determinar se uma série temporal é útil para prever outra. Aqui está um exemplo passo a passo:

1. **Carregando os Dados:** Vamos usar um conjunto de dados que contém valores para o número de ovos produzidos nos EUA juntamente com o número de galinhas de 1930 a 1983. Primeiro, carregue os dados:

```
import pandas as pd

# Defina a URL onde o conjunto de dados está localizado
url = "https://raw.githubusercontent.com/Statology/Miscellaneous/main/chicken_egg.txt"

# Leia o conjunto de dados como um DataFrame do pandas
df = pd.read_csv(url, sep="\t")

# Visualize as primeiras cinco linhas do DataFrame
df.head()
```

1. **Realizando o Teste de Granger-Causality:** Agora, usaremos a função `grangercausalitytests()` para verificar se o número de ovos produzidos é preditivo do número futuro de galinhas. Vamos executar o teste com três lags:

```
from statsmodels.tsa.stattools import grangercausalitytests

# Realize o teste de Granger-Causality
grangercausalitytests(df[['chicken', 'egg']], maxlag=3)
```

O resultado inclui estatísticas F e valores p correspondentes para diferentes testes. Se o valor p for menor que 0,05, podemos rejeitar a hipótese nula e concluir que saber o número de ovos é útil para prever o futuro número de galinhas. No nosso exemplo, o valor p é 0,0030, o que nos permite rejeitar a hipótese nula. Portanto, o número de ovos Granger-causa o número de galinhas. Lembre-se de substituir os dados pelos seus próprios, se necessário! 😊

Copilot

O parâmetro “maxlag” no teste de Granger-Causality especifica o número máximo de lags (atrasos) a serem considerados no teste. Em outras palavras, ele determina quantos períodos anteriores da série temporal devem ser incluídos na análise para verificar se uma série é preditiva da outra. Um valor maior para “maxlag” permite explorar relações causais em períodos mais distantes, mas também pode aumentar a complexidade do teste. Portanto, é importante escolher um valor adequado com base no contexto dos dados e na frequência das observações. 😊

Copilot

Claro! Vou explicar o output gerado pela função `grangercausalitytests()` no contexto do teste de Granger-Causality.

O teste de Granger-Causality avalia se uma série temporal é útil para prever outra série temporal. Ele compara dois modelos:

1. Modelo Restrito (sem causalidade):

- Este modelo inclui apenas os valores passados da série que estamos testando (por exemplo, a série de ovos).
- Não inclui os valores passados da outra série (por exemplo, a série de galinhas).

2. Modelo Completo (com causalidade):

- Este modelo inclui tanto os valores passados da série que estamos testando quanto os valores passados da outra série.
- Ele permite que a série de ovos seja preditiva da série de galinhas.

A função `grangercausalitytests()` realiza vários testes estatísticos para verificar se o Modelo Completo é significativamente melhor do que o Modelo Restrito. Vamos entender os testes específicos:

1. Teste F (F-test):

- O teste F compara a variância explicada pelo Modelo Restrito com a variância explicada pelo Modelo Completo.
- Se o valor p do teste F for menor que um nível de significância (geralmente 0,05), rejeitamos a hipótese nula de que o Modelo Restrito é igualmente bom ou melhor que o Modelo Completo.
- Em outras palavras, se o valor p for baixo, há evidência de que a série de ovos Granger-causa a série de galinhas.

2. Teste Qui-Quadrado (Chi-squared test):

- O teste qui-quadrado verifica se os coeficientes dos lags adicionados no Modelo Completo são estatisticamente diferentes de zero.
- Se o valor p do teste qui-quadrado for baixo, isso sugere que os lags adicionados são significativos e melhoram o modelo.

3. Teste Wald (Wald test):

- O teste Wald verifica se os coeficientes dos lags adicionados são diferentes de zero.
- Se o valor p do teste Wald for baixo, isso também sugere que os lags adicionados são relevantes.

Em resumo, se algum desses testes indicar significância estatística (valor p baixo), podemos concluir que a série de ovos Granger-causa a série de galinhas. Caso contrário, não há evidência suficiente para afirmar a causalidade. Espero que isso esclareça! 😊

Copilot

Claro! Vou explicar o significado de “df_denom” e “df_num” no contexto do teste de Granger-Causality.

- **df_denom (Graus de Liberdade do Denominador):**

- O termo “denominador” refere-se ao denominador da estatística F usada no teste.
- Representa o número de graus de liberdade associados ao modelo restrito (sem causalidade).
- É calculado como o número total de observações menos o número de parâmetros estimados no modelo restrito.
- Em outras palavras, é a quantidade de informação disponível para estimar os parâmetros do modelo.

- **df_num (Graus de Liberdade do Numerador):**

- O termo “numerador” refere-se ao numerador da estatística F.
- Representa o número de graus de liberdade associados ao modelo completo (com causalidade).
- É calculado como o número de parâmetros estimados no modelo completo.
- Reflete a quantidade de informação usada para estimar os parâmetros adicionais no modelo completo.

Em resumo, os graus de liberdade indicam quantos parâmetros foram estimados nos modelos e quantos dados estão disponíveis para avaliar a significância estatística. Espero que isso esclareça sua dúvida! 😊