**Công tác thí nghiệm số 1. Phần 2**

**Phần 2. Phương pháp bình phương tối thiểu tổng quát.**

Mục tiêu: củng cố kiến thức lý thuyết về hồi quy tuyến tính, rèn luyện kỹ năng thực hành xây dựng mô hình hồi quy tuyến tính.

Nhiệm vụ:

* Ghi nhớ và hiểu các khái niệm cơ bản của mô hình hồi quy tuyến tính;
* Tìm hiểu cách xây dựng mô hình hồi quy tuyến tính bằng phương pháp bình phương tối thiểu tổng quát;
* Tìm hiểu cách sử dụng thư viện statsmodels ;

Bình phương tối thiểu tổng quát

Bình phương tối thiểu tổng quát (GLS) là một phương pháp ước tính các tham số của mô hình hồi quy tuyến tính được sử dụng khi các giả định về bình phương tối thiểu tiêu chuẩn (OLS) không đúng. Đặc biệt, OLS được sử dụng trong trường hợp lỗi mô hình có tính phương sai thay đổi (các phương sai khác nhau) hoặc tương quan.

Để tìm hệ số hồi quy cho OLS, chúng tôi sử dụng công thức sau:

ma trận hiệp phương sai của lỗi mô hình ở đâu

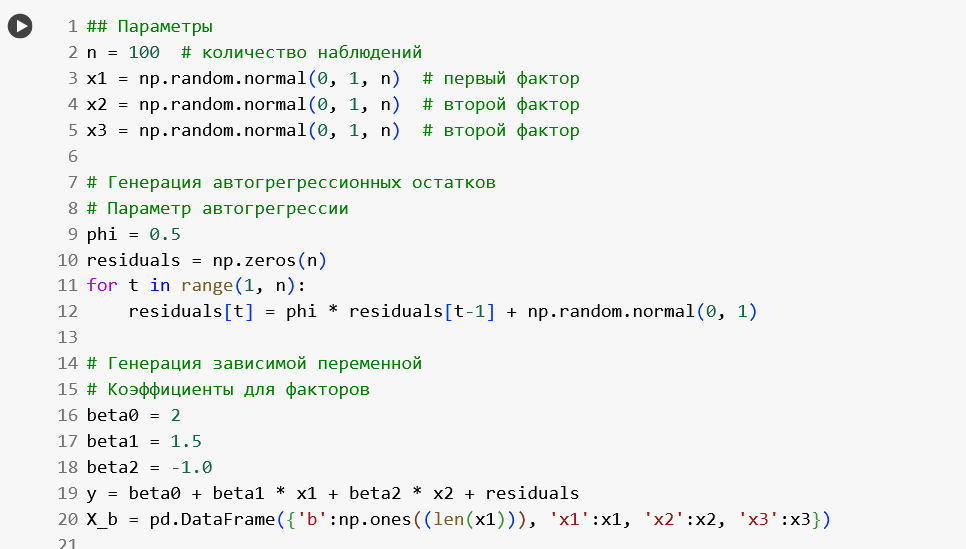
Để xây dựng mô hình này cần tính toán ma trận hiệp phương sai.

Đầu tiên, hãy nhập tất cả các thư viện cần thiết:

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Hãy xem một ví dụ. Hãy tạo dữ liệu với số dư tự hồi quy:



Tiếp theo, chúng tôi xây dựng hồi quy bằng phương pháp bình phương tối thiểu tiêu chuẩn:

Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описание

Chúng tôi nhận được phần còn lại của mô hình này.



Tiếp theo, cần phải có được sự đánh giá về bản chất của mối quan hệ tự tương quan của phần dư. Người ta thường cho rằng mối quan hệ này có thể được thể hiện bằng mô hình tự hồi quy bậc nhất ( AR (1)). Điều này có nghĩa là lỗi có thể được thể hiện như sau:

đâu là hệ số tự tương quan của sai số thứ tự đầu tiên.

Để tìm, chúng ta sẽ xây dựng mô hình hồi quy tuyến tính cho phần dư.

*Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание*

Hệ số sẽ là tham số cho biến độc lập. Hãy viết nó thành một biến:

Изображение выглядит как текст, Шрифт, дизайн

Автоматически созданное описание

Như chúng ta đã biết, quy trình AR(1) có nghĩa là các láng giềng gần nhau có kết nối mạnh hơn, vì vậy chúng ta có thể xác định cấu trúc này bằng ma trận Toeplitz.



Ma trận Toeplitz là ma trận vuông trong đó các phần tử chỉ phụ thuộc vào hiệu chỉ số. Nghĩa là phần tử ma trận a ij​ chỉ phụ thuộc vào i−j. Điều này có nghĩa là các giá trị trên mỗi đường chéo song song với đường chéo chính là như nhau.

Để xây dựng ma trận này, bạn cần nhập nó từ thư viện scipy.

Tiếp theo, chúng tôi xây dựng ma trận hiệp phương sai:



Trong ví dụ này, chúng tôi đã nâng hệ số tự tương quan lên lũy thừa từ ma trận Toeplitz và ghi nó vào một mảng.

Khi đó các hệ số hồi quy, theo OLS, có thể được xác định như sau:



Điều tương tự có thể được thực hiện bằng thư viện statsmodels

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Bài tập

1. Tạo tập dữ liệu với biến phụ thuộc y và biến độc lập x1 và x2. Thêm hiện tượng tự tương quan vào phần dư.
2. Sử dụng bình phương tối thiểu để ước lượng mô hình và báo cáo kết quả.
3. Sử dụng bình phương tối thiểu tổng quát để ước lượng mô hình và báo cáo kết quả.
4. Vẽ biểu đồ phần dư cho cả hai mô hình để đánh giá trực quan sự phân bố của chúng.
5. Tải xuống dữ liệu Boston . csv . Tạo một DataFrame thư viện gấu trúc bằng mã sau:

nhập gấu trúc dưới dạng pd

df = pd.read\_csv('boston.csv')

in(df.head())

1. Thực hiện một nghiên cứu phân phối dữ liệu. Trực quan hóa sự phụ thuộc. Để làm điều này, hãy tính toán các hệ số tương quan và trực quan hóa chúng bằng bản đồ nhiệt. Để thực hiện việc này, hãy sử dụng đoạn mã sau:

nhập khẩu seaborn dưới dạng sns

sns.heatmap(df.corr().round(1), annot = True, cbar=False)

plt.show()

1. Chia dữ liệu thành X và y :

X = df [[' CRIM ', ' ZN ', ' INDUS ', ' CHAS ', ' NOX ', ' RM ', ' TUỔI ', ' DIS ', ' RAD ', ' THUẾ ', ' PTRATIO ', ' B ', ' LSTAT ']]

y = df['MEDV']

1. Xây dựng mô hình hồi quy tuyến tính bội sử dụng OLS và GLS bằng thư viện statsmodels .
2. Xây dựng mô hình sử dụng GLS cho dữ liệu bất động sản.
3. Rút ra kết luận về chất lượng của mô hình đã xây dựng.