Кузнецов Михаил Пи19-4

Тема 7: "Прогнозирование значений эндогенной переменной, проверка её адекватности и ошибки спецификации"

Задание 1.

1. Используя 90% выборки постройте множественную модель линейной регрессии:

𝑦𝑦𝑖𝑖=𝛼𝛼+𝛽𝛽1𝑥𝑥1𝑖𝑖+𝛽𝛽2𝑥𝑥2𝑖𝑖+𝛽𝛽3𝑥𝑥3𝑖𝑖+𝛽𝛽4𝑥𝑥4𝑖𝑖+𝜀𝜀𝑖𝑖

2. Проверьте ее адекватность на основе оставшихся 10% наблюдений.

3. Изучите выборку на однородность, используя тест Чоу.

4. Постройте модель парной регрессии с наиболее влияющим фактором. Сравните исходную модель с парной моделью. Выберите лучшую.

5. Выбранную модель проверьте на наличие пропущенных переменных с помощью теста Рамсея.

6. Выполните задания, используя инструменты R и Python.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Объем денежных накоплений в РФ миллиард руб. (y) | Сводные данные о продаже наличной иностранной валюты кредитными организациями физическим лицам (x1) | Курс рубля к доллару США (x2) | Цена нефти марки «Юралс» (x3) | Динамика потребительских цен по группам товаров и услуг (месяц к соответствующему месяцу предыдущего года,%) (x4) |
| 12 345,00 | 2769 | 29,67 | 94,2 | 9,6 |
| 12 091,90 | 3065 | 28,94 | 101,1 | 9,5 |
| 12 339,70 | 3899 | 28,43 | 111,6 | 9,5 |
| 12 450,60 | 3790 | 27,5 | 119,7 | 9,6 |
| 12 832,10 | 3963 | 28,07 | 112 | 9,6 |
| 12 892,10 | 4224 | 28,08 | 112 | 9,4 |
| 13 224,50 | 4645 | 27,68 | 115,3 | 9 |
| 13 449,90 | 4914 | 28,86 | 109,6 | 8,2 |
| 13 476,80 | 4830 | 31,88 | 112,6 | 7,2 |
| 13 588,90 | 4815 | 29,9 | 108,7 | 7,2 |
| 13 600,40 | 4401 | 31,32 | 110,9 | 6,8 |
| 13 768,50 | 5189 | 32,2 | 108 | 6,1 |
| 14 796,60 | 3434 | 30,36 | 109,8 | 4,2 |
| 14 424,20 | 5195 | 28,95 | 119,2 | 3,7 |
| 14 607,30 | 5401 | 29,33 | 123,3 | 3,7 |
| 14 688,00 | 4239 | 29,36 | 117,8 | 3,6 |
| 15 057,30 | 4558 | 32,45 | 109,2 | 3,6 |
| 15 161,00 | 4675 | 32,82 | 93,5 | 4,3 |
| 15 534,60 | 6558 | 32,19 | 102,7 | 5,6 |
| 15 544,50 | 6253 | 32,92 | 113,5 | 5,9 |
| 15 644,40 | 5410 | 30,92 | 112 | 6,6 |
| 15 717,80 | 5023 | 31,53 | 110,8 | 6,5 |
|  |  |  |  |  |

1. Используя 90% выборки постройте множественную модель линейной регрессии:

𝑦𝑦𝑖𝑖=𝛼𝛼+𝛽𝛽1𝑥𝑥1𝑖𝑖+𝛽𝛽2𝑥𝑥2𝑖𝑖+𝛽𝛽3𝑥𝑥3𝑖𝑖+𝛽𝛽4𝑥𝑥4𝑖𝑖+𝜀𝜀𝑖𝑖

Обучающая выборка

Изображение выглядит как стол

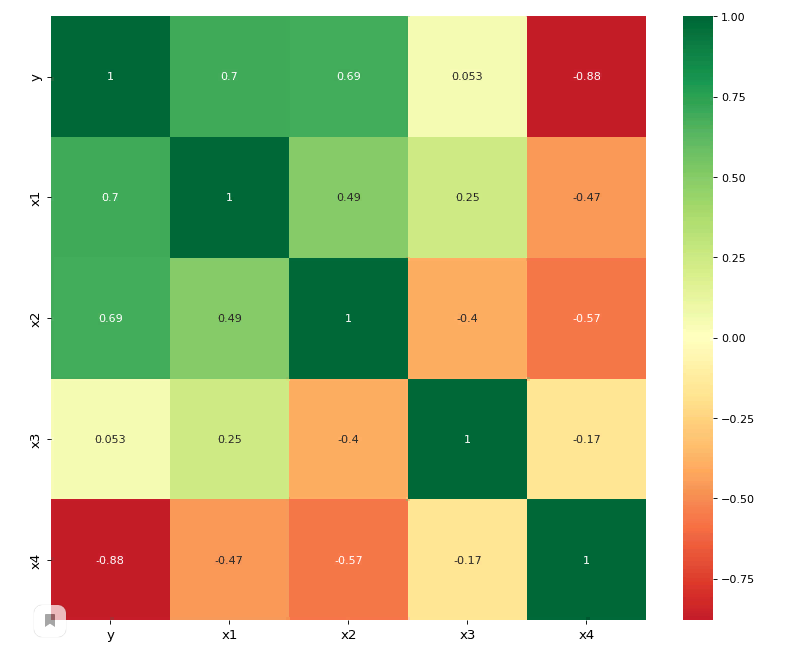
Автоматически созданное описание

Контролирующая выборка

Изображение выглядит как текст, стол

Автоматически созданное описание

Корреляционная матрица



Проверка на значимость модели в целом и каждого коэффициента по отдельности

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

|  |  |  |
| --- | --- | --- |
| Коэффициент | Значение | Вывод |
| R^2 | 0.906017889736171 | Модель очень качественная (R2 близка к 1) |
| R^2 скорректированный | 0.8809559936658167 | Модель очень качественная (R2 корр близка к 1) |
| Стандартная ошибка Se | 385.3485703727518 | Модель очень качественная, малая ошибка относительно дельты по y |
| Ошибка аппроксимации | 1.987952 | Малая ошибка, модель качественная |
| F-статистика | 36.15121087378124 |  |
| P-value | 1.548e-07 |  |

Можем сделать вывод о том, что модель очень хорошего качества и значима

Прогнозирование с учетом всех факторов

Изображение выглядит как текст, стол

Автоматически созданное описание

Y\_20=15644.4 ∉ (13385.995225;15115.148696) – предиктивный интервал

Y\_20=15644.4 ∉ (13980.617503;14520.526417) – доверительный интервал

Y\_21=15717.8 ∉ (13284.313696;15026.211982) – предиктивный интервал

Y\_21=15717.8 ∉ (13865.547656;14444.978022) – доверительный интервал

Y\_20 и Y\_21 не входят ни в один из интервалов (предиктивный и доверительный).  
После всего вышеперечисленного, сделаем вывод, что модель неадекватна.

Прогнозирование без учета фактора x3

Изображение выглядит как текст, стол

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Y\_20=15644.4 ∉ (13410.459547;15140.912739) – предиктивный интервал

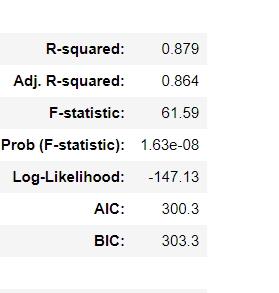
Y\_20=15644.4 ∉ (14009.406423;14541.965863) – доверительный интервал

Y\_21=15717.8 ∉ (13366.205596;15089.802348) – предиктивный интервал

Y\_21=15717.8 ∉ (13973.083853;14482.924092)– доверительный интервал

Y\_20 и Y\_21 не входят ни в один из интервалов (предиктивный и доверительный).  
Сделаем вывод, что модель опять же неадекватна

Прогнозирование без учета фактора x3 и х2



Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Y\_20=15644.4 ∉ (13331.321865;15152.829562) – предиктивный интервал

Y\_20=15644.4 ∉ (13964.776816;14519.374611) – доверительный интервал

Y\_21=15717.8 ∉ (13212.525975;15002.577309) – предиктивный интервал

Y\_21=15717.8 ∉ (13887.340267;14327.763017)– доверительный интервал

Y\_20 и Y\_21 не входят ни в один из интервалов (предиктивный и доверительный).  
Сделаем вывод, что модель опять же неадекватна

Тест Чоу

Делим выборку на 2 равнозначные части ( по 10 наблюдений )

Изображение выглядит как стол

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

Строим по этим выборкам модели

Изображение выглядит как текст, квитанция, снимок экрана

Автоматически созданное описание

Изображение выглядит как стол

Автоматически созданное описание

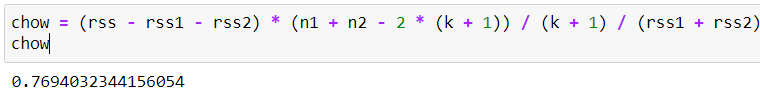
H0: выборки однородны, нет структурного сдвига.

H1: выборки не однородны, есть структурный сдвиг.

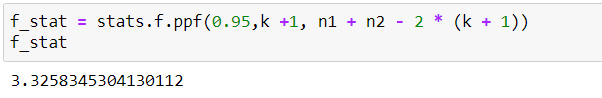
Если Fнабл < Fкр -> h0, нет структурного сдвига, выборки однородны

Если Fнабл > Fкр -> h1, есть структурный сдвиг, выборки не однородны

Ищем Fнабл



Ищем Fкр



У нас Fнабл < Fкр -> h0, нет структурного сдвига, выборки однородны.