Лабораторная работа №2

Марковский Валентин Александрович

25 12 2020

## Изначальная модель

Модель 0:

## Список столбцов файла:

* fo - название федерального округа
* iso\_code - код региона субъекта РФ
* region - название субъекта РФ
* Y - Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по видам экономической деятельности: обрабатывающие производства за 2014 год
* X1 - Индексы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства за 2014 год
* X2 - Изменение среднегодовой численности занятых за 2014 год
* X3 - Индекс физического объема ВРП за 2013 год
* X4 - Сумма убытка организаций по отдельным видам экономической деятельности: обрабатывающие производства за 2013 год

Данные по пяти показателям по обрабатывающим производствам за 2013-2014 гг. по регионам РФ. Данные собраны из файлов, загруженных из базы Росстата <https://www.gks.ru/folder/210/document/13204> 25 декабря 2020 года. После загрузки файлы, каждый за отдельный год, были объединены в одну таблицу. Пропущенные наблюдения обозначить как NA.

### Раздел 1

## Сравнение нескольких моделей по качеству

**Модели с фактором X4**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
| 83 | 2.147871e+13 | NA | NA | NA | NA |
| 81 | 1.358195e+13 | 2 | 7.896758e+12 | 23.54733 | 0 |

## Вывод по сравнению нескольких моделей по качеству

Видно, что для рассматриваемого фактора добавление фиктивных переменных уменьшает остаточную дисперсию (RSS) и тем самым значимо повышает качество моделей (p-значение меньше 0,05)

## Таблица с характеристиками качества моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Ср.ошибка.Аппрокс |
| fit.X4 | 0.000 | 1.03 | 9600.7 |
| fit.X4.fo | 0.352 | 16.23 | 7336.6 |

## Вывод

Судя по таблице с характеристиками качества моделей видно, что лучшей моделью является fit.X4.fo, c фиктивными переменными, так как имеет наибольший скорректированный R-квадрат, а так же наименьшую среднюю ошибку аппроксимации.

### Раздел 2

## Сравнение нескольких моделей по качеству (с исходными данными и прологарифмированными)

**Модели с фактором X4**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
| 83 | 2.147871e+13 | NA | NA | NA | NA |
| 81 | 1.358195e+13 | 2 | 7.896758e+12 | 1.101714e+12 | 0 |
| 83 | 5.031351e+02 | -2 | 1.358195e+13 | NA | NA |
| 76 | 2.723726e+02 | 7 | 2.307625e+02 | 9.198507e+00 | 0 |

## Вывод по сравнению нескольких моделей по качеству

Видно, что для рассматриваемого фактора добавление фиктивных переменных уменьшает остаточную дисперсию (RSS) и тем самым значимо повышает качество моделей (p-значение меньше 0,05)

## Таблица с характеристиками качества моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Ср.ошибка.Аппрокс |
| fit.X4 | 0.000 | 1.03 | 9600.7 |
| fit.X4.fo | 0.352 | 16.23 | 7336.6 |

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Ср.ошибка.Аппрокс |
| fit.log.X4 | 0.030 | 3.64 | 23.8 |
| fit.log.X4.fo | 0.427 | 8.82 | 16.0 |

## Вывод

Судя по таблице с характеристиками качества моделей видно, что лучшими моделями являются fit.X4.fo (с исходными данными) и fit.log.X4.fo (с прологарифмированными данными), c фиктивными переменными, так как имеет наибольший скорректированный R-квадрат, а так же наименьшую среднюю ошибку аппроксимации.