Лабораторная работа №3

Марковский Валентин Александрович

25 12 2020

## Лучшие модели

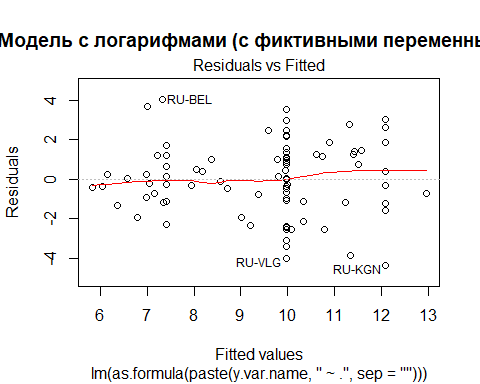
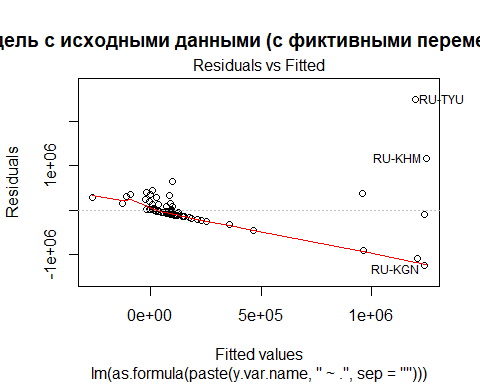
1. ()
2. ()

## 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.

## График остатков “Residuals vs Fitted” для моделей



## Тест на равенство среднего остатков нулю

|  |  |  |
| --- | --- | --- |
| Модель | Статистика | P.значение |
| fit.X4.fo | 0 | 1 |
| fit.log.X4.fo | 0 | 1 |

## Вывод

Для обеих моделей принимается нулевая гипотеза (среднее значение остатков равно 0), так как p-значения моделей (1) больше заданного уровня значимости 0,05.

## Тест Бройша-Пагана на гетероскедастичность

|  |  |  |
| --- | --- | --- |
| Модель | Статистика | P.значение |
| fit.X4.fo | 34.799 | 0.000 |
| fit.log.X4.fo | 7.840 | 0.449 |

## Вывод

Для первой модели нулевая гипотеза отвергается, т.е. остатки являются гетероскедастичными, поскольку p-значение = 0.000 меньше заданного уровня значимости 0.05. Для второй модели нулевая гипотеза принимается, т.е. остатки не являются гетероскедастичными, поскольку p-значение = 0.449 больше заданного уровня значимости.

## Тест Голдфельда-Квандта на гетероскедастичность

|  |  |  |
| --- | --- | --- |
| Модель | Статистика | P.значение |
| fit.X4.fo | 9.380 | 0.000 |
| fit.log.X4.fo | 1.306 | 0.255 |

## Вывод

Для первой модели нулевая гипотеза отвергается, т.е. остатки являются гетероскедастичными, поскольку p-значение = 0.000 меньше заданного уровня значимости 0.05. Для второй модели нулевая гипотеза принимается, т.е. остатки не являются гетероскедастичными, поскольку p-значение = 0.255 больше заданного уровня значимости.

## Тест Дабрина-Уотсона на автокрреляцию

|  |  |  |
| --- | --- | --- |
| Модель | Статистика | P.значение |
| fit.X4.fo | 1.605 | 0.051 |
| fit.log.X4.fo | 1.828 | 0.170 |

## Вывод

Для обеех моделей нулевая гипотеза принимается, т.е. коэффициент автокорреляции незначим, поскольку p-значение (0.051 и 0.170 соответственно) больше заданного уровня значимости 0.05. Так как условия Гаусса-Маркова не выполнены ни для одной модели (по тесту Бройша-Пагана и Голдфельда-Квандта), то возьмём лучшее уравнение регрессии fit.log.X4.fo (скорректированный R-квадрат = 0,4264) и переоценим его параметры с учётом ошибок спецификации.

## Переоценка коэффициентов регрессии

##   
## t test of coefficients:  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.96974 0.44294 22.5082 < 2.2e-16 \*\*\*  
## foПФО -24.68386 3.10212 -7.9571 1.330e-11 \*\*\*  
## foСКФО -2.54793 0.69233 -3.6802 0.0004332 \*\*\*  
## foСФО -4.85498 1.07908 -4.4992 2.414e-05 \*\*\*  
## foУФО 2.12554 1.26352 1.6822 0.0966279 .   
## foЦФО -11.31815 3.00466 -3.7669 0.0003242 \*\*\*  
## X4.foПФО 2.82558 0.37351 7.5650 7.458e-11 \*\*\*  
## X4.foСФО 0.71830 0.14879 4.8276 6.984e-06 \*\*\*  
## X4.foЦФО 1.12743 0.35073 3.2146 0.0019185 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Вывод

После переоценки лучшей модели все параметры остались значимыми.