Лабораторная работа № 2

Ямпольский Антон

27 12 2020

##Импорт данных Импортируем объекты, сохраненные в рабочем пространстве по итогу ЛР№1

## [1] "coef.vars" "coef.vars1" "DF" "DF1" "file.path"   
## [6] "i" "matrix.cor" "matrix.p" "mns" "mns1"   
## [11] "p" "pic.num" "r.corr" "reg.df" "sds"   
## [16] "sds1" "smm" "smm1" "table" "table.num"   
## [21] "table1" "W" "x"

## [1] 83 6

## FO IPI.2013 PIM.2013 DDFA.2013 FCI.2012 DLR.2013  
## 3 ЦФО 102.1 96.0 43.3 88929 240495  
## 4 ЦФО 96.8 106.6 45.6 36974 50809  
## 5 ЦФО 108.0 99.8 43.7 42761 69259  
## 6 ЦФО 106.1 101.6 43.0 78223 209873  
## 7 ЦФО 106.2 103.4 42.2 27353 41361  
## 8 ЦФО 105.9 103.1 35.8 95312 84824

## 'data.frame': 83 obs. of 6 variables:  
## $ FO : Factor w/ 8 levels "ДФО","ПФО","СЗФО",..: 7 7 7 7 7 7 7 7 7 7 ...  
## $ IPI.2013 : num 102.1 96.8 108 106.1 106.2 ...  
## $ PIM.2013 : num 96 106.6 99.8 101.6 103.4 ...  
## $ DDFA.2013: num 43.3 45.6 43.7 43 42.2 35.8 48.4 47.9 49.2 40.5 ...  
## $ FCI.2012 : int 88929 36974 42761 78223 27353 95312 32059 59477 80173 73822 ...  
## $ DLR.2013 : int 240495 50809 69259 209873 41361 84824 33399 143835 98075 790946 ...

# Раздел I.

## Изначальная регрессионная модель, основанная на ЛР№1

Модель 0: , где

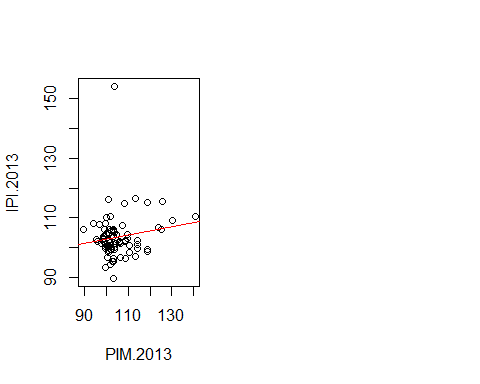
* Y *IPI.2013* – Индексы промышленного производства.
* X1 *PIM.2013* – Индесы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства.
* X2 *DDFA.2013* – Степень износа основных фондов.
* X3 *FCI.2012* – Инвестиции в основвной капитал на душу населения.
* X4 *DLR.2013* – Задолжность по кредитам в рублях, предоставленым кредитными организациями юридическим лицам.

По количеству 83-x наблюдений.

## Оценка параметров этой моделей

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 89.0251 | 10.4829 | 8.4924 | 0.0000 |
| PIM.2013 | 0.1387 | 0.0994 | 1.3957 | 0.1666 |



#### Рис. 2. график разброса начальной модели

**Проверка значимости для коэффициента при PIM.2013.**

H0: (параметр) коэфф. при PIM.2013 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при PIM.2013 не равен 0 в генеральной совокупности (значим).

Проверим значимость при помощи p-значения.

**Напоминание:** *Сравниваем p-значение и (Уровень значимости = 0,05);* *Если p-значение > , то принимается гипотеза H0, в ином случае принимается противоположная гипотеза H1.*

P-значение при PIM.2014 = => принимается гипотеза H0. **Параметр незначим.**

## Пошаговое исключение регрессоров

Исключаем DDFA.2013 вторым так как у него самое большое p-значение (0.708) Исключаем FCI.2012 первым так как у него самое большое p-значение (0.598) Исключаем DLR.2013 третьим так как у него самое большое p-значение (0.224)

Явный вид модели 1:

#### Рис. 2. график разброса исправленной модели

## модель с переменной структурой по федеральным округам

Построим модель с переменной структурой, используя принадлежность каждого региона к одному из восьми федеральных округов. Включим фиктивные переменные как в константу, так и в коэффициенты. Общий вид модели с переменной структурой.

#### Таблица 2 - описательные статистики модели по федеральным округам

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 89.1514 | 45.3782 | 1.9646 | 0.0536 |
| PIM.2013 | 0.1320 | 0.3954 | 0.3338 | 0.7396 |
| FOПФО | 3.8515 | 100.8740 | 0.0382 | 0.9697 |
| FOСЗФО | 66.0239 | 133.8435 | 0.4933 | 0.6234 |
| FOСКФО | -306.9909 | 119.9295 | -2.5598 | 0.0127 |
| FOСФО | -2.6671 | 50.1594 | -0.0532 | 0.9578 |
| FOУФО | 2.1967 | 54.4481 | 0.0403 | 0.9679 |
| FOЦФО | 48.3428 | 90.6178 | 0.5335 | 0.5955 |
| FOЮФО | -18.3854 | 54.8022 | -0.3355 | 0.7383 |
| PIM.2013:FOПФО | -0.0503 | 0.9658 | -0.0521 | 0.9586 |
| PIM.2013:FOСЗФО | -0.6674 | 1.2958 | -0.5151 | 0.6082 |
| PIM.2013:FOСКФО | 3.1350 | 1.1760 | 2.6657 | 0.0096 |
| PIM.2013:FOСФО | 0.0242 | 0.4376 | 0.0552 | 0.9561 |
| PIM.2013:FOУФО | -0.0129 | 0.4876 | -0.0264 | 0.9790 |
| PIM.2013:FOЦФО | -0.4695 | 0.8691 | -0.5402 | 0.5909 |
| PIM.2013:FOЮФО | 0.2070 | 0.4923 | 0.4205 | 0.6755 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё немного выше, чем у модели по всем регионам (23.1%). У неё несколько незначимых параметров. Исключать их последовательно вручную трудоёмко, поэтому мы воспользуемся пользовательской функцией, которая проводит процедуру последовательного исключения регрессоров.

Сначала сгенерируем матрицу независимых переменных функцией *model.matrix()*. После загружаем функцию для исключения незначимых регрессоров из файла «removeFactorsByPValue.R» в рабочей директории и применяем её к модели с переменной структурой.

### Модель без поправки:

#### Таблица 3 - описательные статистики модели по федеральным округам без поправки

## IPI.2013 FOПФО FOСЗФО FOСКФО FOСФО FOУФО FOЦФО FOЮФО PIM.2013 FOПФО.PIM.2013  
## 3 102.1 0 0 0 0 0 1 0 96.0 0  
## 4 96.8 0 0 0 0 0 1 0 106.6 0  
## 5 108.0 0 0 0 0 0 1 0 99.8 0  
## 6 106.1 0 0 0 0 0 1 0 101.6 0  
## 7 106.2 0 0 0 0 0 1 0 103.4 0  
## 8 105.9 0 0 0 0 0 1 0 103.1 0  
## FOСЗФО.PIM.2013 FOСКФО.PIM.2013 FOСФО.PIM.2013 FOУФО.PIM.2013 FOЦФО.PIM.2013  
## 3 0 0 0 0 96.0  
## 4 0 0 0 0 106.6  
## 5 0 0 0 0 99.8  
## 6 0 0 0 0 101.6  
## 7 0 0 0 0 103.4  
## 8 0 0 0 0 103.1  
## FOЮФО.PIM.2013  
## 3 0  
## 4 0  
## 5 0  
## 6 0  
## 7 0  
## 8 0

## IPI.2013 FOПФО FOСЗФО FOСКФО FOСФО FOУФО FOЦФО FOЮФО PIM.2013 FOПФО.PIM.2013  
## 87 102.2 0 0 0 0 0 0 0 108.6 0  
## 88 106.9 0 0 0 0 0 0 0 123.9 0  
## 89 103.0 0 0 0 0 0 0 0 109.7 0  
## 90 99.5 0 0 0 0 0 0 0 119.0 0  
## 91 102.6 0 0 0 0 0 0 0 108.7 0  
## 92 116.6 0 0 0 0 0 0 0 113.3 0  
## FOСЗФО.PIM.2013 FOСКФО.PIM.2013 FOСФО.PIM.2013 FOУФО.PIM.2013 FOЦФО.PIM.2013  
## 87 0 0 0 0 0  
## 88 0 0 0 0 0  
## 89 0 0 0 0 0  
## 90 0 0 0 0 0  
## 91 0 0 0 0 0  
## 92 0 0 0 0 0  
## FOЮФО.PIM.2013  
## 87 0  
## 88 0  
## 89 0  
## 90 0  
## 91 0  
## 92 0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.0671 | 0.8031 | 128.3314 | 0.0000 |
| FOСКФО | -320.9066 | 106.3810 | -3.0166 | 0.0034 |
| FOСКФО.PIM.2013 | 3.2670 | 1.0613 | 3.0782 | 0.0029 |

Все коэффициенты модели значимы, но она имеет низкий уровень коэффициента детерминации. ( 0.157)

Значима константа для Северо-Кавказского федерального округа, а также коэффициент при независимых переменных для некоторых округов.

*( PIM.2013 с Северно-Кавказскими федеральным округом;*

### Модель с поправкой Бонферрони:

Явный вид модели 2:

#### Таблица 4 - описательные статистики модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.0671 | 0.8031 | 128.3314 | 0.0000 |
| FOСКФО | -320.9066 | 106.3810 | -3.0166 | 0.0034 |
| FOСКФО.PIM.2013 | 3.2670 | 1.0613 | 3.0782 | 0.0029 |

Коэффициенты модели при *PIM.2013* значимы, однако коэффициент детерминации не изменился ( 0.157).

# строим ПЛР на второй по силе корреляции фактор

Построим модель с переменной структурой, используя принадлежность каждого региона к одному из восьми федеральных округов. Включим фиктивные переменные как в константу, так и в коэффициенты. Общий вид модели с переменной структурой.

Модель со 2 фактором по силе корреляции не значима, и имеет очень слабый коэффицент детерминации ( 0.02)

#### Таблица 5 - описательные статистики модели по федеральным округам

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 104.8812 | 3.7465 | 27.9945 | 0.0000 |
| FOПФО | -4.5571 | 5.1141 | -0.8911 | 0.3761 |
| FOСЗФО | -4.1445 | 4.5739 | -0.9061 | 0.3681 |
| FOСКФО | 7.0358 | 5.6093 | 1.2543 | 0.2141 |
| FOСФО | 0.4640 | 5.1078 | 0.0908 | 0.9279 |
| FOУФО | -0.0429 | 6.4702 | -0.0066 | 0.9947 |
| FOЦФО | -0.9382 | 4.2155 | -0.2226 | 0.8246 |
| FOЮФО | 2.9296 | 5.6884 | 0.5150 | 0.6082 |
| DLR.2013 | 0.0000 | 0.0001 | -0.2227 | 0.8245 |
| FOПФО:DLR.2013 | 0.0000 | 0.0001 | 0.3252 | 0.7461 |
| FOСЗФО:DLR.2013 | 0.0000 | 0.0001 | 0.2006 | 0.8416 |
| FOСКФО:DLR.2013 | -0.0001 | 0.0001 | -0.5382 | 0.5922 |
| FOСФО:DLR.2013 | 0.0000 | 0.0001 | 0.0200 | 0.9841 |
| FOУФО:DLR.2013 | 0.0000 | 0.0001 | 0.1140 | 0.9096 |
| FOЦФО:DLR.2013 | 0.0000 | 0.0001 | 0.1983 | 0.8434 |
| FOЮФО:DLR.2013 | 0.0000 | 0.0001 | 0.0526 | 0.9582 |

Модель в целом незначима, и скорректированный коэффициент детерминации у неё ниже, чем у модели по всем регионам (15.7%). У неё много незначимых параметров. Исключать их последовательно вручную трудоёмко, поэтому мы воспользуемся пользовательской функцией, которая проводит процедуру последовательного исключения регрессоров.

Сначала сгенерируем матрицу независимых переменных функцией *model.matrix()*. После загружаем функцию для исключения незначимых регрессоров из файла «removeFactorsByPValue.R» в рабочей директории и применяем её к модели с переменной структурой

### Модель без поправки:

#### Таблица 6 - описательные статистики модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.0671 | 0.8441 | 122.1022 | 0.0000 |
| FOСКФО | 6.4472 | 2.9066 | 2.2181 | 0.0293 |

Все коэффициенты модели значимы и она имеет малый уровень коэффициента детерминации. ( 0.057)

### Модель с поправкой Бонферрони:

#### Таблица 7 - описательные статистики модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.0671 | 0.8441 | 122.1022 | 0.0000 |
| FOСКФО | 6.4472 | 2.9066 | 2.2181 | 0.0293 |

Коэффициент модели значим, однако коэффициент детерминации остался очень слабым ( 0.057).

## Сравнение моделей по качеству.

Сравним три полученные модели: изначальную, с поправкой по ФО и без поправки по ФО.

#### Таблица 8 - сравнение трёх моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.1 | 0.011 | 1.95 | 7.5 |
| fit.1.foBonferroni | 0.136 | 7.46 | 7.0 |
| fit.1.fo | 0.136 | 7.46 | 7.0 |

Результат:

Среднее по Y = 103.6108;

По столбцу больше всего подходит вторая и третья модель; По столбцу F.расч - вторая и третья; По минимальной Стандартной ошибке -вторая и третья.

Таким образом, модель по федеральным округам без поправки (fit.1.fo) наиболее предпочтительна.

Явный вид модели 4:

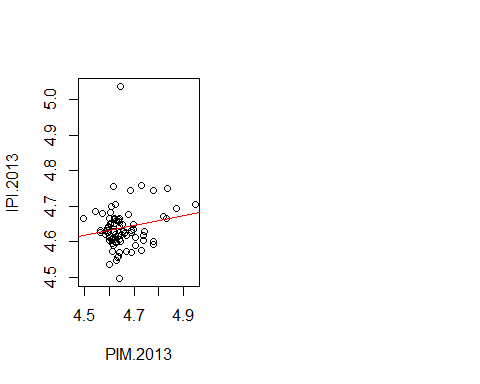
# Раздел II

## Изначаьная регрессионная модель для логарифмированных данных, основанная на ЛР№1

Модель 0: , где

* Y *IPI.2013* – Индексы промышленного производства.
* X1 *PIM.2013* – Индесы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства.
* X2 *DDFA.2013* – Степень износа основных фондов.
* X3 *FCI.2012* – Инвестиции в основвной капитал на душу населения.
* X4 *DLR.2013* – Задолжность по кредитам в рублях, предоставленым кредитными организациями юридическим лицам. По количеству 83-x наблюдений.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 3.9966 | 0.4489 | 8.9023 | 0.0000 |
| PIM.2013 | 0.1379 | 0.0965 | 1.4297 | 0.1566 |

 #### Рис. 3. график разброса начальной логарифмированной модели

## Проверка значимости для логарифмированных значений:

**Проверка значимости для коэффициента при PIM.2013.**

H0: (параметр) коэфф. при PIM.2013 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при PIM.2013 не равен 0 в генеральной совокупности (значим).

Проверим значимость при помощи p-значения. ( )

P-значение при PIM.2013 = => принимается гипотеза H0. **Параметр незначим.**

# Модель с переменной структурой по федеральным округам (логарифмированные данные).

Построим модель с переменной структурой, используя принадлежность каждого региона к одному из восьми федеральных округов. Включим фиктивные переменные как в константу, так и в коэффициенты. Общий вид модели с переменной структурой.

#### Таблица 10 - описательные статистики логарифмированной модели по федеральным округам

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 3.9420 | 1.9511 | 2.0204 | 0.0473 |
| FOПФО | 0.3027 | 4.2089 | 0.0719 | 0.9429 |
| FOСЗФО | 3.1538 | 5.5565 | 0.5676 | 0.5722 |
| FOСКФО | -10.0099 | 4.9787 | -2.0106 | 0.0484 |
| FOСФО | -0.0879 | 2.1664 | -0.0406 | 0.9677 |
| FOУФО | 0.2829 | 2.3057 | 0.1227 | 0.9027 |
| FOЦФО | 2.2569 | 3.7766 | 0.5976 | 0.5521 |
| FOЮФО | -0.8028 | 2.3706 | -0.3387 | 0.7359 |
| PIM.2013 | 0.1485 | 0.4116 | 0.3608 | 0.7194 |
| FOПФО:PIM.2013 | -0.0677 | 0.9050 | -0.0748 | 0.9406 |
| FOСЗФО:PIM.2013 | -0.6860 | 1.1978 | -0.5727 | 0.5688 |
| FOСКФО:PIM.2013 | 2.1849 | 1.0761 | 2.0304 | 0.0463 |
| FOСФО:PIM.2013 | 0.0189 | 0.4572 | 0.0413 | 0.9672 |
| FOУФО:PIM.2013 | -0.0589 | 0.4891 | -0.1203 | 0.9046 |
| FOЦФО:PIM.2013 | -0.4867 | 0.8122 | -0.5993 | 0.5510 |
| FOЮФО:PIM.2013 | 0.1799 | 0.5035 | 0.3572 | 0.7221 |

Модель в целом значима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (19.1%). У неё несколько незначимых параметров. Исключать их последовательно вручную трудоёмко, поэтому мы воспользуемся пользовательской функцией, которая проводит процедуру последовательного исключения регрессоров.

Сначала сгенерируем матрицу независимых переменных функцией *model.matrix()*. После загружаем функцию для исключения незначимых регрессоров из файла «removeFactorsByPValue.R» в рабочей директории и применяем её к модели с переменной структурой.

### Модель без поправки:

#### Таблица 11 - описательные статистики логарифмированной модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.6344 | 0.0072 | 643.3334 | 0.0000 |
| FOСКФО | -10.7023 | 4.4059 | -2.4291 | 0.0174 |
| FOСКФО.PIM.2013 | 2.3334 | 0.9564 | 2.4398 | 0.0169 |

Все коэффициенты модели значимы и она имеет слабый уровень коэффициента детерминации. 0.107.

Значимы константы для Северо-Кавказского федеральных округов, коэффициент при PIM.2013, а также коэффициенты при независимых переменных для некоторых округов.

*(PIM.2013 с Северо-Кавказским федеральным округом;)*

### Модель с поправкой Бонферрони:

Явный вид модели 3: .

#### Таблица 12 - описательные статистики логарифмированной модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.6344 | 0.0072 | 643.3334 | 0.0000 |
| FOСКФО | -10.7023 | 4.4059 | -2.4291 | 0.0174 |
| FOСКФО.PIM.2013 | 2.3334 | 0.9564 | 2.4398 | 0.0169 |

Коэффициент модели при *PIM.2013* значим, однако коэффициент детерминации заметно понизился ( 0.107).

Сравним три полученные модели: изначальную, с поправкой по ФО и без поправки по ФО.

#### Таблица 13 - сравнение трёх моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.11 | 0.013 | 2.04 | 0.1 |
| fit.11.foBonferroni | 0.084 | 4.78 | 0.1 |
| fit.11.fo | 0.084 | 4.78 | 0.1 |

## [1] 4.638374

Результат:

Среднее по Y = 4.63874;

По столбцу больше всего подходит вторая и третья; По столбцу F.расч - вторая и третья; По минимальной Стандартной ошибке - вторая и третья, из-за того что у обеих моделей она приблизительно похожа, выбор наилучшей становится сложнее.

Выберем третью модель. **Явный вид модели: ..**