Лабораторная работа №2

Кравчук Д. А.

27 12 2020

## Импорт данных

Импортируем объекты, сохранённые в рабочем пространстве по итогу лабораторной №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"

# Раздел I.

## Изначальная регрессионная модель, основанная на Лабораторной №1

Модель 0: , где

* Y (*IndProd*) – Индексы промышленного производства.
* X1 (*PromPrices*) – Индексы цен производителей промышленных товаров по видам экономической деятельности: обрабатывающие производства.
* X2 (*FundsDeter*) – Степень износа основных фондов.
* X3 (*CapInvests*) – Инвестиции в основной капитал на душу населения
* X4 (*CreditsIndebted*) – Задолженность по кредитам в рублях, предоставленным кредитными организациями юридическим лицам

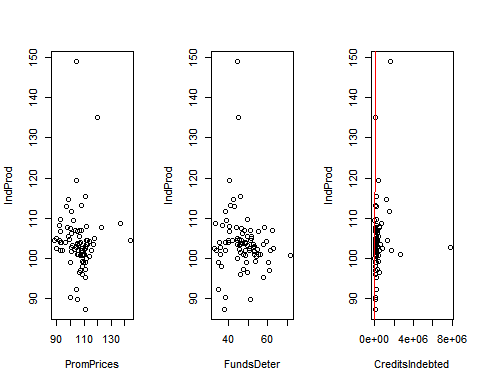
По количеству 83-x наблюдений.

## Оценка параметров этой модели

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 107.9896 | 11.4937 | 9.3956 | 0.0000 |
| PromPrices | 0.0050 | 0.0983 | 0.0505 | 0.9599 |
| FundsDeter | -0.0989 | 0.1128 | -0.8764 | 0.3835 |
| CreditsIndebted | 0.0000 | 0.0000 | 0.9002 | 0.3708 |

## Warning in abline(reg, col = "red"): only using the first two of 4 regression  
## coefficients



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

**Проверка значимости для коэффициента при PromPrices.**

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

H1: (параметр) коэфф. при PromPrices 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 PromPrices = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при FundsDeter.**

P-значение при FundsDeter = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при CreditsIndebted.**

P-значение при CreditsIndebted = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.022. 97% исходного разброса зависимой переменной Y (ВРП) объясняет разброс объясняющих переменных X1 (Инвестиции в основной капитал), X2 (Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику) и X4 (Внутренние затраты на научные исследования и разработки).

Явный вид модели 1: .

## 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) | -38.0394 | 131.2337 | -0.2899 | 0.7731 |
| FOПФО | 200.3180 | 137.2093 | 1.4599 | 0.1504 |
| FOСЗФО | 142.6570 | 132.6185 | 1.0757 | 0.2871 |
| FOСКФО | 29.3984 | 135.9511 | 0.2162 | 0.8297 |
| FOСФО | 182.7839 | 140.9355 | 1.2969 | 0.2005 |
| FOУФО | 108.6023 | 140.1149 | 0.7751 | 0.4419 |
| FOЦФО | 154.2140 | 132.9158 | 1.1602 | 0.2514 |
| FOЮФО | 151.3460 | 141.3051 | 1.0711 | 0.2892 |
| PromPrices | 1.3469 | 1.3006 | 1.0356 | 0.3053 |
| FundsDeter | -0.1038 | 0.3636 | -0.2854 | 0.7765 |
| CreditsIndebted | 0.0000 | 0.0000 | -0.9435 | 0.3499 |
| FOПФО:PromPrices | -1.7515 | 1.3506 | -1.2968 | 0.2005 |
| FOСЗФО:PromPrices | -1.2874 | 1.3092 | -0.9833 | 0.3301 |
| FOСКФО:PromPrices | -0.0875 | 1.3335 | -0.0656 | 0.9479 |
| FOСФО:PromPrices | -1.6910 | 1.3896 | -1.2169 | 0.2292 |
| FOУФО:PromPrices | -1.0985 | 1.3682 | -0.8028 | 0.4258 |
| FOЦФО:PromPrices | -1.1460 | 1.3261 | -0.8642 | 0.3915 |
| FOЮФО:PromPrices | -1.3565 | 1.3208 | -1.0270 | 0.3093 |
| FOПФО:FundsDeter | -0.1907 | 0.5122 | -0.3723 | 0.7112 |
| FOСЗФО:FundsDeter | -0.1190 | 0.4390 | -0.2710 | 0.7875 |
| FOСКФО:FundsDeter | -0.2802 | 0.5864 | -0.4778 | 0.6349 |
| FOСФО:FundsDeter | -0.1310 | 0.4519 | -0.2899 | 0.7731 |
| FOУФО:FundsDeter | 0.1851 | 0.7680 | 0.2410 | 0.8105 |
| FOЦФО:FundsDeter | -0.5349 | 0.5371 | -0.9960 | 0.3240 |
| FOЮФО:FundsDeter | -0.0518 | 0.8044 | -0.0644 | 0.9489 |
| FOПФО:CreditsIndebted | 0.0000 | 0.0000 | 0.9158 | 0.3641 |
| FOСЗФО:CreditsIndebted | 0.0001 | 0.0000 | 1.4182 | 0.1622 |
| FOСКФО:CreditsIndebted | 0.0000 | 0.0000 | 1.4978 | 0.1404 |
| FOСФО:CreditsIndebted | 0.0001 | 0.0000 | 2.0173 | 0.0489 |
| FOУФО:CreditsIndebted | 0.0000 | 0.0000 | 0.9151 | 0.3645 |
| FOЦФО:CreditsIndebted | 0.0000 | 0.0000 | 0.8944 | 0.3753 |
| FOЮФО:CreditsIndebted | 0.0000 | 0.0000 | 0.9807 | 0.3313 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (62.7%). 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:

Явный вид модели 2: .

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 5.3397 | 23.8305 | 0.2241 | 0.8234 |
| FOПФО | 147.5151 | 44.0940 | 3.3455 | 0.0013 |
| FOСЗФО | 98.2747 | 27.9394 | 3.5174 | 0.0008 |
| FOСКФО | 10.9205 | 3.1271 | 3.4923 | 0.0008 |
| FOСФО | 144.3127 | 55.2017 | 2.6143 | 0.0110 |
| FOЦФО | 104.6970 | 31.2408 | 3.3513 | 0.0013 |
| FOЮФО | 96.0530 | 30.0540 | 3.1960 | 0.0021 |
| PromPrices | 0.8766 | 0.2166 | 4.0467 | 0.0001 |
| FOПФО.PromPrices | -1.3444 | 0.4065 | -3.3073 | 0.0015 |
| FOСЗФО.PromPrices | -0.8774 | 0.2551 | -3.4392 | 0.0010 |
| FOСФО.PromPrices | -1.3541 | 0.5022 | -2.6961 | 0.0088 |
| FOЦФО.PromPrices | -0.9246 | 0.2970 | -3.1131 | 0.0027 |
| FOЮФО.PromPrices | -0.8285 | 0.2742 | -3.0216 | 0.0035 |
| FOСФО.CreditsIndebted | 0.0000 | 0.0000 | 6.3147 | 0.0000 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.521)

### Модель с поправкой Бонферрони:

Явный вид модели 3: $IndProd = 78240.44 + 1.9575 PromPrices + 1.6602 FOПФО.PromPrices - 2.2150 FOЦФО.PromPrices + 15.7997 FOУФО.FundsDeter + 20.7943 FOЦФО.FundsDeter + 20.1379 FOСЗФО.CreditsIndebted + 31.8390 FOСФО.CreditsIndebted + 19.7434 FOЦФО.CreditsIndebted + 41.2278 FOЮФО.CreditsIndebted $.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.6410 | 0.7275 | 142.4578 | 0.0000 |
| FOСКФО | -129.6884 | 28.3082 | -4.5813 | 0.0000 |
| FOСКФО.PromPrices | 1.2867 | 0.2735 | 4.7052 | 0.0000 |
| FOСФО.FundsDeter | -0.1602 | 0.0510 | -3.1389 | 0.0024 |
| FOСФО.CreditsIndebted | 0.0000 | 0.0000 | 7.2883 | 0.0000 |

Коэффициенты модели значимы, однако коэффициент детерминации слегка понизился ( 0.5).

## Сравнение моделей по качеству.

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

#### Таблица 5 - сравнение трёх моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.1 | -0.015 | 0.60 | 8.1 |
| fit.1.foBonferroni | 0.475 | 19.52 | 5.8 |
| fit.1.fo | 0.431 | 5.77 | 6.1 |

Результат:

Среднее по Y = 653467.5;

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

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

**Явный вид модели 3: .**

# Раздел II.

## Изначальная регрессионная модель для логарифмированных данных, основанная на Лабораторной №1

Модель 0: , где

* Y (*IndProd*) – ВРП.
* X1 (*PromPrices*) – Инвестиции в основной капитал.
* X2 (*FundsDeter*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику.
* X3 (*CapInvests*) – Использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* X4 (*CreditsIndebted*) – Внутренние затраты на научные исследования и разработки.

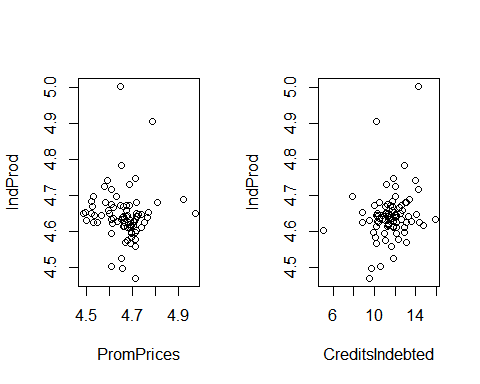
По количеству 83-x наблюдений.

## Оценка параметров этой модели

#### Таблица 6 - описательные статистики логарифмированной модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.7030 | 0.4324 | 10.8755 | 0.0000 |
| PromPrices | -0.0394 | 0.0933 | -0.4218 | 0.6743 |
| CreditsIndebted | 0.0107 | 0.0049 | 2.1731 | 0.0327 |

## Warning in abline(reg, col = "red"): only using the first two of 4 regression  
## coefficients



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

## Проверка значимости для логарифмированных значений:

**Проверка значимости для коэффициента при PromPrices.**

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

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

Проверим значимость при помощи p-значения. ( )

P-значение при PromPrices = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при CreditsIndebted.**

P-значение при CreditsIndebted = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.056. 94% исходного разброса зависимой переменной Y (ВРП) объясняет разброс объясняющих переменных X1 (Инвестиции в основной капитал), X4 (Внутренние затраты на научные исследования и разработки).

Явный вид модели 1: .

## 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 7 - описательные статистики логарифмированной модели по федеральным округам

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.9364 | 4.5583 | 0.2054 | 0.8379 |
| FOПФО | 5.8449 | 4.8975 | 1.1934 | 0.2375 |
| FOСЗФО | 3.4815 | 4.6377 | 0.7507 | 0.4558 |
| FOСКФО | -1.3369 | 4.8106 | -0.2779 | 0.7821 |
| FOСФО | 9.9760 | 5.1627 | 1.9323 | 0.0581 |
| FOУФО | 2.4286 | 5.1095 | 0.4753 | 0.6363 |
| FOЦФО | 3.8984 | 4.6744 | 0.8340 | 0.4077 |
| FOЮФО | 3.7291 | 4.6856 | 0.7959 | 0.4293 |
| PromPrices | 0.8291 | 1.0110 | 0.8201 | 0.4154 |
| CreditsIndebted | -0.0186 | 0.0289 | -0.6420 | 0.5234 |
| FOПФО:PromPrices | -1.2939 | 1.0825 | -1.1952 | 0.2368 |
| FOСЗФО:PromPrices | -0.8011 | 1.0260 | -0.7808 | 0.4380 |
| FOСКФО:PromPrices | 0.2248 | 1.0668 | 0.2107 | 0.8338 |
| FOСФО:PromPrices | -2.2449 | 1.1364 | -1.9755 | 0.0529 |
| FOУФО:PromPrices | -0.5625 | 1.1303 | -0.4976 | 0.6206 |
| FOЦФО:PromPrices | -0.8881 | 1.0361 | -0.8571 | 0.3948 |
| FOЮФО:PromPrices | -0.8471 | 1.0420 | -0.8129 | 0.4195 |
| FOПФО:CreditsIndebted | 0.0204 | 0.0335 | 0.6084 | 0.5453 |
| FOСЗФО:CreditsIndebted | 0.0270 | 0.0308 | 0.8764 | 0.3844 |
| FOСКФО:CreditsIndebted | 0.0352 | 0.0344 | 1.0214 | 0.3112 |
| FOСФО:CreditsIndebted | 0.0519 | 0.0308 | 1.6833 | 0.0976 |
| FOУФО:CreditsIndebted | 0.0195 | 0.0355 | 0.5476 | 0.5860 |
| FOЦФО:CreditsIndebted | 0.0263 | 0.0315 | 0.8371 | 0.4059 |
| FOЮФО:CreditsIndebted | 0.0259 | 0.0329 | 0.7891 | 0.4332 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (38.6%). 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:

Явный вид модели 2:

#### Таблица 8 - описательные статистики логарифмированной модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.8092 | 1.1269 | 0.7181 | 0.4752 |
| FOПФО | 5.9641 | 2.0487 | 2.9112 | 0.0048 |
| FOСЗФО | 3.8693 | 1.3556 | 2.8543 | 0.0057 |
| FOСФО | 10.1033 | 2.5768 | 3.9209 | 0.0002 |
| FOЦФО | 4.0416 | 1.4999 | 2.6946 | 0.0088 |
| FOЮФО | 3.6237 | 1.4574 | 2.4864 | 0.0153 |
| PromPrices | 0.8110 | 0.2400 | 3.3797 | 0.0012 |
| FOПФО.PromPrices | -1.2694 | 0.4373 | -2.9024 | 0.0050 |
| FOСЗФО.PromPrices | -0.8195 | 0.2891 | -2.8341 | 0.0060 |
| FOСФО.PromPrices | -2.2268 | 0.5511 | -4.0407 | 0.0001 |
| FOЦФО.PromPrices | -0.8534 | 0.3225 | -2.6465 | 0.0101 |
| FOЮФО.PromPrices | -0.7606 | 0.3109 | -2.4462 | 0.0170 |
| FOСКФО.CreditsIndebted | 0.0089 | 0.0030 | 3.0010 | 0.0037 |
| FOСФО.CreditsIndebted | 0.0333 | 0.0102 | 3.2597 | 0.0017 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.344.

### Модель с поправкой Бонферрони:

Явный вид модели 3: .

#### Таблица 9 - описательные статистики логарифмированной модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.6398 | 0.0078 | 597.5356 | 0.0000 |
| FOСКФО | -5.2294 | 1.4300 | -3.6570 | 0.0005 |
| FOСФО | 6.2727 | 2.2760 | 2.7560 | 0.0073 |
| FOСКФО.PromPrices | 1.1331 | 0.3086 | 3.6722 | 0.0004 |
| FOСФО.PromPrices | -1.4158 | 0.4873 | -2.9055 | 0.0048 |
| FOСФО.CreditsIndebted | 0.0333 | 0.0100 | 3.3188 | 0.0014 |

Коэффициент модели при *FOСКФО.RKBS.2012* значим, однако коэффициент детерминации заметно понизился ( 0.294).

## Сравнение моделей по качеству.

Сравним две полученные модели: изначальную и без поправки по ФО.

#### Таблица 10 - сравнение двух логарифмированных моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.11 | 0.033 | 2.38 | 0.070 |
| fit.11.fo | 0.220 | 2.78 | 0.063 |
| fit.11.foB | 0.475 | 19.52 | 5.835 |

Результат:

Среднее по Y = 12.70829 ;

По столбцу больше всего подходит первая и вторая модель; По столбцу F.расч - первая и вторая; По минимальной Стандартной ошибке - первая и вторая, но при том, что у обеих моделей она приблизительно похожа, выбор наилучшей становится сложнее.

**Явный вид модели:**

Сохраним нужные данные для дальнейших лабораторных и пойдём спатеньки uwu.