Лабораторная работа № 2

Никишов Дмитрий

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 (*GRP.2013*) – ВРП.
* X1 (*IFA.2012*) – Инвестиции в основной капитал.
* X2 (*ERF.2012*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику.
* X3 (*UPC.2012*) – Использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* X4 (*IRDC.2012*) – Внутренние затраты на научные исследования и разработки.

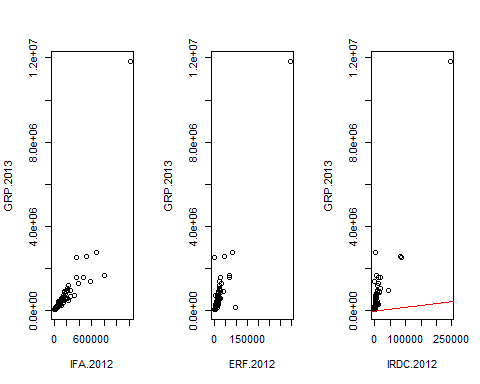
По количеству 81-x наблюдений.

## Оценка параметров этой модели

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -24012.1002 | 33051.9997 | -0.7265 | 0.4697 |
| IFA.2012 | 1.7950 | 0.2096 | 8.5619 | 0.0000 |
| ERF.2012 | 11.2027 | 1.3042 | 8.5896 | 0.0000 |
| IRDC.2012 | 21.9548 | 1.5368 | 14.2857 | 0.0000 |

## 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 IFA.2012.**

H0: (параметр) коэфф. при IFA.2012 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при IFA.2012 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 IFA.2012 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при ERF.2012.**

P-значение при ERF.2012 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при IRDC.2012.**

P-значение при IRDC.2012 = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.975. 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) | -48749.4128 | 67314.8237 | -0.7242 | 0.4724 |
| FOПФО | 108457.1858 | 79888.2079 | 1.3576 | 0.1808 |
| FOСЗФО | 177791.5357 | 91195.0175 | 1.9496 | 0.0570 |
| FOСКФО | 14916.5002 | 107132.2792 | 0.1392 | 0.8898 |
| FOСФО | 45033.4704 | 86899.5193 | 0.5182 | 0.6066 |
| FOУФО | -86688.6774 | 120139.2899 | -0.7216 | 0.4740 |
| FOЦФО | 69217.4728 | 85573.6327 | 0.8089 | 0.4225 |
| FOЮФО | 116619.7192 | 88245.4961 | 1.3215 | 0.1925 |
| IFA.2012 | 2.1224 | 1.0868 | 1.9530 | 0.0565 |
| ERF.2012 | 13.7056 | 13.0274 | 1.0521 | 0.2979 |
| IRDC.2012 | -22.2409 | 33.0293 | -0.6734 | 0.5039 |
| FOПФО:IFA.2012 | 5.2675 | 1.5117 | 3.4845 | 0.0010 |
| FOСЗФО:IFA.2012 | -0.8033 | 1.1459 | -0.7010 | 0.4866 |
| FOСКФО:IFA.2012 | -0.4172 | 2.1893 | -0.1906 | 0.8497 |
| FOСФО:IFA.2012 | -0.8560 | 1.2996 | -0.6587 | 0.5132 |
| FOУФО:IFA.2012 | 0.2138 | 1.1188 | 0.1911 | 0.8492 |
| FOЦФО:IFA.2012 | -1.7213 | 1.2842 | -1.3404 | 0.1863 |
| FOЮФО:IFA.2012 | -1.6871 | 2.4325 | -0.6935 | 0.4912 |
| FOПФО:ERF.2012 | -40.8944 | 14.8785 | -2.7486 | 0.0084 |
| FOСЗФО:ERF.2012 | -14.2716 | 13.0863 | -1.0906 | 0.2808 |
| FOСКФО:ERF.2012 | -9.0689 | 21.3370 | -0.4250 | 0.6727 |
| FOСФО:ERF.2012 | 8.3701 | 16.8804 | 0.4958 | 0.6222 |
| FOУФО:ERF.2012 | 0.1329 | 13.2502 | 0.0100 | 0.9920 |
| FOЦФО:ERF.2012 | 7.4113 | 13.0913 | 0.5661 | 0.5739 |
| FOЮФО:ERF.2012 | 1.8569 | 30.5098 | 0.0609 | 0.9517 |
| FOПФО:IRDC.2012 | 16.4073 | 33.1742 | 0.4946 | 0.6231 |
| FOСЗФО:IRDC.2012 | 44.5821 | 33.0694 | 1.3481 | 0.1838 |
| FOСКФО:IRDC.2012 | 219.2339 | 169.6732 | 1.2921 | 0.2024 |
| FOСФО:IRDC.2012 | 29.7214 | 34.3885 | 0.8643 | 0.3916 |
| FOУФО:IRDC.2012 | 41.3166 | 33.6471 | 1.2279 | 0.2253 |
| FOЦФО:IRDC.2012 | 38.4964 | 33.2592 | 1.1575 | 0.2527 |
| FOЮФО:IRDC.2012 | 69.7271 | 40.1299 | 1.7375 | 0.0886 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (99.7%). 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) | 24825.4640 | 18579.0134 | 1.3362 | 0.1861 |
| IFA.2012 | 1.7319 | 0.1887 | 9.1795 | 0.0000 |
| ERF.2012 | 12.4521 | 1.7850 | 6.9758 | 0.0000 |
| IRDC.2012 | -6.3987 | 2.9721 | -2.1529 | 0.0350 |
| FOПФО.IFA.2012 | 6.0506 | 0.9598 | 6.3043 | 0.0000 |
| FOУФО.IFA.2012 | 0.4731 | 0.1964 | 2.4092 | 0.0188 |
| FOЦФО.IFA.2012 | -1.3805 | 0.3845 | -3.5902 | 0.0006 |
| FOЮФО.IFA.2012 | -1.0176 | 0.2271 | -4.4810 | 0.0000 |
| FOПФО.ERF.2012 | -41.3695 | 7.1481 | -5.7875 | 0.0000 |
| FOСЗФО.ERF.2012 | -11.7535 | 1.8915 | -6.2139 | 0.0000 |
| FOЦФО.ERF.2012 | 8.6405 | 2.1482 | 4.0222 | 0.0002 |
| FOСЗФО.IRDC.2012 | 28.2599 | 3.2401 | 8.7219 | 0.0000 |
| FOСФО.IRDC.2012 | 22.0211 | 5.9086 | 3.7269 | 0.0004 |
| FOУФО.IRDC.2012 | 20.8084 | 5.8824 | 3.5374 | 0.0008 |
| FOЦФО.IRDC.2012 | 22.9172 | 3.8367 | 5.9731 | 0.0000 |
| FOЮФО.IRDC.2012 | 60.7067 | 12.2124 | 4.9709 | 0.0000 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.996)

### Модель с поправкой Бонферрони:

Явный вид модели 3: $GRP.2013 = 78240.44 + 1.9575 IFA.2012 + 1.6602 FOПФО.IFA.2012 - 2.2150 FOЦФО.IFA.2012 + 15.7997 FOУФО.ERF.2012 + 20.7943 FOЦФО.ERF.2012 + 20.1379 FOСЗФО.IRDC.2012 + 31.8390 FOСФО.IRDC.2012 + 19.7434 FOЦФО.IRDC.2012 + 41.2278 FOЮФО.IRDC.2012 $.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 78240.4445 | 19918.0297 | 3.9281 | 0.0002 |
| IFA.2012 | 1.9575 | 0.1297 | 15.0972 | 0.0000 |
| FOПФО.IFA.2012 | 1.6602 | 0.2066 | 8.0342 | 0.0000 |
| FOЦФО.IFA.2012 | -2.2150 | 0.4389 | -5.0471 | 0.0000 |
| FOУФО.ERF.2012 | 15.7997 | 1.5909 | 9.9316 | 0.0000 |
| FOЦФО.ERF.2012 | 20.7943 | 1.5179 | 13.6993 | 0.0000 |
| FOСЗФО.IRDC.2012 | 20.1379 | 1.4778 | 13.6272 | 0.0000 |
| FOСФО.IRDC.2012 | 31.8390 | 5.8408 | 5.4511 | 0.0000 |
| FOЦФО.IRDC.2012 | 19.7434 | 2.9661 | 6.6564 | 0.0000 |
| FOЮФО.IRDC.2012 | 41.2278 | 12.1024 | 3.4066 | 0.0011 |

Коэффициенты модели значимы, однако коэффициент детерминации слегка понизился ( 0.993).

## 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 | 0.974 | 981.00 | 223516.7 |
| fit.1.foBonferroni | 0.993 | 1192.44 | 118179.8 |
| fit.1.fo | 0.995 | 1098.77 | 95490.5 |

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 (*GRP.2013*) – ВРП.
* X1 (*IFA.2012*) – Инвестиции в основной капитал.
* X2 (*ERF.2012*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на национальную экономику.
* X3 (*UPC.2012*) – Использование информационных и коммуникационных технологий в организациях: персональные компьютеры.
* X4 (*IRDC.2012*) – Внутренние затраты на научные исследования и разработки.

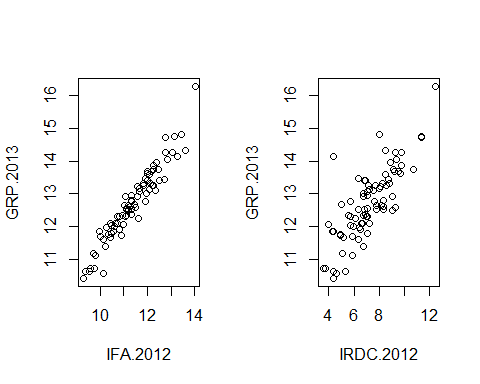
По количеству 81-x наблюдений.

## Оценка параметров этой модели

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 2.1863 | 0.3724 | 5.8712 | 0 |
| IFA.2012 | 0.8454 | 0.0416 | 20.3387 | 0 |
| IRDC.2012 | 0.1249 | 0.0232 | 5.3878 | 0 |

## 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 IFA.2012.**

H0: (параметр) коэфф. при IFA.2012 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при IFA.2012 не равен 0 в генеральной совокупности (значим).

Проверим значимость при помощи p-значения. ( )

P-значение при IFA.2012 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при IRDC.2012.**

P-значение при IRDC.2012 = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.942. 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) | 1.2379 | 1.0995 | 1.1259 | 0.2649 |
| FOПФО | -0.3576 | 1.6477 | -0.2170 | 0.8289 |
| FOСЗФО | 3.5713 | 1.6147 | 2.2117 | 0.0310 |
| FOСКФО | 1.5702 | 1.6130 | 0.9735 | 0.3344 |
| FOСФО | 0.7919 | 1.4239 | 0.5562 | 0.5803 |
| FOУФО | 1.2986 | 1.7114 | 0.7588 | 0.4511 |
| FOЦФО | -0.3118 | 1.5264 | -0.2043 | 0.8388 |
| FOЮФО | 2.7078 | 1.5891 | 1.7040 | 0.0938 |
| IFA.2012 | 0.8890 | 0.1306 | 6.8054 | 0.0000 |
| IRDC.2012 | 0.1548 | 0.0869 | 1.7805 | 0.0803 |
| FOПФО:IFA.2012 | 0.1776 | 0.1887 | 0.9413 | 0.3505 |
| FOСЗФО:IFA.2012 | -0.3327 | 0.1791 | -1.8578 | 0.0684 |
| FOСКФО:IFA.2012 | -0.1720 | 0.1915 | -0.8981 | 0.3729 |
| FOСФО:IFA.2012 | -0.0169 | 0.1700 | -0.0995 | 0.9211 |
| FOУФО:IFA.2012 | -0.0407 | 0.1675 | -0.2431 | 0.8088 |
| FOЦФО:IFA.2012 | 0.1396 | 0.1808 | 0.7722 | 0.4432 |
| FOЮФО:IFA.2012 | -0.3156 | 0.2119 | -1.4896 | 0.1418 |
| FOПФО:IRDC.2012 | -0.1822 | 0.1091 | -1.6704 | 0.1003 |
| FOСЗФО:IRDC.2012 | 0.0591 | 0.1026 | 0.5757 | 0.5671 |
| FOСКФО:IRDC.2012 | 0.0972 | 0.1432 | 0.6788 | 0.5000 |
| FOСФО:IRDC.2012 | -0.0478 | 0.1084 | -0.4409 | 0.6610 |
| FOУФО:IRDC.2012 | -0.0596 | 0.1013 | -0.5884 | 0.5586 |
| FOЦФО:IRDC.2012 | -0.1259 | 0.1051 | -1.1973 | 0.2362 |
| FOЮФО:IRDC.2012 | 0.1549 | 0.1572 | 0.9853 | 0.3286 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (96.4%). 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) | 2.1863 | 0.3724 | 5.8712 | 0 |
| IFA.2012 | 0.8454 | 0.0416 | 20.3387 | 0 |
| IRDC.2012 | 0.1249 | 0.0232 | 5.3878 | 0 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.942.

### 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) | 2.1863 | 0.3724 | 5.8712 | 0 |
| IFA.2012 | 0.8454 | 0.0416 | 20.3387 | 0 |
| IRDC.2012 | 0.1249 | 0.0232 | 5.3878 | 0 |

Коэффициент модели при *FOСКФО.RKBS.2012* значим, однако коэффициент детерминации заметно понизился ( 0.942).

## 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 | 0.940 | 628.15 | 0.268 |
| fit.11.fo | 0.940 | 628.15 | 0.268 |
| fit.11.foB | 0.993 | 1192.44 | 118179.789 |

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.