Лабораторная работа №2

Кравчук Д. А.

28 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

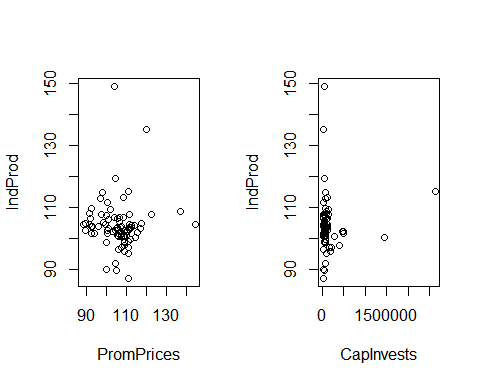
* Y (*IndProd*) – Среднедушевые денежные доходы населения;
* X1 (*PromPrices*) – ВРП на душу населения;
* X3 (*CapInvests*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику;

По количеству 83-x наблюдений.

## Оценка параметров этой модели

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 104.8177 | 10.4946 | 9.9878 | 0.0000 |
| PromPrices | -0.0088 | 0.0990 | -0.0893 | 0.9291 |
| CapInvests | 0.0000 | 0.0000 | 0.6862 | 0.4946 |



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

**Проверка значимости для коэффициента при PromPrices**

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 CapInvests.**

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

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.006.

## 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) | 60.6910 | 98.6413 | 0.6153 | 0.5407 |
| FOПФО | 104.1355 | 110.0311 | 0.9464 | 0.3478 |
| FOСЗФО | 43.1538 | 100.3699 | 0.4299 | 0.6688 |
| FOСКФО | -108.6294 | 107.4247 | -1.0112 | 0.3160 |
| FOСФО | 180.8256 | 116.5728 | 1.5512 | 0.1262 |
| FOУФО | 11.0586 | 113.0202 | 0.0978 | 0.9224 |
| FOЦФО | 49.3562 | 101.9316 | 0.4842 | 0.6300 |
| FOЮФО | 41.0370 | 101.3575 | 0.4049 | 0.6870 |
| PromPrices | 0.3694 | 0.8962 | 0.4122 | 0.6817 |
| CapInvests | 0.0000 | 0.0000 | -0.0409 | 0.9675 |
| FOПФО:PromPrices | -0.8991 | 0.9993 | -0.8998 | 0.3719 |
| FOСЗФО:PromPrices | -0.3877 | 0.9125 | -0.4248 | 0.6725 |
| FOСКФО:PromPrices | 0.9896 | 0.9688 | 1.0214 | 0.3112 |
| FOСФО:PromPrices | -1.6422 | 1.0564 | -1.5546 | 0.1254 |
| FOУФО:PromPrices | -0.0804 | 1.0264 | -0.0783 | 0.9379 |
| FOЦФО:PromPrices | -0.4224 | 0.9341 | -0.4522 | 0.6528 |
| FOЮФО:PromPrices | -0.3025 | 0.9229 | -0.3278 | 0.7442 |
| FOПФО:CapInvests | -0.0001 | 0.0001 | -0.9394 | 0.3513 |
| FOСЗФО:CapInvests | 0.0000 | 0.0000 | 0.3262 | 0.7454 |
| FOСКФО:CapInvests | 0.0003 | 0.0003 | 1.0867 | 0.2816 |
| FOСФО:CapInvests | 0.0000 | 0.0001 | 0.5232 | 0.6028 |
| FOУФО:CapInvests | 0.0000 | 0.0000 | -0.0989 | 0.9215 |
| FOЦФО:CapInvests | 0.0000 | 0.0001 | 0.1054 | 0.9164 |
| FOЮФО:CapInvests | 0.0000 | 0.0001 | -0.3323 | 0.7408 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (33.7%). 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 3 - описательные статистики модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 103.6109 | 0.9093 | 113.9518 | 0.0000 |
| FOСКФО | -129.6584 | 35.2879 | -3.6743 | 0.0004 |
| FOСФО | 146.3632 | 56.5090 | 2.5901 | 0.0114 |
| FOСКФО.PromPrices | 1.2867 | 0.3409 | 3.7746 | 0.0003 |
| FOСФО.PromPrices | -1.3253 | 0.5182 | -2.5577 | 0.0125 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.224)

Значимы константы для Северо-Западного, Северо-Кавказского и Центрального федеральных округов, а также коэффициенты при независимых переменных для некоторых округов.

### 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) | 103.6109 | 0.9093 | 113.9518 | 0.0000 |
| FOСКФО | -129.6584 | 35.2879 | -3.6743 | 0.0004 |
| FOСФО | 146.3632 | 56.5090 | 2.5901 | 0.0114 |
| FOСКФО.PromPrices | 1.2867 | 0.3409 | 3.7746 | 0.0003 |
| FOСФО.PromPrices | -1.3253 | 0.5182 | -2.5577 | 0.0125 |

Коэффициент модели при *PromPrices* значим, однако коэффициент детерминации заметно понизился ( 0.224).

## 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 | -0.019 | 0.24 | 8.1 |
| fit.1.foBonferroni | 0.184 | 5.61 | 7.3 |
| fit.1.fo | 0.184 | 5.61 | 7.3 |

Результат:

Среднее по Y = 138420.6 (~ 16% стандартной ошибки);

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

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

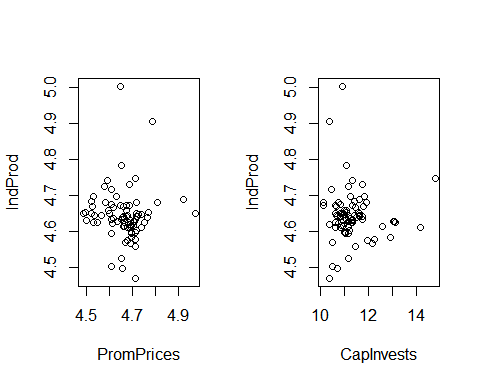
* Y (*IndProd*) – Среднедушевые денежные доходы населения;
* X1 (*PromPrices*) – ВРП на душу населения;
* X3 (*CapInvests*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику;

По количеству 83-x наблюдений.

## Оценка параметров этой модели

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.7276 | 0.4498 | 10.5111 | 0.0000 |
| PromPrices | -0.0160 | 0.0960 | -0.1663 | 0.8684 |
| CapInvests | -0.0009 | 0.0100 | -0.0880 | 0.9301 |



#### 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 CapInvests.**

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

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0. 51% исходного разброса зависимой переменной Y (оборота розничной торговли на душу населения) объясняет разброс объясняющих переменных X1 (число малых предприятий на 10000 человек населения), X3 (расходы консолидированных бюджетов субъектов Российской Федерации: всего).

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) | 2.3665 | 4.2072 | 0.5625 | 0.5759 |
| FOПФО | 5.4842 | 4.7098 | 1.1644 | 0.2489 |
| FOСЗФО | 1.8531 | 4.3000 | 0.4310 | 0.6681 |
| FOСКФО | -4.7305 | 4.8472 | -0.9759 | 0.3331 |
| FOСФО | 6.9197 | 4.9861 | 1.3878 | 0.1704 |
| FOУФО | 0.9662 | 4.8202 | 0.2005 | 0.8418 |
| FOЦФО | 2.4532 | 4.3514 | 0.5638 | 0.5750 |
| FOЮФО | 2.2265 | 4.3457 | 0.5124 | 0.6103 |
| PromPrices | 0.4704 | 0.8777 | 0.5359 | 0.5940 |
| CapInvests | 0.0033 | 0.0317 | 0.1025 | 0.9187 |
| FOПФО:PromPrices | -1.0083 | 0.9672 | -1.0425 | 0.3014 |
| FOСЗФО:PromPrices | -0.4475 | 0.8955 | -0.4997 | 0.6192 |
| FOСКФО:PromPrices | 0.7448 | 0.9462 | 0.7872 | 0.4343 |
| FOСФО:PromPrices | -1.6200 | 1.0323 | -1.5693 | 0.1219 |
| FOУФО:PromPrices | -0.1817 | 1.0123 | -0.1794 | 0.8582 |
| FOЦФО:PromPrices | -0.5184 | 0.9099 | -0.5698 | 0.5710 |
| FOЮФО:PromPrices | -0.4060 | 0.9046 | -0.4488 | 0.6552 |
| FOПФО:CapInvests | -0.0665 | 0.0663 | -1.0040 | 0.3195 |
| FOСЗФО:CapInvests | 0.0235 | 0.0372 | 0.6329 | 0.5293 |
| FOСКФО:CapInvests | 0.1274 | 0.1367 | 0.9321 | 0.3551 |
| FOСФО:CapInvests | 0.0656 | 0.0598 | 1.0962 | 0.2775 |
| FOУФО:CapInvests | -0.0082 | 0.0389 | -0.2099 | 0.8345 |
| FOЦФО:CapInvests | 0.0019 | 0.0520 | 0.0361 | 0.9713 |
| FOЮФО:CapInvests | -0.0237 | 0.0606 | -0.3913 | 0.6970 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (32.9%). 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 8 - описательные статистики логарифмированной модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 4.6398 | 0.0082 | 562.5148 | 0.0000 |
| FOСКФО | -5.2294 | 1.5190 | -3.4427 | 0.0009 |
| FOСФО | 5.8303 | 2.4136 | 2.4156 | 0.0180 |
| FOСКФО.PromPrices | 1.1331 | 0.3278 | 3.4569 | 0.0009 |
| FOСФО.PromPrices | -1.2405 | 0.5146 | -2.4108 | 0.0183 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.193.

Значимы константы для Приволжского, Северо-Западного, Северо-Кавказского и Северного федеральных округов, коэффициент при CapInvests, а также коэффициенты при независимых переменных для некоторых округов.

*(PromPrices с Южным федеральным округом;* *CapInvests с Северо-Западным и Южным федеральными округами)*

### 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) | 4.6416 | 0.0078 | 597.9815 | 0.0000 |
| FOСКФО | -5.2313 | 1.5577 | -3.3583 | 0.0012 |
| FOСКФО.PromPrices | 1.1331 | 0.3361 | 3.3710 | 0.0012 |

Коэффициент модели при *FOСКФО.CapInvests* значим, однако коэффициент детерминации заметно понизился ( 0.129).

Сравнивать эту модель с двумя другими нет смысла, так как в ней отсутствуют главные независимые переменные.

## 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 | -0.025 | 0.02 | 0.072 |
| fit.11.fo | 0.151 | 4.65 | 0.066 |

Результат:

Среднее по Y = 11,7907 (~ 1% стандартной ошибки);

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

При отношении F расчётного к скорректированному выбор падает на первую модель. Хоть её коэффициент детерминации ниже, чем у второй, но F в 2,3 раза больше.

**Явный вид модели: .**

Сохраним нужные данные для дальнейших лабораторных и пойдём спатеньки uwu.