Лабораторная работа №2

Ларшина М.А.

21 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

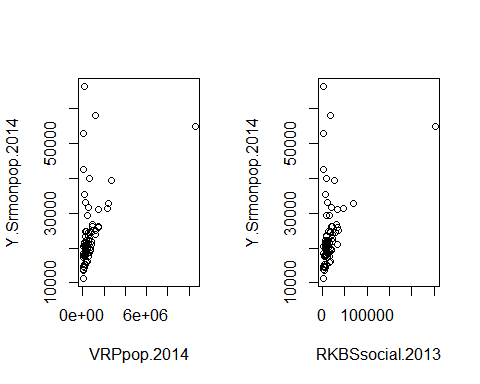
* Y (*Y.Srmonpop.2014*) – Среднедушевые денежные доходы населения;
* X1 (*VRPpop.2014*) – ВРП на душу населения;
* X3 (*RKBSsocial.2013*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику;

По количеству 83-x наблюдений.

## Оценка параметров этой модели

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 22200.8549 | 1101.7566 | 20.1504 | 0.0000 |
| VRPpop.2014 | 0.0126 | 0.0037 | 3.3764 | 0.0011 |
| RKBSsocial.2013 | -0.3908 | 0.1601 | -2.4412 | 0.0168 |



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

**Проверка значимости для коэффициента при VRPpop.2014**

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 VRPpop.2014 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при RKBSsocial.2013.**

P-значение при RKBSsocial.2013 = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.264.

## 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) | 19397.0729 | 2133.2923 | 9.0926 | 0.0000 |
| FO‘”Ћ | -3255.2055 | 4344.4075 | -0.7493 | 0.4567 |
| FO‘‡”Ћ | 13803.8744 | 4049.9837 | 3.4084 | 0.0012 |
| FO‘Љ”Ћ | -4750.5364 | 5132.3501 | -0.9256 | 0.3584 |
| FO“”Ћ | 6892.5834 | 7874.1825 | 0.8753 | 0.3849 |
| FO„‚”Ћ | 21312.6064 | 4870.7458 | 4.3756 | 0.0001 |
| FOћ”Ћ | -4057.9952 | 4964.0800 | -0.8175 | 0.4169 |
| FOЏ”Ћ | -6143.6066 | 4495.9893 | -1.3665 | 0.1770 |
| VRPpop.2014 | 0.0007 | 0.0135 | 0.0489 | 0.9612 |
| RKBSsocial.2013 | 0.1166 | 0.5671 | 0.2056 | 0.8378 |
| FO‘”Ћ:VRPpop.2014 | 0.0051 | 0.0301 | 0.1702 | 0.8655 |
| FO‘‡”Ћ:VRPpop.2014 | 0.0469 | 0.0222 | 2.1175 | 0.0384 |
| FO‘Љ”Ћ:VRPpop.2014 | 0.0179 | 0.0400 | 0.4463 | 0.6570 |
| FO“”Ћ:VRPpop.2014 | 0.0110 | 0.0145 | 0.7570 | 0.4521 |
| FO„‚”Ћ:VRPpop.2014 | 0.0148 | 0.0201 | 0.7369 | 0.4641 |
| FOћ”Ћ:VRPpop.2014 | 0.0072 | 0.0216 | 0.3323 | 0.7408 |
| FOЏ”Ћ:VRPpop.2014 | 0.0051 | 0.0159 | 0.3207 | 0.7496 |
| FO‘”Ћ:RKBSsocial.2013 | -0.0688 | 1.1196 | -0.0614 | 0.9512 |
| FO‘‡”Ћ:RKBSsocial.2013 | -2.5117 | 1.0435 | -2.4070 | 0.0192 |
| FO‘Љ”Ћ:RKBSsocial.2013 | -0.2082 | 1.1994 | -0.1736 | 0.8628 |
| FO“”Ћ:RKBSsocial.2013 | -0.5018 | 0.7372 | -0.6807 | 0.4987 |
| FO„‚”Ћ:RKBSsocial.2013 | -1.2503 | 0.7492 | -1.6687 | 0.1005 |
| FOћ”Ћ:RKBSsocial.2013 | -0.1478 | 0.8579 | -0.1722 | 0.8638 |
| FOЏ”Ћ:RKBSsocial.2013 | 0.1140 | 0.6999 | 0.1629 | 0.8711 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (62.4%). 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:

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 17736.9634 | 980.6187 | 18.0875 | 0.0000 |
| FO…Ћ | 15463.9839 | 3382.0689 | 4.5723 | 0.0000 |
| FO…Ћ.1 | 23780.5888 | 4168.6460 | 5.7046 | 0.0000 |
| RKBSsocial.2013 | 0.1567 | 0.0265 | 5.9083 | 0.0000 |
| FO…Ћ.VRPpop.2014 | 0.0476 | 0.0165 | 2.8852 | 0.0051 |
| FO..Ћ.VRPpop.2014.1 | 0.0088 | 0.0020 | 4.4211 | 0.0000 |
| FO…Ћ.RKBSsocial.2013 | -2.5519 | 0.8240 | -3.0969 | 0.0028 |
| FO…Ћ.RKBSsocial.2013.1 | -0.9255 | 0.3222 | -2.8724 | 0.0053 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.577)

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) | 17736.9634 | 980.6187 | 18.0875 | 0.0000 |
| FO…Ћ | 15463.9839 | 3382.0689 | 4.5723 | 0.0000 |
| FO…Ћ.1 | 23780.5888 | 4168.6460 | 5.7046 | 0.0000 |
| RKBSsocial.2013 | 0.1567 | 0.0265 | 5.9083 | 0.0000 |
| FO…Ћ.VRPpop.2014 | 0.0476 | 0.0165 | 2.8852 | 0.0051 |
| FO..Ћ.VRPpop.2014.1 | 0.0088 | 0.0020 | 4.4211 | 0.0000 |
| FO…Ћ.RKBSsocial.2013 | -2.5519 | 0.8240 | -3.0969 | 0.0028 |
| FO…Ћ.RKBSsocial.2013.1 | -0.9255 | 0.3222 | -2.8724 | 0.0053 |

Коэффициент модели при *VRPpop.2014* значим, однако коэффициент детерминации заметно понизился ( 0.577).

## 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 | 0.246 | 14.35 | 8565.9 |
| fit.1.foBonferroni | 0.538 | 14.64 | 6704.5 |
| fit.1.fo | 0.538 | 14.64 | 6704.5 |

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

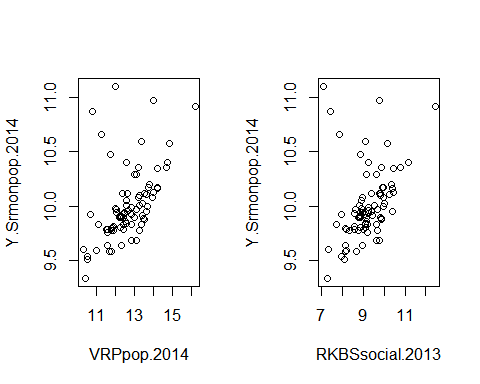
* Y (*Y.Srmonpop.2014*) – Среднедушевые денежные доходы населения;
* X1 (*VRPpop.2014*) – ВРП на душу населения;
* X3 (*RKBSsocial.2013*) – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику;

По количеству 83-x наблюдений.

## Оценка параметров этой модели

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 8.0160 | 0.3516 | 22.7960 | 0e+00 |
| VRPpop.2014 | 0.3801 | 0.0715 | 5.3151 | 0e+00 |
| RKBSsocial.2013 | -0.3061 | 0.0872 | -3.5107 | 7e-04 |



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

## Проверка значимости для логарифмированных значений:

**Проверка значимости для коэффициента при VRPpop.2014.**

H0: (параметр) коэфф. при VRPpop.2014 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при VRPpop.2014 не равен 0 в генеральной совокупности (значим).

Проверим значимость при помощи p-значения. ( )

P-значение при VRPpop.2014 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при RKBSsocial.2013.**

P-значение при RKBSsocial.2013 = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.34. 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) | 6.9444 | 0.5976 | 11.6197 | 0.0000 |
| FO‘”Ћ | 1.3213 | 0.8101 | 1.6310 | 0.1082 |
| FO‘‡”Ћ | 1.9997 | 1.0105 | 1.9788 | 0.0525 |
| FO‘Љ”Ћ | 0.8256 | 1.0817 | 0.7633 | 0.4483 |
| FO“”Ћ | 1.0612 | 1.4448 | 0.7345 | 0.4656 |
| FO„‚”Ћ | 3.7811 | 0.8987 | 4.2073 | 0.0001 |
| FOћ”Ћ | 0.2077 | 1.2333 | 0.1684 | 0.8668 |
| FOЏ”Ћ | -0.2533 | 0.9467 | -0.2676 | 0.7900 |
| VRPpop.2014 | 0.2279 | 0.1566 | 1.4551 | 0.1509 |
| RKBSsocial.2013 | 0.0155 | 0.1719 | 0.0899 | 0.9287 |
| FO‘”Ћ:VRPpop.2014 | -0.0235 | 0.2232 | -0.1052 | 0.9166 |
| FO‘‡”Ћ:VRPpop.2014 | 0.2283 | 0.1883 | 1.2127 | 0.2301 |
| FO‘Љ”Ћ:VRPpop.2014 | -0.0602 | 0.2171 | -0.2772 | 0.7826 |
| FO“”Ћ:VRPpop.2014 | 0.5118 | 0.2454 | 2.0854 | 0.0414 |
| FO„‚”Ћ:VRPpop.2014 | 0.1035 | 0.1957 | 0.5288 | 0.5989 |
| FOћ”Ћ:VRPpop.2014 | 0.2772 | 0.4242 | 0.6535 | 0.5160 |
| FOЏ”Ћ:VRPpop.2014 | -0.0345 | 0.2220 | -0.1554 | 0.8771 |
| FO‘”Ћ:RKBSsocial.2013 | -0.1211 | 0.2658 | -0.4557 | 0.6503 |
| FO‘‡”Ћ:RKBSsocial.2013 | -0.5208 | 0.1989 | -2.6187 | 0.0112 |
| FO‘Љ”Ћ:RKBSsocial.2013 | -0.0148 | 0.2552 | -0.0578 | 0.9541 |
| FO“”Ћ:RKBSsocial.2013 | -0.8127 | 0.3654 | -2.2243 | 0.0300 |
| FO„‚”Ћ:RKBSsocial.2013 | -0.5064 | 0.2180 | -2.3224 | 0.0237 |
| FOћ”Ћ:RKBSsocial.2013 | -0.4169 | 0.4751 | -0.8774 | 0.3838 |
| FOЏ”Ћ:RKBSsocial.2013 | 0.0567 | 0.2674 | 0.2121 | 0.8327 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (81.4%). 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:

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 7.6268 | 0.2441 | 31.2467 | 0.0000 |
| FO…Ћ.1 | 3.5760 | 0.6001 | 5.9586 | 0.0000 |
| VRPpop.2014 | 0.1795 | 0.0193 | 9.2992 | 0.0000 |
| FO…Ћ.VRPpop.2014 | 0.4018 | 0.0716 | 5.6093 | 0.0000 |
| FO..Ћ.VRPpop.2014.1 | 0.5473 | 0.1890 | 2.8950 | 0.0050 |
| FO…Ћ.RKBSsocial.2013 | -0.5354 | 0.1003 | -5.3403 | 0.0000 |
| FO..Ћ.RKBSsocial.2013.1 | -0.7407 | 0.2632 | -2.8139 | 0.0062 |
| FO…Ћ.RKBSsocial.2013.1 | -0.3373 | 0.0666 | -5.0677 | 0.0000 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.755.

Значимы константы для Приволжского, Северо-Западного, Северо-Кавказского и Северного федеральных округов, коэффициент при RKBSsocial.2013, а также коэффициенты при независимых переменных для некоторых округов.

*(VRPpop.2014 с Южным федеральным округом;* *RKBSsocial.2013 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) | 7.9773 | 0.2405 | 33.1652 | 0e+00 |
| FO…Ћ.1 | 3.7898 | 0.6617 | 5.7269 | 0e+00 |
| RKBSsocial.2013 | 0.2077 | 0.0260 | 7.9799 | 0e+00 |
| FO…Ћ.VRPpop.2014 | 0.5480 | 0.0805 | 6.8114 | 0e+00 |
| FO..Ћ.VRPpop.2014.1 | 0.7388 | 0.2020 | 3.6578 | 5e-04 |
| FO…Ћ.RKBSsocial.2013 | -0.7352 | 0.1125 | -6.5334 | 0e+00 |
| FO..Ћ.RKBSsocial.2013.1 | -1.0007 | 0.2815 | -3.5547 | 7e-04 |
| FO…Ћ.RKBSsocial.2013.1 | -0.3635 | 0.0734 | -4.9494 | 0e+00 |

Коэффициент модели при *FOСКФО.RKBSsocial.2013* значим, однако коэффициент детерминации заметно понизился ( 0.714).

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 | 0.324 | 20.63 | 0.273 |
| fit.11.fo | 0.732 | 32.98 | 0.172 |

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.