Лабораторная работа №2

Бессмертная М.М.

20 Декабрь 2020

## Импорт данных

Импортируем объекты, сохранённые в рабочем пространстве по итогу лабораторной №1.

## [1] "DF" "pic.num" "reg.df" "reg.df1" "table.num"

# Раздел I.

## Изначальная регрессионная модель, основанная на Лабораторной №1

Модель 0: , где

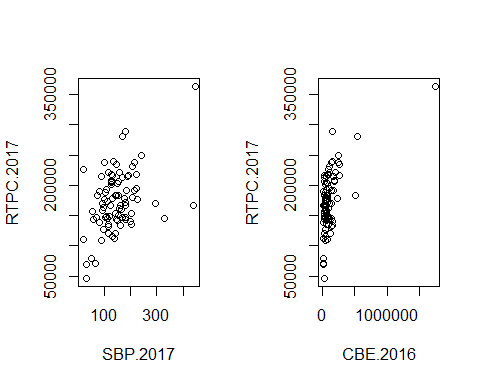
* Y (*RTPC.2017*) – оборот розничной торговли на душу населения;
* X1 (*SBP.2017*) – число малых предприятий на 10000 человек населения;
* X3 (*CBE.2016*) – расходы консолидированных бюджетов субъектов Российской Федерации: всего.

По количеству 85-x наблюдений.

## Оценка параметров этой модели

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 140385.4806 | 10007.5930 | 14.0279 | 0.000 |
| SBP.2017 | 134.6743 | 68.2968 | 1.9719 | 0.052 |
| CBE.2016 | 0.1186 | 0.0247 | 4.7925 | 0.000 |



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

**Проверка значимости для коэффициента при SBP.2017.**

H0: (параметр) коэфф. при SBP.2017 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при SBP.2017 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 SBP.2017 = => принимается гипотеза H0. **Параметр не значим.**

**Проведём аналогичную проверку коэффициента линейной регрессии при CBE.2016.**

P-значение при CBE.2016 = => принимается гипотеза H1. **Параметр значим.**

## Исключение незначимого параметра из модели (Х1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 157477.1930 | 5088.682 | 30.9466 | 0 |
| CBE.2016 | 0.1423 | 0.022 | 6.4670 | 0 |

Явный вид модели 1: .

P-значение при CBE.2016 = => принимается гипотеза H1. **Параметр значим.**

## Модель с переменной структурой по федеральным округам.

Построим модель с переменной структурой, используя в качестве параметра принадлежность каждого региона к федеральному округу. Для этого включим соответсвующие фиктивные переменные как в константу, так и в коэффициенты.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 166384.7486 | 17629.1014 | 9.4381 | 0.0000 |
| FOПФО | -60475.8830 | 23265.8704 | -2.5993 | 0.0114 |
| FOСЗФО | 25852.6520 | 21118.6347 | 1.2242 | 0.2251 |
| FOСКФО | -95762.8187 | 28341.1924 | -3.3789 | 0.0012 |
| FOСФО | -46704.6330 | 23475.0866 | -1.9895 | 0.0506 |
| FOУФО | -41730.2503 | 35677.1675 | -1.1697 | 0.2462 |
| FOЦФО | 6404.9257 | 19282.1681 | 0.3322 | 0.7408 |
| FOЮФО | -41651.4872 | 23970.5795 | -1.7376 | 0.0867 |
| CBE.2016 | 0.4465 | 0.1670 | 2.6743 | 0.0093 |
| FOПФО:CBE.2016 | 0.0562 | 0.2064 | 0.2723 | 0.7862 |
| FOСЗФО:CBE.2016 | -0.4732 | 0.1809 | -2.6155 | 0.0109 |
| FOСКФО:CBE.2016 | 0.7052 | 0.3874 | 1.8205 | 0.0730 |
| FOСФО:CBE.2016 | -0.2073 | 0.2167 | -0.9567 | 0.3421 |
| FOУФО:CBE.2016 | 0.0383 | 0.2389 | 0.1603 | 0.8731 |
| FOЦФО:CBE.2016 | -0.3289 | 0.1679 | -1.9582 | 0.0542 |
| FOЮФО:CBE.2016 | -0.0873 | 0.2088 | -0.4180 | 0.6773 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (69.3%). В приведенной таблице видно, что у модели много незначимых параметров. Исключать их последовательно вручную трудоёмко, поэтому мы воспользуемся пользовательской функцией для проведения процедуры последовательного исключения регрессоров.

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 4 - описательные статистики модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 176743.1442 | 5596.7618 | 31.5795 | 0.0000 |
| FOПФО | -58357.2356 | 10681.5350 | -5.4634 | 0.0000 |
| FOСКФО | -106121.2143 | 22721.2454 | -4.6706 | 0.0000 |
| FOСФО | -70611.0121 | 10883.6353 | -6.4878 | 0.0000 |
| FOУФО | -35482.2169 | 15570.0432 | -2.2789 | 0.0255 |
| FOЮФО | -54538.2825 | 12569.6643 | -4.3389 | 0.0000 |
| CBE.2016 | 0.3851 | 0.0567 | 6.7976 | 0.0000 |
| FOСЗФО.CBE.2016 | -0.3539 | 0.0772 | -4.5862 | 0.0000 |
| FOСКФО.CBE.2016 | 0.7666 | 0.3515 | 2.1811 | 0.0323 |
| FOЦФО.CBE.2016 | -0.2712 | 0.0578 | -4.6891 | 0.0000 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.672)

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

*(CBE.2016 с Западным, Северо-Кавказким и Центральным федеральными округами)*

### Модель с поправкой Бонферрони:

Явный вид модели 2: .

#### Таблица 5 - описательные статистики модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 173372.6075 | 5775.9876 | 30.0161 | 0e+00 |
| FOПФО | -49656.4250 | 10351.4617 | -4.7970 | 0e+00 |
| FOСКФО | -58259.4643 | 13039.3648 | -4.4680 | 0e+00 |
| FOСФО | -62577.0026 | 10749.4163 | -5.8214 | 0e+00 |
| FOЮФО | -46255.7343 | 12584.0656 | -3.6757 | 4e-04 |
| CBE.2016 | 0.3349 | 0.0508 | 6.5870 | 0e+00 |
| FOСЗФО.CBE.2016 | -0.2911 | 0.0733 | -3.9736 | 2e-04 |
| FOЦФО.CBE.2016 | -0.2178 | 0.0517 | -4.2153 | 1e-04 |

Коэффициент модели при *SBP.2017* значим, однако коэффициент детерминации заметно понизился ( 0.624).

## Сравнение моделей по качеству.

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

#### Таблица 6 - сравнение трёх моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.1 | 0.327 | 41.82 | 40482.8 |
| fit.1.foBonferroni | 0.590 | 18.26 | 31602.6 |
| fit.1.fo | 0.632 | 17.04 | 29929.8 |

Результат:

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) является наиболее подходящей.

**Явный вид модели 3: .**

# Раздел II.

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

Модель 0: , где

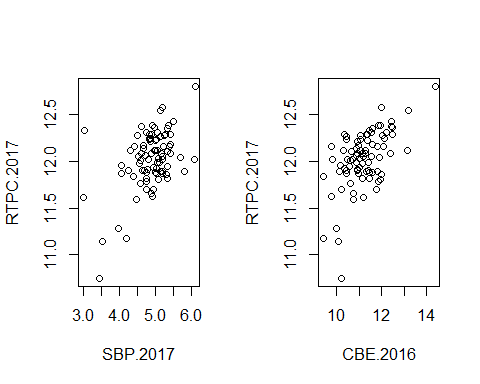
* Y (*RTPC.2017*) – оборот розничной торговли на душу населения;
* X1 (*SBP.2017*) – число малых предприятий на 10000 человек населения;
* X3 (*CBE.2016*) – расходы консолидированных бюджетов субъектов Российской Федерации: всего;

По количеству 85 наблюдений.

## Оценка параметров этой модели

#### Таблица 7 - описательные статистики логарифмированной модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 9.2969 | 0.3611 | 25.7437 | 0.000 |
| SBP.2017 | 0.1816 | 0.0531 | 3.4198 | 0.001 |
| CBE.2016 | 0.1645 | 0.0331 | 4.9722 | 0.000 |



#### Рис. 3. график разброса модели на логарифмированных данных

## Проверка значимости для логарифмированных значений:

**Проверка значимости для коэффициента при SBP.2017.**

H0: (параметр) коэфф. при SBP.2017 равен 0 в генеральной совокупности (не значим);

H1: (параметр) коэфф. при SBP.2017 не равен 0 в генеральной совокупности (значим).

Проверим значимость при помощи p-значения. ( )

P-значение при SBP.2017 = => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при CBE.2016.**

P-значение при CBE.2016 = => принимается гипотеза H1. **Параметр значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.416. 42% 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 в качестве параметра принадлежность каждого региона к федеральному округу. Для этого включим соответсвующие фиктивные переменные как в константу, так и в коэффициенты.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 10.1995 | 0.9778 | 10.4307 | 0.0000 |
| FOПФО | -2.5757 | 1.5372 | -1.6756 | 0.0989 |
| FOСЗФО | 3.0680 | 1.5239 | 2.0132 | 0.0485 |
| FOСКФО | -5.2539 | 1.8609 | -2.8233 | 0.0064 |
| FOСФО | -0.6153 | 1.2967 | -0.4745 | 0.6368 |
| FOУФО | -1.8987 | 1.9458 | -0.9758 | 0.3330 |
| FOЦФО | -0.1501 | 1.2261 | -0.1224 | 0.9030 |
| FOЮФО | -1.4719 | 1.5166 | -0.9705 | 0.3356 |
| SBP.2017 | 0.0515 | 0.2278 | 0.2261 | 0.8219 |
| CBE.2016 | 0.1589 | 0.1018 | 1.5603 | 0.1239 |
| FOПФО:SBP.2017 | 0.1702 | 0.3371 | 0.5050 | 0.6154 |
| FOСЗФО:SBP.2017 | -0.1905 | 0.2638 | -0.7223 | 0.4729 |
| FOСКФО:SBP.2017 | 0.1928 | 0.2707 | 0.7124 | 0.4790 |
| FOСФО:SBP.2017 | 0.2006 | 0.2660 | 0.7542 | 0.4536 |
| FOУФО:SBP.2017 | -0.0860 | 0.3400 | -0.2530 | 0.8011 |
| FOЦФО:SBP.2017 | -0.1781 | 0.2992 | -0.5951 | 0.5540 |
| FOЮФО:SBP.2017 | 0.3294 | 0.3705 | 0.8892 | 0.3774 |
| FOПФО:CBE.2016 | 0.1244 | 0.1422 | 0.8753 | 0.3848 |
| FOСЗФО:CBE.2016 | -0.1959 | 0.1260 | -1.5543 | 0.1253 |
| FOСКФО:CBE.2016 | 0.3858 | 0.1637 | 2.3558 | 0.0217 |
| FOСФО:CBE.2016 | -0.0666 | 0.1386 | -0.4805 | 0.6326 |
| FOУФО:CBE.2016 | 0.1838 | 0.1819 | 1.0103 | 0.3164 |
| FOЦФО:CBE.2016 | 0.0843 | 0.1218 | 0.6924 | 0.4913 |
| FOЮФО:CBE.2016 | -0.0312 | 0.1291 | -0.2414 | 0.8100 |

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (69.6%).В приведенной таблице видно, что у модели много незначимых параметров. Исключать их последовательно вручную трудоёмко, поэтому мы воспользуемся пользовательской функцией, которая проводит процедуру последовательного исключения регрессоров.

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 9 - описательные статистики логарифмированной модели по федеральным округам без поправки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 9.3106 | 0.3564 | 26.1244 | 0.0000 |
| FOПФО | -0.1988 | 0.0627 | -3.1685 | 0.0022 |
| FOСЗФО | 3.9568 | 1.2214 | 3.2397 | 0.0018 |
| FOСКФО | -3.9839 | 1.4049 | -2.8357 | 0.0059 |
| FOСФО | -0.2547 | 0.0668 | -3.8117 | 0.0003 |
| SBP.2017 | 0.1774 | 0.0694 | 2.5549 | 0.0126 |
| CBE.2016 | 0.1719 | 0.0332 | 5.1772 | 0.0000 |
| FOСЗФО.SBP.2017 | -0.3164 | 0.1500 | -2.1095 | 0.0382 |
| FOСЗФО.CBE.2016 | -0.2090 | 0.0813 | -2.5691 | 0.0122 |
| FOСКФО.CBE.2016 | 0.3604 | 0.1319 | 2.7313 | 0.0079 |

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.627.

Значимы константы для Приволжского, Северо-Западного, Северо-Кавказского и Северного федеральных округов, коэффициенты при SBP.2017 и CBE.2016, а также коэффициенты при независимых переменных для некоторых округов.

*(SBP.2017 с Северо-Западным федеральным округом;* *CBE.2016 с Северо-Западным и Северо-Кавказским федеральным округом)*

### Модель с поправкой Бонферрони:

Явный вид модели 3: .

#### Таблица 10 - описательные статистики логарифмированной модели по федеральным округам с поправкой Бонферрони

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 9.4863 | 0.3329 | 28.4981 | 0.0000 |
| FOСЗФО | 2.9039 | 0.9243 | 3.1416 | 0.0024 |
| FOСКФО | -0.3056 | 0.0891 | -3.4302 | 0.0010 |
| CBE.2016 | 0.2344 | 0.0295 | 7.9503 | 0.0000 |
| FOПФО.CBE.2016 | -0.0173 | 0.0058 | -2.9629 | 0.0040 |
| FOСЗФО.CBE.2016 | -0.2567 | 0.0828 | -3.1015 | 0.0027 |
| FOСФО.CBE.2016 | -0.0242 | 0.0063 | -3.8515 | 0.0002 |

Коэффициент модели при *FOСКФО.CBE.2016* значим, однако коэффициент детерминации заметно понизился ( 0.563).

## Сравнение моделей по качеству.

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

#### Таблица 11 - сравнение двух логарифмированных моделей

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.11 | 0.402 | 29.22 | 0.242 |
| fit.11.fo | 0.582 | 13.99 | 0.202 |
| fit.11.foBonferroni | 0.529 | 16.75 | 0.215 |

Результат:

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

**Явный вид модели: .**