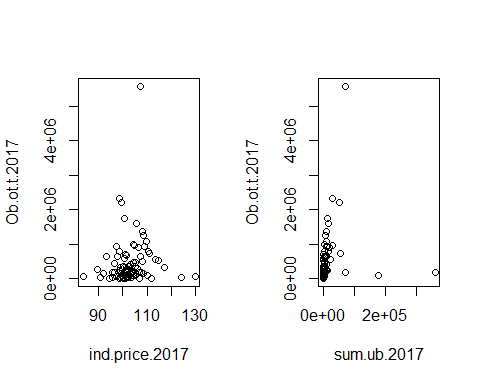
Lab\_test\_2

Блинов И И

01 12 2020

#### Таблица 1 - описательные статистики модели 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | -3163979.0483 | 4255151.2525 | -0.7436 | 0.4594 |
| ind.price.2017 | 8810.0698 | 12622.2407 | 0.6980 | 0.4873 |
| izm.sr.2017 | 16611.2289 | 38954.0484 | 0.4264 | 0.6710 |
| ind.phis.2016 | 10278.3703 | 30837.7519 | 0.3333 | 0.7398 |
| sum.ub.2017 | 2.6067 | 1.8774 | 1.3885 | 0.1689 |

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

**Проверка значимости для коэффициента при sum.ub.2017.**

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

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

Проверим значимость при помощи p-значения.

**Параметр не значим.**

**Проведём похожую проверку коэффициента при ind.price.2017.**

P-значение при ind.price.2017 > 0,05 => **Параметр не значим.**

**Проведём похожую проверку коэффициента при ind.price.2017.**

P-значение при ind.price.2017 = => отверается гипотеза H1. **Параметр не значим.**

## 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.

fit.1.fo <- lm(Ob.ot.t.2017 ~ FO\*(ind.price.2017 + sum.ub.2017),  
 data = reg.df)   
kable(round(summary(fit.1.fo)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 305584.4685 | 2820259.2063 | 0.1084 | 0.9141 |
| FOПФО | -3506888.1661 | 3321347.5044 | -1.0559 | 0.2953 |
| FOСЗФО | -119140.1679 | 4279525.9067 | -0.0278 | 0.9779 |
| FOСКФО | -320432.2259 | 3224879.4592 | -0.0994 | 0.9212 |
| FOСФО | -1681687.2307 | 3166147.9103 | -0.5311 | 0.5973 |
| FOУФО | 2115070.9371 | 3933526.1956 | 0.5377 | 0.5928 |
| FOЦФО | 928777.7722 | 3401070.9852 | 0.2731 | 0.7857 |
| FOЮФО | 834527.0877 | 3036041.0646 | 0.2749 | 0.7844 |
| ind.price.2017 | -2657.7620 | 27474.6644 | -0.0967 | 0.9233 |
| sum.ub.2017 | 14.9314 | 21.8353 | 0.6838 | 0.4968 |
| FOПФО:ind.price.2017 | 39101.0505 | 32171.2240 | 1.2154 | 0.2291 |
| FOСЗФО:ind.price.2017 | 2389.3176 | 41382.3128 | 0.0577 | 0.9542 |
| FOСКФО:ind.price.2017 | 2656.5513 | 30909.7199 | 0.0859 | 0.9318 |
| FOСФО:ind.price.2017 | 18748.6568 | 30924.5357 | 0.6063 | 0.5467 |
| FOУФО:ind.price.2017 | -11468.8844 | 37886.4185 | -0.3027 | 0.7632 |
| FOЦФО:ind.price.2017 | -10492.3682 | 33406.2668 | -0.3141 | 0.7546 |
| FOЮФО:ind.price.2017 | -6692.0761 | 29549.7391 | -0.2265 | 0.8216 |
| FOПФО:sum.ub.2017 | -15.8555 | 21.8568 | -0.7254 | 0.4711 |
| FOСЗФО:sum.ub.2017 | 23.4121 | 22.8325 | 1.0254 | 0.3094 |
| FOСКФО:sum.ub.2017 | 55.4034 | 146.3067 | 0.3787 | 0.7063 |
| FOСФО:sum.ub.2017 | -5.4020 | 27.5798 | -0.1959 | 0.8454 |
| FOУФО:sum.ub.2017 | -19.1374 | 21.9326 | -0.8726 | 0.3864 |
| FOЦФО:sum.ub.2017 | 63.0265 | 22.3961 | 2.8142 | 0.0066 |
| FOЮФО:sum.ub.2017 | -0.1934 | 22.8233 | -0.0085 | 0.9933 |

table.num <- table.num + 1

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (86.1%). 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

X.matrix <- model.matrix(Ob.ot.t.2017 ~ FO\*(ind.price.2017 + sum.ub.2017),   
 data = reg.df)  
   
   
 # присоединяем независимую переменную  
data.fit <- cbind(Ob.ot.t.2017 = reg.df$Ob.ot.t.2017,   
 data.frame(X.matrix)[, -1])  
   
#сохраняем для следующей лабораторной  
data.fit.1.fo <- data.fit  
  
# функция с последовательным исключением незначимых регрессоров  
source('https://raw.githubusercontent.com/aksyuk/R-Practice-basics/master/user\_functions/removeFactorsByPValue.R')  
  
# применяем процедуру, сначала без поправок на p-значения  
fit.1.fo <- removeFactorsByPValue(data = data.fit,   
 y.var.name = 'Ob.ot.t.2017')  
kable(round(summary(fit.1.fo)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 150669.1043 | 51326.3860 | 2.9355 | 0.0045 |
| FOПФО | -3351972.8019 | 1656961.6104 | -2.0230 | 0.0467 |
| FOУФО | 787362.7268 | 151552.9746 | 5.1953 | 0.0000 |
| sum.ub.2017 | 15.3600 | 4.9992 | 3.0725 | 0.0030 |
| FOПФО.ind.price.2017 | 36443.2885 | 15801.0530 | 2.3064 | 0.0239 |
| FOЦФО.ind.price.2017 | -2329.1439 | 1006.4654 | -2.3142 | 0.0235 |
| FOПФО.sum.ub.2017 | -16.2841 | 5.0823 | -3.2041 | 0.0020 |
| FOСЗФО.sum.ub.2017 | 23.2109 | 7.0838 | 3.2766 | 0.0016 |
| FOУФО.sum.ub.2017 | -19.5927 | 5.3654 | -3.6517 | 0.0005 |
| FOЦФО.sum.ub.2017 | 61.7233 | 6.7206 | 9.1842 | 0.0000 |

table.num <- table.num + 1  
#summary(fit.1.fo)

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. ( 0.847) Значимы константы для Приволжского и Уральского федеральных округов, а также коэффициенты при независимых переменных для некоторых округов.

### Модель с поправкой Бонферрони:

Явный вид модели : .

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

# теперь с поправкой Бонферрони  
fit.1.foB <- removeFactorsByPValue(data = data.fit,   
 y.var.name = 'Ob.ot.t.2017',  
 p.adj.method = 'bonferroni')  
kable(round(summary(fit.1.foB)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 315261.0952 | 52863.3117 | 5.9637 | 0 |
| FOЦФО.sum.ub.2017 | 65.8587 | 5.8084 | 11.3385 | 0 |

table.num <- table.num + 1

Коэффициент модели при *FOЦФО.sum.ub.2017* значим, однако коэффициент детерминации заметно понизился ( 0.613).

## Сравнение моделей по качеству.

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

#### Таблица 5 - сравнение трёх моделей

# модели с фактором sum.ub.2017  
#anova(fit.1, fit.1.foB, fit.1.fo)  
# список построенных моделей  
models.list <- list(fit.1, fit.1.foB, fit.1.fo)  
names(models.list) <- c('fit.1', 'fit.1.foBonferroni', 'fit.1.fo')  
# фрейм с характеристиками четырёх моделей  
df.goodness.of.fit <- data.frame(Модель = names(models.list),   
 R.2.скорр = 0,  
 F.расч = 0,  
 Станд.Ошибка = 0)  
for (i in 1:length(models.list)) {  
 # скорректированный R-квадрат  
 df.goodness.of.fit[i, 'R.2.скорр'] <-   
 round(summary(models.list[[i]])$adj.r.squared, 3)  
 # F расчётное  
 df.goodness.of.fit[i, 'F.расч'] <-   
 round(summary(models.list[[i]])$fstatistic[1], 2)  
 # стандартная ошибка  
 df.goodness.of.fit[i, 'Станд.Ошибка'] <-   
 round(summary(models.list[[i]])$sigma, 1)  
}  
kable(df.goodness.of.fit)

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.1 | -0.013 | 0.73 | 749906.5 |
| fit.1.foBonferroni | 0.609 | 128.56 | 466006.2 |
| fit.1.fo | 0.828 | 44.88 | 308900.1 |

table.num <- table.num + 1  
mean(reg.df$Ob.ot.t.2017)

## [1] 466588.5

Результат:

Среднее по Y = 466588.5 ;

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 : $Y.ORTorg.2013 = 150669.1043 - 3351972.8019 FOПФО + 787362.7268 FOУФО + 15.3600 sum.ub.2017 + 36443.2885 FOПФО.ind.price.2017 - 2329.1439 FOЦФО.ind.price.2017 -16.2841 FOПФО.sum.ub.2017 + 23.2109 FOСЗФО.sum.ub.2017 -19.5927 FOУФО.sum.ub.2017 + 61.7233 FOЦФО.sum.ub.2017**

# Раздел II.

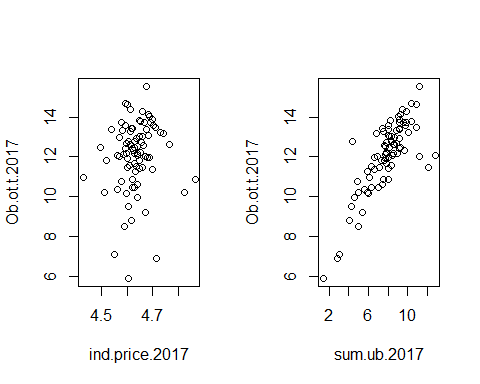
## Изначальная регрессионная модель для логарифмированных данных, основанная на Лабораторной №1

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

# множественная регрессия для всех регионов на логарифмированных данных  
fit.11 <- lm(Ob.ot.t.2017 ~ ind.price.2017 + sum.ub.2017,   
 data = DF1)  
kable(round(summary(fit.11)$coef, 4)) # незначимые параметры

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 3.6101 | 8.4812 | 0.4257 | 0.6715 |
| ind.price.2017 | 0.6913 | 1.8387 | 0.3760 | 0.7079 |
| sum.ub.2017 | 0.6698 | 0.0575 | 11.6493 | 0.0000 |

table.num <- table.num + 1  
#построим график  
par(mfrow = c(1, 2))  
plot(Ob.ot.t.2017 ~ ind.price.2017 + sum.ub.2017,   
 data = DF1)



par(mfrow = c(1, 1))  
pic.num <- pic.num + 1

## Проверка значимости для логарифмированных значений:

**Проверка значимости для коэффициента при ind.price.2017.** Проверим значимость при помощи p-значения. ( ) P-значение при sum.ub.2017 = 0.0000<$ => принимается гипотеза H1. **Параметр значим.**

**Проведём похожую проверку коэффициента при ind.price.2017.**

P-значение при ind.price.2017 = 0,7079<$ => отвергается гипотеза H1. **Параметр не значим.**

Все имеющиеся параметры значимы, исключать регрессоры не требуется. 0.634.

Явный вид модели : .

## 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

fit.11.fo <- lm(Ob.ot.t.2017 ~ FO\*(ind.price.2017 + sum.ub.2017),   
 data = DF1)   
kable(round(summary(fit.11.fo)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 50.2492 | 35.4463 | 1.4176 | 0.1616 |
| FOПФО | -78.2257 | 40.4989 | -1.9316 | 0.0582 |
| FOСЗФО | -55.2593 | 49.3346 | -1.1201 | 0.2672 |
| FOСКФО | -43.6832 | 40.3789 | -1.0818 | 0.2837 |
| FOСФО | -42.5773 | 38.3729 | -1.1096 | 0.2717 |
| FOУФО | -32.8185 | 46.9553 | -0.6989 | 0.4873 |
| FOЦФО | -40.0896 | 40.6476 | -0.9863 | 0.3280 |
| FOЮФО | -19.6330 | 38.0958 | -0.5154 | 0.6082 |
| ind.price.2017 | -9.5118 | 7.5673 | -1.2570 | 0.2137 |
| sum.ub.2017 | 0.6526 | 0.1433 | 4.5550 | 0.0000 |
| FOПФО:ind.price.2017 | 18.1947 | 8.6468 | 2.1042 | 0.0396 |
| FOСЗФО:ind.price.2017 | 12.1073 | 10.5484 | 1.1478 | 0.2557 |
| FOСКФО:ind.price.2017 | 9.4013 | 8.4755 | 1.1092 | 0.2718 |
| FOСФО:ind.price.2017 | 8.8255 | 8.2348 | 1.0717 | 0.2882 |
| FOУФО:ind.price.2017 | 8.7383 | 10.0681 | 0.8679 | 0.3890 |
| FOЦФО:ind.price.2017 | 8.3636 | 8.7572 | 0.9551 | 0.3434 |
| FOЮФО:ind.price.2017 | 4.1230 | 8.1673 | 0.5048 | 0.6156 |
| FOПФО:sum.ub.2017 | -0.5955 | 0.2099 | -2.8369 | 0.0062 |
| FOСЗФО:sum.ub.2017 | 0.0393 | 0.2080 | 0.1889 | 0.8508 |
| FOСКФО:sum.ub.2017 | 0.0368 | 0.4006 | 0.0919 | 0.9271 |
| FOСФО:sum.ub.2017 | 0.3484 | 0.1847 | 1.8863 | 0.0642 |
| FOУФО:sum.ub.2017 | -0.7261 | 0.1941 | -3.7411 | 0.0004 |
| FOЦФО:sum.ub.2017 | 0.2614 | 0.2475 | 1.0560 | 0.2953 |
| FOЮФО:sum.ub.2017 | 0.1577 | 0.1815 | 0.8692 | 0.3882 |

table.num <- table.num + 1

Модель в целом незначима, но скорректированный коэффициент детерминации у неё выше, чем у модели по всем регионам (85.8%). 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

# создаём фрейм со всеми переменными-факторами (создаём фиктивные)  
X.matrix <- model.matrix(Ob.ot.t.2017 ~ FO\*(ind.price.2017 + sum.ub.2017),   
 data = DF1)  
# присоединяем независимую переменную  
data.fit <- cbind(Ob.ot.t.2017 = DF1$Ob.ot.t.2017,   
 data.frame(X.matrix)[, -1])  
# сохраняем для следующей лабораторной  
data.fit.11.fo <- data.fit  
# функция с последовательным исключением незначимых регрессоров  
source('https://raw.githubusercontent.com/aksyuk/R-Practice-basics/master/user\_functions/removeFactorsByPValue.R')  
# применяем процедуру, сначала без поправок на p-значения  
fit.11.fo <- removeFactorsByPValue(data = data.fit,   
 y.var.name = 'Ob.ot.t.2017')  
kable(round(summary(fit.11.fo)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 5.4440 | 0.3873 | 14.0577 | 0 |
| sum.ub.2017 | 0.8453 | 0.0499 | 16.9271 | 0 |
| FOПФО.ind.price.2017 | 1.5646 | 0.3241 | 4.8278 | 0 |
| FOУФО.ind.price.2017 | 1.8097 | 0.2589 | 6.9899 | 0 |
| FOПФО.sum.ub.2017 | -0.8245 | 0.1623 | -5.0787 | 0 |
| FOУФО.sum.ub.2017 | -0.9230 | 0.1419 | -6.5045 | 0 |

table.num <- table.num + 1  
#summary(fit.11.fo)

Все коэффициенты модели значимы и она имеет высокий уровень коэффициента детерминации. 0.808.

*(sum.ub.2017, sum.ub.2017 с Приволжским и Уральским федеральными округами;* *ind.price с Приволжским и Уральским федеральными округами)*

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

# теперь с поправкой Бонферрони  
fit.11.foB <- removeFactorsByPValue(data = data.fit,   
 y.var.name = 'Ob.ot.t.2017',  
 p.adj.method = 'bonferroni')  
kable(round(summary(fit.11.foB)$coef, 4))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 6.2847 | 0.4604 | 13.6517 | 0.0000 |
| sum.ub.2017 | 0.7435 | 0.0588 | 12.6472 | 0.0000 |
| FOПФО.ind.price.2017 | 1.3855 | 0.4083 | 3.3929 | 0.0011 |
| FOПФО.sum.ub.2017 | -0.7236 | 0.2040 | -3.5460 | 0.0007 |

table.num <- table.num + 1

Коэффициент модели при *sum.ub.2017*, *FOПФО.ind.price.2017*, *FOПФО.sum.ub.2017* значимы, однако коэффициент детерминации заметно понизился ( 0.685).

Явный вид модели : .

#### Таблица 10 - сравнение трёх моделей

# модели с фактором sum.ub.2017  
#anova(fit.11, fit.11.foB, fit.11.fo)  
# список построенных моделей  
models.list <- list(fit.11, fit.11.foB, fit.11.fo)  
names(models.list) <- c('fit.11', 'fit.11.foB', 'fit.11.fo')  
# фрейм с характеристиками четырёх моделей  
df.goodness.of.fit <- data.frame(Модель = names(models.list),   
 R.2.скорр = 0,  
 F.расч = 0,  
 Станд.Ошибка = 0)  
for (i in 1:length(models.list)) {  
 # скорректированный R-квадрат  
 df.goodness.of.fit[i, 'R.2.скорр'] <-   
 round(summary(models.list[[i]])$adj.r.squared, 3)  
 # F расчётное  
 df.goodness.of.fit[i, 'F.расч'] <-   
 round(summary(models.list[[i]])$fstatistic[1], 2)  
 # стандартная ошибка  
 df.goodness.of.fit[i, 'Станд.Ошибка'] <-   
 round(summary(models.list[[i]])$sigma, 1)  
}  
kable(df.goodness.of.fit)

|  |  |  |  |
| --- | --- | --- | --- |
| Модель | R.2.скорр | F.расч | Станд.Ошибка |
| fit.11 | 0.625 | 69.23 | 1.1 |
| fit.11.foB | 0.673 | 57.26 | 1.0 |
| fit.11.fo | 0.796 | 64.84 | 0.8 |

table.num <- table.num + 1  
mean(reg.df1$Ob.ot.t.2017)

## [1] 12.07549

Результат:

Среднее по Y = 12,07907 ;

По столбцу больше всего подходит третья модель; По столбцу F.расч - первая; По минимальной Стандартной ошибке - третья. Таким образом, модель по федеральным округам (fit.11.fo) наиболее предпочтительна. **Явный вид модели: Ob.ot.t.2017=5.4440+0.8453 sum.ub.2017+1.5646 ind.price.2017+FOПФО.ind.price.2017 sum.ub.2017 + 1.8097 ind.price -0.8245sum.ub.2017 -0.9230 FOУФО.sum.ub.2017**.

Сохраним нужные данные для дальнейших лабораторных работ…….

# 4. Сохранение нужных объектов рабочего пространства -------------------------  
save(list = c('data.fit.1.fo', 'data.fit.11.fo', 'fit.1.fo', 'fit.11.fo', 'DF', 'DF1', 'reg.df'),   
 file = 'test\_lab2\_Блинов.RData')