Балашов Д. С. КМБО-03-20(Вариант 1)

Задача №1

Необходимо загрузить данные из указанного набора и произвести следующие действия.

Набор данных: Swiss.

Объясняемая переменная: *Agriculture*.

Регрессоры: *Fertility*, *Examination*.

***№1***. Оценить среднее значение, дисперсию и СКО переменных, указанных во втором и в третьем столбце.

* *Agriculture*:
* Дисперсия = 515.799 (большой разброс)
* Ско = 22.71
* Среднее арифметическое = 50.66
* *Fertility*:
* Дисперсия = 156.04 (средний разброс)
* Ско = 12.49
* Среднее арифметическое = 70.14
* *Examination*:
* Дисперсия = 63.65 (маленький разброс)
* Ско = 7.98
* Среднее арифметическое = 16.49

***№2***. Построить зависимости вида y = a + bx, где y – объясняемая переменная, x – регрессор.

*Таблица 1*. Характеристики модели зависимости параметра *Agriculture* от параметра *Examination* в наборе данных Swiss.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Параметр \ Характеристики | Значение | Std. Error | t value | Pr(>|t|) | Уровень  значимости |
| (Intercept) | 82.88 | 5.64 | 14.69 | < 2e-16 | \*\*\* |
| *Examination* | -1.95 | 0.31 | -6.33 | 9.95e - 08 | \*\*\* |

*# Agriculture = -1.95\* Examination + 82.88*

*Таблица 2*. Характеристики модели зависимости параметра *Agriculture* от параметра *Fertility* в наборе данных Swiss.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Параметр \ Характеристики | Значение | Std. Error | t value | Pr(>|t|) | Уровень  значимости |
| (Intercept) | 5.63 | 18.06 | 0.31 | 0.76 |  |
| *Fertility* | 0.64 | 0.25 | 2.53 | 0.01 | \* |

*# Agriculture = 0.64\* Fertility + 5.63*

***№3***. Оценить, насколько «хороша» модель по коэффициенту детерминации R^2.

* *Agriculture = -1.95\* Examination + 82.88:*

# R^2 = 47%, что довольно неплохой показатель, так что можно сказать, что *Сельское Хозяйство* отрицательно зависит от *Оценок на Экзамене* (т.к. коэффициент отрицательный). Однако данная модель нуждается в обработке, поскольку она станет по-настоящему “*хорошей”* (то есть, по которой можно сделать однозначные прогнозы) только, когда R^2 > 70%.

* *Agriculture = 0.64\* Fertility + 5.63:*

# Поскольку R^2 = 0.12% (что << 30%), то можно с уверенность сказать, что данная модель плоха, и делать по ней какие-либо выводы нельзя.

***№4***. Оценить, есть ли взаимосвязь между объясняемой переменной и объясняющей

переменной.

* *Agriculture = -1.95\* Examination + 82.88:*

# Поскольку при обоих коэффициентах значение p-статистики очень хорошее (3 звезды, *см. Таблица 1*) , то можно сделать вывод, что существует взаимосвязь между Объясняемой Переменной*(Agriculture)* и Регрессом(*Examination).*Объяснить данную зависимость чисто логически довольно просто: «Чем выше оценки на экзамене, тем на более престижную вакансию претендует человек, вследствие чего, количество людей заинтересованных в работе в сфере сельского хозяйства снижается».

* *Agriculture = 0.64\* Fertility + 5.63:*

# Поскольку при обоих коэффициентах значение p-статистики плохое (0 звезд / 1 звезда, *см. Таблица 2*), то взаимосвязь между Объясняемой Переменной*(Agriculture)* и Регрессом(*Fertility)* практически отсутствует, поэтому делать какие-либо выводы по этой зависимости нельзя.

Приложение1

library("lmtest")

library ("GGally")

library ("car")

data = Swiss

ggpairs(Swiss)

# Вычислить Дисперсию, СКО и Среднее Арифметическое для *Agriculture, Fertility*, *Examination:*

* var (data$Agriculture)
* var (data$Fertility)
* var (data$Examination)
* sd (data$Agriculture)
* sd (data$Fertility)
* sd (data$Examination)
* mean (data$Agriculture)
* mean (data$Fertility)
* mean (data$Examination)

# Модели, которые необходимо исследовать:

model\_agr\_ex = lm (Agriculture ~ Examination, data)

model\_agr\_F= lm (Agriculture ~ Fertility, data)

summary (model\_agr\_ex)

summary (model\_agr\_F)

Приложение2

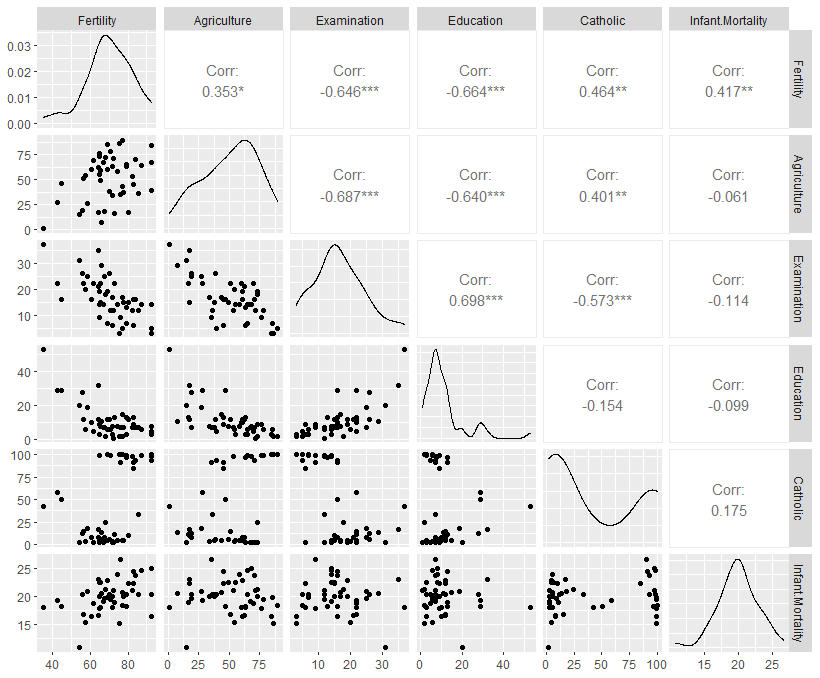


Рисунок 1. Результат работы команды ggpairs() – графики зависимостей между парами

переменных наборе данных Swiss.

Заключение

Были построены две зависимости с одинаковой Объясняемой Переменной*(Agriculture)* и разными Объясняющими Переменными (*Examination /Fertility).* Повыводам из пунктов ***№3*** и ***№4*** видно, что обе исследуемые модели нуждаются в доработке, и ни по одной из них нельзя сделать однозначных прогнозов.

*Приложение*

*Код решения всех задач:*

library("lmtest")

library("GGally")

library("car")

data = swiss

help(swiss)

data

summary(data)

ggpairs(data)

model\_agr\_ex = lm(Agriculture~Examination, data)

model\_agr\_ex

summary(model\_agr\_ex)

plot(model\_agr\_ex) + abline(a = 82.88, b = -1.95, col = "red")

model\_agr\_F = lm(Agriculture~Fertility, data)

model\_agr\_F

summary(model\_agr\_F)

plot(model\_agr\_F) + abline(a = 5.63, b = 0.64, col = "red")

var(data$Agriculture)

sd(data$Agriculture)

mean(data$Agriculture)

var(data$Examination)

sd(data$Examination)

mean(data$Examination)

var(data$Fertility)

sd(data$Fertility)

mean(data$Fertility)