Задача 1

Необходимо загрузить данные из указанного набора и произвести следующие действия.

Набор данных: Swiss.

Объясняемая переменная: Catholic.

Регрессоры: Agriculture, Examination.

1. Оцените среднее значение, дисперсию и СКО переменных, указанных во втором и третьем столбце.

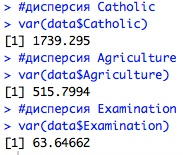
Оценим среднее значение:



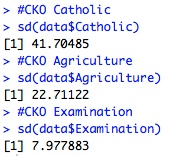




Оценим дисперсию:

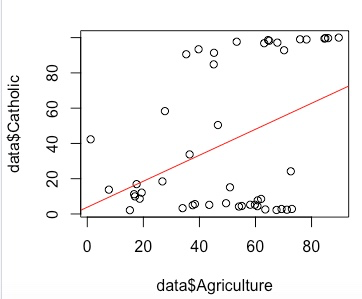


Оценим СКО переменных:

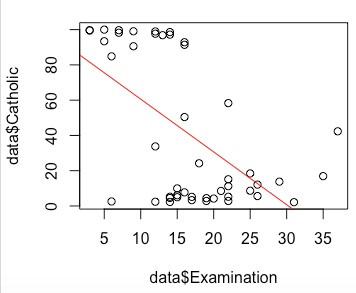


2. Постройте зависимости вида y = a + bx, где y – объясняемая переменная, x – регрессор (для каждого варианта по две зависимости).

Построим зависимость от y = Catholic и x = Agriculture:



Построим зависимость от y = Catholic и x = Examination:

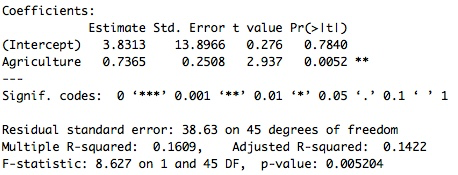


3. Оцените, насколько «хороша» модель по коэффициенту детерминации R2, есть ли взаимосвязь между объясняемой переменной и объясняющей

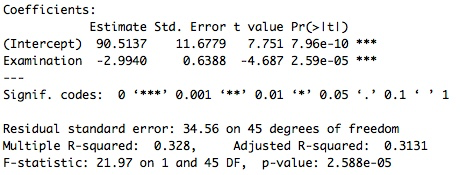
переменной (по значению p-статистики, «количеству звездочек» у регрессора в

модели).

Зависимость Catholic ~ Agriculture. R-squared = 16,09%. Значение коэффициента детерминации 16,09%, поэтому заключаем, что модель очень плохая. Из количества звёздочек у регрессора следует зависимость Catholic от Agriculture средняя.



Зависимость Catholic ~ Examination. R-squared = 32,8%. Значение коэффициента детерминации 32,8%, поэтому заключаем, что модель плоха. Из количества звёздочек у регрессора следует зависимость Catholic от Examination сильная.



Код решения задачи и сведения о проверенных моделях приведены в Приложении 1.

Приложение 1

library("lmtest")

data=swiss

help(swiss)

data$Catholic

mean(data$Catholic)

var(data$Catholic)

sd(data$Catholic)

data$Agriculture

mean(data$Agriculture)

var(data$Agriculture)

sd(data$Agriculture)

data$Examination

mean(data$Examination)

var(data$Examination)

sd(data$Examination)

data

plot(data$Catholic,data$Agriculture)

plot(data$Catholic,data$Examination)

data["Catholic1"]=data$Catholic - mean(data$Catholic)

data$Catholic1

data["Agriculture1"]=data$Agriculture - mean(data$Agriculture)

data$Agriculture1

data["Examination1"]=data$Examination - mean(data$Examination)

data$Examination1

plot(data$Catholic1,data$Agriculture1)

plot(data$Catholic1,data$Examination1)

data["Catholic2"]=data$Catholic1/sqrt(var(data$Catholic))

data$Catholic2

data["Agriculture2"]=data$Agriculture1/sqrt(var(data$Agriculture))

data$Agriculture2

data["Examination2"]=data$Examination1/sqrt(var(data$Examination))

data$Examination2

plot(data$Catholic2~data$Agriculture2) + abline(a = 0, b = 0, col = "red")

plot(data$Catholic2~data$Examination2) + abline(a = 0, b = 0, col = "red")

model1=lm(Catholic~Agriculture,data)

model1

summary(model1)

plot(data$Catholic~data$Agriculture) + abline(a = 3.8313, b = 0.7365 , col = "red")

model2=lm(Catholic~Examination,data)

model2

summary(model2)

plot(data$Catholic~data$Examination) + abline(a = 90.514, b = -2.994, col = "red")