Титульный лист

# **Оглавление**

[Оглавление 1](#_Toc1411991669)

[Задача №1 2](#_Toc715264381)

[Заключение к заданию №1 2](#_Toc1991766249)

[Список литературы 3](#_Toc2078410582)

[Приложение 4](#_Toc1546170724)

[Приложение к задаче №1 4](#_Toc1130227294)

[Приложение к задаче №2 6](#_Toc1602653238)

# **Задача №1**

Необходимо загрузить данные из указанного набора и произвести следующие действия.

Набор данных: Swiss.

Объясняемая переменная: *Infant.Mortality.*

Регрессоры: *Agriculture, Examination*.

1. Оценить среднее значение, дисперсию и СКО переменных.

Используя встроенную функцию *mean*, *var*, *sd*, получаем

Таблица 1. Среднее значение, дисперсия и СКО *Agriculture, Examination* и *Infant.Mortality.*

|  |  |  |  |
| --- | --- | --- | --- |
| Переменная | Среднее значение | Дисперсия | СКО |
| *Agriculture* | 50.65957 | 515.7994 | 22.71122 |
| *Examination* | 16.48936 | 63.64662 | 7.977883 |
| *Infant.Mortality* | 19.94255 | 8.483802 | 2.912697 |

1. Построить зависимости вида y = a + bx, где y – объясняемая переменная, x – регрессор. Оценить, насколько «хороша» модель по коэффициенту детерминации R2. Оценить, есть ли взаимосвязь между объясняемой переменной и объясняющей переменной

Строим модель *Infant.Mortality ~ Agriculture.* Получаем уравнение вида

*Infant*.*Mortality* = 20.338 - 0.008 \* *Agriculure.* R2 равен 0.003704 (0,4%), количество звездочек у регрессора равно нулю. Это говорит о том, что между регрессором и объясняемым значением нет зависимости.

Теперь построим модель *Infant.Mortality~Examination.* Получаем уравнение вида

*Infant.Mortality = 20.62899 -0.04163 \* Examination*. R2 равен 0.013 (1,3 %), количество звездочек у регрессора равно нулю. Это указывает на то, что модель также не объясняет нашу переменную

Код решения задачи и сведения о проверенных моделях приведены в [Приложении к задаче 1](#Bookmark1)

# **Заключение к заданию №1**

1. Построенные модели не отображают зависимости детской смертности от процента мужского населения, занимающегося сельским хозяйством, и от процента людей, получивших максимальный балл на экзаменах при поступлении на военную службу.
2. Улучшить модели невозможно -- нужно менять регрессоры.
3. Из модели мы можем понять, что причиной низкой или высокой детской смертности может быть что-то другое, а не данные параметры.

# **Список** **литературы**

1

# **Приложение**

# Приложение к задаче №1

|  |
| --- |
| library("lmtest")    data = swiss  #используем набор данных swiss  #переменные:  # объясняемая: Infant.Mortality  # объясняющие: Agriculture | Examination  #задания:  #1) Оценить среднее значение, дисперсию и СКО переменных  mean(data$Infant.Mortality)  # 19.94255  mean(data$Agriculture)  # 50.65957  mean(data$Examination)  # 16.48936 процент людей занимающихся фермерским хозяйством больше, чем получивших высокие баллы на экзамене    var(data$Infant.Mortality)  # 8.483802 дисперсия мала, отклонение от среднего не велико  var(data$Agriculture)  # 515.7994 большая (первышает 100) дисперсия, отклонение от среднего может быть велико  var(data$Examination)  # 63.64662 средняя (меньше 100, больше 10) дисперсия...    sd(data$Infant.Mortality)  # 2.912697 СКО малое...  sd(data$Agriculture)  # 22.71122 СКО большое...  sd(data$Examination)  # 7.977883 СКО не большое...    #2) Построить зависимости вида y = a + bx 3) и 4) анализ моделей, оценка зависимости объясняемой переменной от объясняющих |

|  |
| --- |
| model1 = lm(Infant.Mortality~Agriculture,data)  summary(model1)  #Infant.Mortality = 20.338 - 0.008 \* Agriculure  #зависимость отрицательная  #Pr(agriculture) = 0.684 (>0.005)  #причинно-следственной связи нет  #Pr(Coef) <2e-16  #есть зависимость от коэф-та    #p-value: 0.6845 > 0.05 (слишком большой)  #Multiple R-squared: 0.003704, Adjusted R-squared: -0.01844  #зависимости нет. модель не отображает действительности  model2 = lm(Infant.Mortality~Examination,data)  summary(model2)  #Infant.Mortality = 20.62899 -0.04163 \* Examination  # зависимость отрицательная  #Pr(Examination) = 0.445 (>0.005)  #причинно-следственной связи нет  #Pr(Coef) <2e-16  #есть зависимость от коэф-та    #p-value: 0.4454 > 0.05 (слишком большой)  #Multiple R-squared: 0.013, Adjusted R-squared: -0.008932  #зависимости нет. модель не отображает действительности    #Вывод:  #у Infant.Mortality нет реальных зависимостей от Examination и Agriculture  #детская смертность абосолютно не зависит от того, как были сданы экзамены, и от того, аграрный район это или нет  #причиной этого может быть то, что уровень медицины не отличался для людей работающих аграрных регионах и в армии    #Результаты:  #мы не получили моделей, способных предсказывать значение Infant.Mortality  #полученные модели (model1 | model2) не отображают действительности |

# Приложение к задаче №2

-

1. -