Задача 2.2

Условие

Набор данных: swiss.

Объясняемая переменная: Infant.Mortality.

Регрессоры: *Fertility, Catholic, Agriculture*.

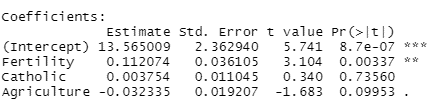
Для зависимости, построенной при решении практического задания №2, оцените:

1. Доверительные интервалы для всех коэффициентов в модели, p = 95%.
2. Сделайте вывод о отвержении или невозможности отвергнуть статистическую гипотезу о том, что коэффициент равен 0
3. Доверительный интервал для одного прогноза (p = 95%, набор значений регрессоров выбираете сами).

Решение

Имеем следующую модель:

Таблица 1. Характеристики модели зависимости параметра: *Infant.Mortality* от параметров *Fertility*, *Catholic*, *Agriculture* в наборе данных Swiss.



1. Оценим доверительные интервалы для всех коэффициентов в модели (для p=95%):

Число степеней свободы в модели df = 43 - 4 = 39, и t-критерий Стьюдента тогда равен 2.022691.

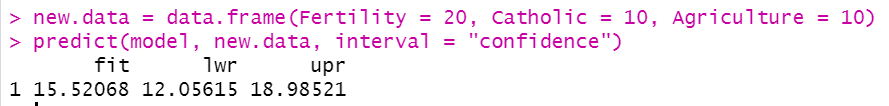
Зная стандартные ошибки (2.3629 для свободного коэффициента, 0.0361 для *Fertility*, 0.0110 для *Catholic*, 0.0192 для *Agriculture*, как видно из таблицы) найдём доверительные интервалы для коэффициентов по формуле Estimate t-критерий Стьюдента \* Std. Error:

* Доверительный интервал свободного коэффициента: [8.79, 18.34]
* Доверительный интервал *Fertility*: [0.04, 0.19]
* Доверительный интервал *Catholic*: [ -0.02, 0.03]
* Доверительный интервал *Agriculture*: [-0.07,0.01]

1. По доверительным интервалам сделаем вывод об отвержении или невозможности отвергнуть статистическую гипотезу о том, что коэффициент равен 0:

* Доверительный интервал свободного коэффициента: [8.79, 18.34], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Fertility*: [0.04, 0.19], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Catholic*: [ -0.02, 0.03], 0 попадает в доверительный интервал, поэтому невозможно отвергнуть статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Agriculture*: [-0.07,0.01], 0 попадает в доверительный интервал, поэтому невозможно отвергнуть статистическую гипотезу о том, что коэффициент равен 0.

1. Оценим доверительный интервал для одного прогноза (p = 95%, *Fertility* = 20, *Catholic* = 10, *Agriculture* = 10), используя команду predict:

  
Рисунок 1. Оценка доверительного интервала с помощью команды predict

Имеем доверительный интервал [12.05615, 18.98521]

Полный код решения задачи приведён в Приложении 1.

Выводы

Есть вероятность, что коэффициенты *Catholic* и *Agriculture* равны нулю, так как ноль попадает в их доверительные интервалы, что означает отсутствие взаимосвязи с объясняемой переменной.

Доверительный интервал для ненулевых регрессоров небольшой, что свидетельствует о сильной взаимосвязи между ними и объясняемой переменной *Infant.Mortality.*

Поскольку доверительный интервал для прогноза со значениями *Fertility* = 20, *Catholic* = 10, *Agriculture* = 10 получился достаточно большой, модель сложно назвать хорошей.

Приложение 1

Код решения задачи:

1. library("lmtest")
2. library("GGally")
3. library("car")
5. data = na.omit(swiss)
6. model = lm(Infant.Mortality ~ Fertility + Catholic + Agriculture, data)
7. summary(model)
8. # Число степеней свободы в моделе df =  43 - 4 = 39
9. t\_critical = qt(0.975, df = 39) #t-критерий Стьюдента - общий для всех коэффициентов
10. # Стандартные ошибки коэффициентов (взяты из summary(model)):
11. Std\_Error\_Intercept = 2.3629
12. Std\_Error\_Fertility = 0.0361
13. Std\_Error\_Catholic = 0.0110
14. Std\_Error\_Agriculture = 0.0192
15. print(paste("Доверительный интервал свободного коэффициента: [", model$coefficients[1] - t\_critical \* Std\_Error\_Intercept,
16. ",", model$coefficients[1] + t\_critical \* Std\_Error\_Intercept, "]"))
18. print(paste("Доверительный интервал Fertility: [", model$coefficients[2] - t\_critical \* Std\_Error\_Fertility,
19. ",", model$coefficients[2] + t\_critical \* Std\_Error\_Fertility, "]"))
21. print(paste("Доверительный интервал Catholic: [", model$coefficients[3] - t\_critical \* Std\_Error\_Catholic,
22. ",", model$coefficients[3] + t\_critical \* Std\_Error\_Catholic, "]"))
24. print(paste("Доверительный интервал Agriculture: [", model$coefficients[4] - t\_critical \* Std\_Error\_Agriculture,
25. ",", model$coefficients[4] + t\_critical \* Std\_Error\_Agriculture, "]"))
26. new.data = data.frame(Fertility = 20, Catholic = 10, Agriculture = 10)
27. predict(model, new.data, interval = "confidence")