Задача 2.2

Набор данных: swiss.

Объясняемая переменная: Infant.Mortality .

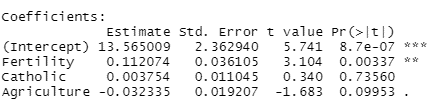
Регрессоры: *Infant.Mortality , Catholic, Catholic*.

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).

Имеем следующую модель:

Таблица 1. Характеристики модели зависимости параметра : *Infant.Mortality* от параметров *Infant.Mortality* , *Catholic*, *Catholic* 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 *Infant.Mortality*, 0.0110 для *Catholic*, 0.0192 для *Catholic*, как видно из таблицы) найдём доверительные интервалы для коэффициентов по формуле Estimate t-критерий Стьюдента \* Std. Error:

* Доверительный интервал свободного коэффициента: [ 8.7855926699545 , 18.3444254198642 ]
* Доверительный интервал *Infant.Mortality* : [ 0.0390550318940447 , 0.185093316320699 ]
* Доверительный интервал *Catholic*: [ -0.0184955828737981 , 0.0260036173670106 ]
* Доверительный интервал *Catholic*: [ -0.0711707027421142 , 0.0065006285872974 ]

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

* Доверительный интервал свободного коэффициента: [ 8.7855926699545 , 18.3444254198642 ], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Infant.Mortality* : [ 0.0390550318940447 , 0.185093316320699 ], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Catholic*: [ -0.0184955828737981 , 0.0260036173670106 ], 0 попадает в доверительный интервал, поэтому невозможно отвергнуть статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Catholic*: [ -0.0711707027421142 , 0.0065006285872974], 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%, *Infant.Mortality*  = 20, *Catholic* = 10, *Catholic* = 10), используя команду predict:

Доверительный интервал - [12.05615, 18.98521]

Полный код решения задачи приведён в Приложении 1.

Приложение 1

library("lmtest")

library("GGally")

library("car")

data = na.omit(swiss)

model = lm(Infant.Mortality ~ Fertility + *Catholic* + *Catholic*, data)

summary(model)

# Число степеней свободы в моделе df = 43 - 4 = 39

t\_critical = qt(0.975, df = 39) #t-критерий Стьюдента - общий для всех коэффициентов

# Стандартные ошибки коэффициентов (взяты из summary(model)):

Std\_Error\_Intercept = 2.3629

Std\_Error\_Fertility = 0.0361

Std\_Error\_*Catholic* = 0.0110

Std\_Error\_*Catholic* = 0.0192

print(paste("Доверительный интервал свободного коэффициента: [", model$coefficients[1] - t\_critical \* Std\_Error\_Intercept,

",", model$coefficients[1] + t\_critical \* Std\_Error\_Intercept, "]"))

print(paste("Доверительный интервал Fertility: [", model$coefficients[2] - t\_critical \* Std\_Error\_Fertility,

",", model$coefficients[2] + t\_critical \* Std\_Error\_Fertility, "]"))

print(paste("Доверительный интервал *Catholic*: [", model$coefficients[3] - t\_critical \* Std\_Error\_*Catholic*,

",", model$coefficients[3] + t\_critical \* Std\_Error\_*Catholic*, "]"))

print(paste("Доверительный интервал *Catholic*: [", model$coefficients[4] - t\_critical \* Std\_Error\_*Catholic*,

",", model$coefficients[4] + t\_critical \* Std\_Error\_*Catholic*, "]"))

new.data = data.frame(Fertility = 20, *Catholic* = 10, *Catholic* = 10)

predict(model, new.data, interval = "confidence")

Заключение

Есть вероятность, что коэффициенты *Catholic* и *Catholic* равны нулю, так как ноль попадает в их доверительные интервалы. В целом, все доверительные интервалы не слишком широкие.