Задача 2.2

Условие

Набор данных: swiss.

Объясняемая переменная: *Agriculture*.

Регрессоры: *Fertility, Catholic, Education*.

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:

Характеристики модели зависимости параметра: *Agriculture* от параметров *Fertility*, *Education*, *Catholic* в наборе данных Swiss. Зная стандартные ошибки (20.8456 для свободного коэффициента, 0.2748 для *Fertility*, 0.0623 для *Catholic*, 0.3201 для *Education*) найдём доверительные интервалы для коэффициентов по формуле Estimate t-критерий Стьюдента \* Std. Error: (*см. Таблица 1*)

* *Fertility*:

Критерий Стьюдента(t) = 2.02

Доверительный интервал: [-1.34, -0.23]

* *Edecation*:

Критерий Стьюдента(t) = 2.02

Доверительный интервал: [-2.66, -1.37]

* *Catholic*:

Критерий Стьюдента(t) = 2.02

Доверительный интервал: [0.13, 0.38]

* *Intercept*:

Критерий Стьюдента(t) = 2.02

Доверительный интервал: [75.1, 159.45]

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.

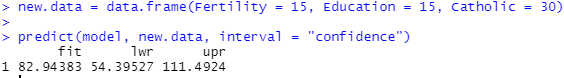
Таблица 1. Характеристики модели зависимости параметра Agriculture от параметров Fertility, Education, Catholic в наборе данных Swiss

|  |  |  |  |
| --- | --- | --- | --- |
| Параметр \ Характеристики | Значение | Доверительный интервал | “Может ли коэффициент быть равен 0?” |
| (Intercept) | 117.2729 | [75.1, 159.45] | Нет |
| *Fertility* | -0.7842 | [-1.34, -0.23] | Нет |
| *Education* | -2.0161 | [-2.66, -1.37] | Нет |
| *Catholic* | 0.2558 | [0.13, 0.38] | Нет |

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

* Доверительный интервал свободного коэффициента: [75.1, 159.45], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Fertility*: [-1.34, -0.23], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Catholic*: [0.13, 0.38], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.
* Доверительный интервал *Education*: [-2.66, -1.37], 0 не попадает в доверительный интервал, поэтому отвергаем статистическую гипотезу о том, что коэффициент равен 0.

1. Оценим доверительный интервал для одного прогноза (p = 95%, *Fertility = 15, Education = 15, Catholic = 30*):

  
Рисунок 1. Оценка доверительного интервалa.

Имеем доверительный интервал [54.39527, 111.4924]

Полный код решения задачи приведён в Приложении 1.

Выводы*:*

Поскольку 0 не попадает в доверительный интервал регрессоров*,* то объясняющая переменная связана с объясняемыми переменными*.*

Прогноз модели Agriculture ~ Fertility, Education, Catholic оценивается как 82.94383 Доверительный интервал для свободного коэффициента имеет вид: [54.4, 111.5].

Приложение 1

Код решения задачи:

library("lmtest")

library("GGally")

library("car")

data = swiss

help(swiss)

#Вариант 2 Бундуки Владислав КМБО-06-20

#Объясняемая переменная Agriculture , регрессоры: Fertility, Education, Catholic

model = lm(Agriculture ~ Fertility + Education + Catholic, data)

model

summary(model)

#Agriculture = 117.2729 + (-0.78)\*Fertility + (-2.01)\*Edu + (0.26)Cat

#R^2 = 0.582

# 43 наблюдений и оценивалось 4 коэффициента: 43 - 4 = 39 степени свободы;

#Оценим доверительные интервалы для Agriculture = 117.2729 + (-0.78)\*Fertility + (-#2.01)\*Edu + (0.26)Cat:

# a) Доверительный интервал для Fertility

Std\_Err = 0.2748

# Критерий Стьюдента: 95%, 39 степени свободы

t\_critical = qt(0.975, df = 39) # ~2.023

model$coefficients[2] - t\_critical \* Std\_Err

model$coefficients[2] + t\_critical \* Std\_Err

# Доверительный интервал для Fertility = [model$coefficients[2] - t\_critical \* Std\_Err , #model$coefficients[2] + t\_critical \* Std\_Err]

# [-0.78 - 2.023 \* 0.2748 , -0.78 + 2.023 \* 0.2748]

# [-1.34 , -0.23] -> коэффициент не может быть равным 0

# Проверка:

confint(model, level = 0.95)

# 2.5 % 97.5 %

#(Intercept) 75.233631 159.3121545

#Fertility -1.338392 -0.2299865

#Education -2.661649 -1.3704634

#Catholic 0.130171 0.3814702

# б) Доверительный интервал для Education

Std\_Err = 0.3201

# Критерий Стьюдента: 95%, 39 степени свободы

model$coefficients[3] - t\_critical \* Std\_Err

model$coefficients[3] + t\_critical \* Std\_Err

# Доверительный интервал для Education = [model$coefficients[3] - t\_critical \* Std\_Err #, model$coefficients[3] + t\_critical \* Std\_Err]

# [-2.01 - 2.023 \* 0.3201 , -2.01 + 2.023 \* 0.3201]

# [-2.66 , -1.37] -> коэффициент не может быть равным 0

# в) Доверительный интервал для Catholic

Std\_Err = 0.0623

# Критерий Стьюдента: 95%, 39 степени свободы

model$coefficients[4] - t\_critical \* Std\_Err

model$coefficients[4] + t\_critical \* Std\_Err

# Доверительный интервал для Education = [model$coefficients[4] - t\_critical \* Std\_Err #, model$coefficients[4] + t\_critical \* Std\_Err]

# [0.26 - 2.023 \* 0.0623 , 0.26 + 2.023 \* 0.0623]

# [0.13 , 0.38] -> коэффициент не может быть равным 0

# г) доверительный интервал для свободного коэф:

Std\_Err = 20.85

# Критерий Стьюдента: 95%, 39 степени свободы

model$coefficients[1] - t\_critical \* Std\_Err

model$coefficients[1] + t\_critical \* Std\_Err

# Доверительный интервал для Intercept = [model$coefficients[1] - t\_critical \* Std\_Err #, model$coefficients[1] + t\_critical \* Std\_Err]

# [117.2729 - 2.023 \* 20.85 , 117.2729 + 2.023 \* 20.85]

# [75.1 , 159.45] -> коэффициент не может быть равным 0

# Вывод: Поскольку 0 не попадает в доверительный интервал регрессоров => все #регрессоры связаны с объясняемой переменной.

#Построим доверительный интервал для прогноза (Объясняемая переменная: #Agriculture, регрессоры: Fertility , Education , Catholic )

model = lm(Agriculture ~ Fertility + Education + Catholic, data)

model

summary(model)

#Agriculture = 117.2729 + (-0.78)\*Fertility + (-2.01)\*Edu + (0.26)Cat

new.data = data.frame(Fertility = 15, Education = 15, Catholic = 30)

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

# fit lwr upr

# 82.94383 54.39527 111.4924

# Вывод: Прогноз модели оценивается как 82.94383;

# Доверительный интервал для свободного коэффициентa = [54.39527,111.4924]