

Формальности

На экзамен выделено 120 минут времени, плюс дополнительные 5 минут на загрузку работы. Можно пользоваться лекциями и любыми источниками, общаться с другими лицами во время экзамена нельзя. Экзамен проводится без прокторинга. Ответ без решения не засчитывается.

Более сложные задачи отмечены звёздочкой в начале условия.

Как сдавать ответ на «ручные» вопросы?

Написать текст на листочках и прикрепить фотографии. Также можно писать текст на планшете и прикрепить готовый pdf.

Как сдавать ответ на компьютерные вопросы?

Скопировать предложенный шаблон ответа в Rstudio. Ниже каждого пункта привести код, требующийся для решения пункта задачи. После кода нужно привести чёткий ответ на поставленный вопрос в виде комментария после #.

Пример условия шаблона:

a) Найдите косинус числа 42.

...

Косинус равен ...

Пример ответа:

a) Найдите косинус числа 42.

`cos(42)`

Косинус равен -0.4.

Затем полученный файл .R надо сохранить и прикрепить на платформе. Один файл .R на всю пятую задачу и один — на шестую.

Ни пуха, ни пера!

1. Исследователь Винни-Пух оценил модель множественной регрессии.

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i.$$

В качестве $\hat{\beta}_2$ возьмите количество букв в своей фамилии, в качестве $se(\hat{\beta}_2)$ — количество букв в своём имени.

У Винни-Пуха 1000 наблюдений и он считает, что выполнены классические предпосылки на ошибки модели, $\mathbb{E}(u_i | X) = 0$, $\text{Var}(u_i | X) = \sigma^2$, $\text{Cov}(u_i, u_j | X) = 0$.

(a) Постройте 90%-й доверительный интервал для β_2 .

(b) Проверьте гипотезу $H_0 : \beta_2 = 0$ против альтернативной $\beta_2 > 0$ на уровне значимости 10%.

2. Заполните пропуски Q1-Q9, пояснив формулами как конкретно, и в каком порядке они заполнялись.

Показатель	Значение					
Выборочная корреляция прогнозов и зависимой переменной	Q1					
R^2	Q2					
Оценка дисперсии ошибки	Q3					
Число наблюдений	800					
RSS	37					
ESS	42.9					
TSS	Q5					

	Coef.	St. error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-25.24	2.0	Q6	0	Q7	-21.31
totspan	1.7	Q8	30.4	0	Q9	Q4

3. (*) Регрессию оценили с помощью МНК по 50 наблюдениям:

$$\hat{y}_i = -17.3 + 3.92x_i - 0.63z_i.$$

В исходной модели $y_i = \beta_1 + \beta_2x_i + \beta_3z_i + u_i$ ошибки удовлетворяют предпосылкам теоремы Гаусса-Маркова и нормально распределены.

Матрица регрессоров X с первым столбцом из единиц такова, что

$$(X^T X)^{-1} = \begin{pmatrix} 0.196 & -0.0114 & -0.008 \\ -0.0114 & 0.00074 & 0.00041 \\ -0.008 & 0.00041 & 0.0208 \end{pmatrix}.$$

Сумма квадратов остатков модели равна 11334.

- (a) Оцените дисперсию случайной ошибки.
 - (b) Найдите стандартную ошибку для $\hat{\beta}_2$.
 - (c) На уровне значимости 10% протестируйте значимость оценки $\hat{\beta}_2$.
 - (d) На уровне значимости 10% протестируйте гипотезу $H_0: \beta_2 + \beta_3 = 3$.
4. (*) Богиня случая Тюхе хочет заглянуть в прошлое и в будущее.
- У нее есть 100 наблюдений за ежегодной урожайностью Волшебной Яблони, $x_1 = 160, x_2 = 142, \dots, x_{99} = 141, x_{100} = 158, \bar{x} = 149.76$.
- Тюхе оценила простую парную регрессию урожая на урожай прошлого года (здесь $t = 2, \dots, 100$):

$$\hat{x}_t = 174.9010 - 0.1687x_{t-1}.$$

- (a) Загляните в будущее найдя прогноз \hat{x}_{101} .
 - (b) По имеющимся данным определите, какие коэффициенты получатся в регрессии $\hat{x}_t = \hat{\gamma}_1 + \hat{\gamma}_2 x_{t+1}$ при $t = 1, \dots, 99$.
 - (c) Загляните в прошлое найдя прогноз \hat{x}_0 .
-

5. Мы будем работать с набором данных `LifeCycleSavings`.

Шаблон с вопросами и пропусками для вашего кода и ответов приведён ниже.

```
library(tidyverse)
library(skimr)

d = LifeCycleSavings

# а) Сколько наблюдений в наборе данных LifeCycleSavings?

# ...

# В наборе данных ... наблюдений.

# б) Чему равно среднее значение личных сбережений sr
# по всем странам?
# (в данных sr измерено в процентах от располагаемого дохода)

# ...

# Среднее значение sr равно ...

# в) Постройте регрессию размера личных сбережений sr на
# константу, долю населения младше 15 лет pop15,
# долю населения старше 75 pop75.
# Выпишите полученное уравнение регрессии.

# ...

#  $\hat{sr}_i = \dots + \dots + \dots$ 

# г) Постройте 99%-й доверительный интервал для коэффициента
# при pop15.

# ...

# Интервал: [..., ...]

# д) Какой прогноз личных сбережений для страны
# с 30% населения младше 15 лет и 3% населения старше 75 лет
# даёт данная модель?

# ...

# Прогноз равен ...
```

6. (*) Мы продолжаем работать с набором данных `LifeCycleSavings`.

Шаблон с вопросами и пропусками для вашего кода и ответов приведён ниже.

```
library(tidyverse)
library(skimr)

d = LifeCycleSavings

# а) Для переменных sr, por15, por75 посчитайте логарифмы.
# В качестве ответа приведите наибольшее значение
# логарифма переменной por15.

# ...

# Наибольшее значение логарифма равно ...

# б) Постройте регрессию логарифма размера личных сбережений  $\log(sr)$  на
# константу,  $\log(por15)$  и  $\log(por75)$ .
# Выпишите полученное уравнение регрессии.

# ...

#  $\log(sr)^{\text{hat}}_i = \dots + \dots + \dots$ 

# в) Найдите оценку ковариационной матрицы оценок коэффициентов
# В качестве ответа выпишите
# оценку ковариации коэффициентов при  $\log(por15)$  и  $\log(por75)$ 

# ...

# Оценка ковариации равна ...

# г) Постройте 99%-й доверительный интервал для
# разницы коэффициентов при  $\log(por15)$  и  $\log(por75)$ 

# ...

# Интервал: [..., ...]

# д) Какой прогноз личных сбережений (не логарифма!) для страны
# с 30% населения младше 15 лет и 3% населения старше 75 лет
# даёт данная модель?

# ...

# Прогноз равен ...
```
