

## 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  $\beta_2$ .

(b) Проверьте гипотезу  $H_0 : \beta_2 = 3$  против альтернативной  $\beta_2 > 3$  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  $\hat{\beta}_3$ .
  - (c) На уровне значимости 10% протестируйте значимость оценки  $\hat{\beta}_3$ .
  - (d) На уровне значимости 10% протестируйте гипотезу  $H_0: \beta_2 = 2\beta_3$ .
4. (\*) Придумайте и явно выпишите набор данных, для которого одновременно выполняются три свойства. Регрессия  $y_i$  на  $x_i$  по первой половине данных даёт положительную оценку коэффициента при  $x_i$ . Регрессия  $y_i$  на  $x_i$  по второй половине данных даёт положительную оценку коэффициента при  $x_i$ . Регрессия  $y_i$  на  $x_i$  по всему набору даёт отрицательную оценку коэффициента при  $x_i$ .

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) Отберите наблюдения с
# подушевым доходом dpi более медианного.
# Сколько наблюдений в полученном наборе данных?

# ...

# В наборе данных ... наблюдений.

# Далее работаем с полученным набором данных.
# б) Чему равно максимальное значение переменной sr?
# (в данных sr измерено в процентах от располагаемого дохода)

# ...

# Максимальное 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 = ... + ... + ...

# г) Постройте 99%-й доверительный интервал для коэффициента
# при pop75.

# ...

# Интервал: [..., ...]

# д) Какой прогноз личных сбережений для страны
# с 20% населения младше 15 лет и 10% населения старше 75 лет
# даёт данная модель?

# ...

# Прогноз равен ...
```

---

6. (\*) Вернёмся к набору данных LifeCycleSavings.

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

```
library(tidyverse)
library(skimr)

d = LifeCycleSavings

# а) Для переменной sr посчитайте логарифм.
# В качестве ответа приведите наибольшее значение
# логарифма.

# ...

# Наибольшее значение логарифма равно ...

# б) Постройте регрессию логарифма размера личных сбережений  $\log(sr)$  на
# константу, pop15, pop75 и их произведение.
# Выпишите полученное уравнение регрессии.

# ...

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

# в) Найдите оценку ковариационной матрицы оценок коэффициентов
# В качестве ответа выпишите
# оценку ковариации коэффициентов при pop15 и pop75.

# ...

# Оценка ковариации равна ...

# г) Постройте 99%-й доверительный интервал для
# разницы коэффициентов при pop15 и pop75.

# ...

# 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 ...
```

---