

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 оценивает модель $y_i = \beta_1 + \beta_2 z_i + u_i$ зависимости роста зелёных человечков y_i от их зелёности z_i по разным наборам данных:

Набор данных	Уравнение	n	RSS
A	$\hat{y}_i = 5 + 2z_i$	100	200
A и ещё одно наблюдение	$\hat{y}_i = 5.2 + 2.3z_i$	101	204

- (а) Проверьте гипотезу о том, что зависимость одинакова для наборов данных A и B .
 (б) Найдите оценки коэффициентов в регрессии

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 z_i + \hat{\beta}_3 b_i,$$

где b_i — дамми-переменная равная 1 для одного добавленного наблюдения.

2. Пират Джо изучает зависимость веса y_i золотой монеты от её номинала x_i . Джо предполагает, что количество золота должно быть прямо пропорционально номиналу, но также допускает, что монеты помимо золота содержат постоянное количество меди. Джо оценивает модель $y_i = \beta_1 + \beta_2 x_i + u_i$ с $\mathbb{E}(u_i | x_i) = 0$.

Также Джо заметил, что абсолютная ошибка весов пропорциональна квадрату весу предмета, а веса равновероятно ошибаются в плюс или в минус.

- (а) Какая зависимость $\text{Var}(u_i | x_i)$ от x_i следует из предположений Джо?
 (б) Какую регрессию стоит оценить Джо, чтобы получить эффективные оценки коэффициентов?
 (с) Как Джо проинтерпретировать оценки $\hat{\beta}_1$ и $\hat{\beta}_2$?
-

3. (*) Исследователь Винни-Пух изучает зависимость бинарной переменной правильности мёда y_i от бинарной переменной правильности пчёл x_i по таблице сопряжённости:

	Правильные пчёлы	Неправильные пчёлы
Правильный мёд	50	60
Неправильный мёд	100	90

- (a) Оцените логистическую модель $\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$.
- (b) Постройте прогноз разницы вероятности правильности мёда у правильных и неправильных пчёл.
4. (*) Широко известно, что оценка студента по эконометрике, res_i , зависит от числа выпитых чашек кофе, cup_i :

$$res_i = \beta_1 + \beta_2 cup_i + u_i.$$

У преподавателя есть данные по res_i , однако величина cup_i не наблюдаема. Преподаватель смог опросить опросить студентов и получил опросные данные по cup_i^* — заявленному числу выпитых чашек кофе, и по $cake_i^*$ — заявленному числу съеденных чиз-кейков.

Других данных у исследователя нет. Исследователь предполагает, что $cup_i^* = cup_i + w_i$, и $cake_i^* = \gamma_1 + \gamma_2 cup_i^* + \nu_i$.

Наблюдения представляют собой случайную выборку. Величины u_i , w_i , ν_i и cup_i независимы и имеют неизвестные дисперсии σ_u^2 , σ_w^2 , σ_ν^2 , σ_c^2 , величина $\gamma_2 \neq 0$ и так же неизвестна.

- (a) Найдите предел по вероятности МНК оценки $\hat{\beta}_2$ в обычной регрессии $\widehat{res}_i = \hat{\beta}_1 + \hat{\beta}_2 cup_i^*$.
- (b) Найдите предел по вероятности IV оценки $\hat{\beta}_2$ в регрессии $\widehat{res}_i = \hat{\beta}_1 + \hat{\beta}_2 cup_i^*$ с инструментом $cake_i^*$.

5. Мы будем работать с набором данных по дефолтам индивидов `Default`.

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

```
library(tidyverse)
library(skimr)
library(ISLR)

d = Default

# а) Оцените логистическую регрессию дефолта индивида default
# от среднего баланса на счету balance и дохода income.

# ...

# Оценка зависимости имеет вид: ... = ... + ... + ...

# б) Постройте 90%-й доверительный интервал
# для предельного эффекта balance
# у среднестатистического наблюдения.

# ...

# Предельный эффект равен ...

# в) Постройте прогноз вероятности дефолта для индивида
# с максимальным доходом и минимальным балансом в выборке.

# ...

# Прогноз вероятности равен ...

# г) Проверьте гипотезу о том, что модель идентична
# для студентов и не студентов против гипотезы о том,
# что модель может отличаться всеми коэффициентами.

# ...

# Значение статистики ..., Р-значение равно ...
# Статистический вывод ...
```

6. Мы продолжаем работать с набором данных по дефолтам индивидов `Default`.

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

```
library(tidyverse)
library(skimr)
library(ISLR)

d = Default

# а) Оцените линейную модель зависимости дохода income от
# баланса на карте balance и статуса студента student.
# Используйте робастные стандартные ошибки.

# ...

# Оценка зависимости имеет вид: ... = ... + ... + ...

# б) Постройте 95%-й робастный доверительный интервал для
# коэффициента при балансе.

# ...

# Интервал имеет вид [...; ...]

# в) Постройте 95%-й робастный доверительный интервал для ожидаемого
# дохода для студента со средним балансом на карточке.

# ...

# Интервал имеет вид [...; ...]

# г) Проверьте гипотезу о том, что модель идентична для индивидов
# с дефолтом и без против гипотезы о том, что модель может
# отличаться константой и коэффициентом при балансе.

# ...

# Значение статистики ..., Р-значение равно ...
# Статистический вывод ...

# д) Проверьте гипотезу об отсутствии гетероскедастичности с помощью
# теста Уайта. Используйте уровень значимости 10%.

# ...

# Значение статистики ..., Р-значение равно ...
# Статистический вывод ...
```
