

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	$\hat{y}_i = 4 + 2z_i$	100	200
B	$\hat{y}_i = 5 + 4z_i$	100	300
A и B	$\hat{y}_i = 2 + 3z_i$	200	700

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

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 a_i + \hat{\beta}_3 z_i + \hat{\beta}_4 z_i b_i,$$

где b_i — дамми-переменная равная 1 для зелёных человечков из группы B , и a_i — дамми-переменная равная 1 для человечков из группы A .

2. Пират Джо изучает зависимость веса y_i ящика с золотыми монетами от количества золотых монет в ящике x_i . Джо предполагает простую зависимость $y_i = \beta_1 + \beta_2 x_i + u_i$ с $\mathbb{E}(u_i | x_i) = 0$.

Также Джо считает, что веса всех монет независимы и одинаково распределены.

- (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	40	70
Неправильный мёд	100	90

- (a) Оцените линейную модель $y_i = \beta_1 + \beta_2 x_i + u_i$ с помощью линейной регрессии.
- (b) Оцените логистическую модель $\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$.
- (c) Постройте прогнозы вероятности правильности мёда у правильных пчёл по обоим моделям.
4. (*) Широко известно, что оценка студента по эконометрике, res_i , зависит от числа выпитых чашек кофе, c_i :

$$res_i = \beta_1 + \beta_2 c_{ip_i} + u_i.$$

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

Других данных у исследователя нет. Исследователь предполагает, что $c_{ip_i}^* = c_{ip_i} + w_i$, и $cake_i^* = \gamma_1 + \gamma_2 c_{ip_i}^2 + \nu_i$.

Наблюдения представляют собой случайную выборку. Величины u_i , w_i , ν_i и c_{ip_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 $\widehat{res}_i = \hat{\beta}_1 + \hat{\beta}_2 c_{ip_i}^*$.
- (b) Предложите состоятельную оценку для неизвестного параметра β_2 и докажите её состоятельность.

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

Шаблон с вопросами и пропусками для вашего кода и ответов приведен 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 для переменной income для медианного наблюд
```

```
# ...
```

```
# Средний предельный эффект равен ...
```

```
# в) Постройте прогноз вероятности дефолта для индивида
# со средним доходом и средним балансом.
```

```
# ...
```

```
# Прогноз вероятности равен ...
```

```
# г) Проверьте гипотезу о том, что модель идентична
# для студентов и не студентов против гипотезы о том,
# что модель может отличаться коэффициентов при константе и при балан
```

```
# ...
```

```
# Значение статистики ..., P-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 90%-й робастный доверительный интервал для
# коэффициента при дамми на статус "студент".

# ...

# Интервал имеет вид [...; ...]

# в) Постройте 90%-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 коэффициентам при константе и балансе.

# ...

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

# д) Проверьте гипотезу об отсутствии гетероскедастичности с помощью
# теста Уайта. Используйте уровень значимости 1%.

# ...

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