Часть 1. Тест.

Вопрос 1 ♣ Если квадраты остатков оценённой с помощью МНК регрессионной модели линейно и значимо зависят от квадрата регрессора Z, то гетероскедастичность можно попытаться устранить,

- A умножив исходное уравнение на Z
- B умножив исходное уравнение на Z^2
- C поделив исходное уравнение на $\sqrt{Z}$
- D умножив исходное уравнение на $\sqrt{Z}$
- [redacted] поделив исходное уравнение на Z
- F поделив исходное уравнение на Z^2
- G *Нет верного ответа.*

Вопрос 2 ♣ Метод максимального правдоподобия для оценки коэффициентов регрессии $Y = X\beta + \varepsilon$ НЕ МОЖЕТ быть применён, если

- A закон распределения вектора ε известен, но не является нормальным
- B $\varepsilon \sim \mathcal{N}(0; \Omega)$ и структура Ω известна, но зависит от набора неизвестных параметров
- C $\varepsilon \sim \mathcal{N}(0; \Omega)$ и $\Omega = b \cdot I$, где b — неизвестный параметр
- [redacted] $\varepsilon \sim \mathcal{N}(0; \Omega)$ и структура Ω неизвестна
- E $\varepsilon \sim \mathcal{N}(0; \Omega)$ и $\Omega = 2017 \cdot I$
- F *Нет верного ответа.*

Вопрос 3 ♣ При наличии сильной практической мультиколлинеарности нарушается следующее свойство МНК-оценок параметров классической регрессии:

- A линейность по зависимой переменой
- B несмещённость
- C эффективность в классе линейных и несмещенных оценок
- D равенство нулю суммы остатков
- [redacted] *Нет верного ответа.*

Вопрос 4 ♣ Оценка максимального правдоподобия параметра λ по случайной выборке X_1, ..., X_n из распределения с функцией плотности

$$f(x|\lambda) = \begin{cases} \lambda^{-1}x^{-1+1/\lambda}, \text{ если } 0 < x < 1; \\ 0, \text{ иначе.} \end{cases}$$

имеет вид:

- A $\hat{\lambda}_{ML} = -\frac{X_1+\ldots+X_n}{n}$
- B $\hat{\lambda}_{ML} = \frac{X_1^2+\ldots+X_n^2}{n}$
- [redacted] $\hat{\lambda}_{ML} = -\frac{\ln X_1+\ldots+\ln X_n}{n}$
- D $\hat{\lambda}_{ML} = \frac{X_1+\ldots+X_n}{n}$
- E $\hat{\lambda}_{ML} = \frac{\ln X_1+\ldots+\ln X_n}{n}$
- F *Нет верного ответа.*

Вопрос 5 ♣ Методом максимального правдоподобия Гоша оценил модель

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_6 X_{i6} + \varepsilon_i,$$

где $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$, по 12 наблюдениям. Оказалось, что $RSS = 24$. Оценка дисперсии случайной составляющей равна

A 0.5

B 0.48

C $24/7$

■ 2

E 2.4

F *Нет верного ответа.*

Вопрос 6 ♣ Имеются данные по 100 работникам: затраты на проезд в общественном транспорте (E_i, руб.), количество часов работы в день (WH_i, руб.), количество часов отдыха в день (LH_i, руб.) и количество часов сна в день (SH_i, руб.). Считая, что всё время суток распределяется между трудом, сном и отдыхом, оценка регрессии в виде

$$E_i = \beta_1 + \beta_2 WH_i + \beta_3 LH_i + \beta_4 SH_i + u_i$$

приведет к тому, что

A МНК-оценки параметров окажутся смещёнными

B коэффициент детерминации R^2 окажется отрицательным

C МНК-оценки параметров регрессии будут несмещенными и эффективными

■ МНК–оценки получить не удастся

E МНК-оценки параметров окажутся неэффективными в классе линейных и несмещённых

F *Нет верного ответа.*

Вопрос 7 ♣ Обобщенный МНК служит для оценивания регрессионных моделей в случае нарушений следующего условия теоремы Гаусса-Маркова:

■ $\mathrm{Var}(u) = \sigma^2 I$

B $\mathrm{rank}\, X = k$

C $\mathbb{E}(u_i) = 0$

D u_i распределены нормально

E Величина Y_i линейна по β_1, β_2, ...

F *Нет верного ответа.*

Вопрос 8 ♣ По $n = 500$ наблюдениям была оценена регрессия:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + u_i.$$

Затем была оценена регрессия $|\hat{u}_i| = \alpha_1 + \alpha_2 \frac{1}{Z_i} + \nu_i$. Оказалось, что $\hat{\alpha}_2 = 24$ и $se(\hat{\alpha}_2) = 5$. Согласно этим данным, на уровне значимости 5% гипотеза о

- A верной функциональной форме не отвергается
- [redacted] гомоскедастичности отвергается
- C пропущенной переменной $1/Z_i$ не отвергается
- D гомоскедастичности не отвергается
- E пропущенной переменной $1/Z_i$ отвергается
- F верной функциональной форме отвергается
- G *Нет верного ответа.*

Вопрос 9 ♣ Василий хочет оценить константу μ в модели $Y_i = \mu + u_i$, где $\mathbb{E}(u_i) = 0, \mathbb{E}(u_i u_j) = 0$ при $i \neq j$, $\mathrm{Var}(u_i) = \sigma^2 X_i$ и $X_i > 0$.
В классе линейных несмещенных оценок наиболее эффективной является:

- [redacted] $\frac{\sum Y_i/\sqrt{X_i}}{\sum 1/X_i}$
- B $\frac{\sum Y_i X_i}{\sum X_i}$
- C $(I'I)^{-1}I'Y$
- D $\frac{\sum Y_i/X_i}{\sum 1/X_i^2}$
- E $\bar{Y}$
- F $\frac{\sum Y_i X_i}{\sum X_i^2}$
- G *Нет верного ответа.*

Вопрос 10 ♣ Оценки коэффициентов линейной регрессии, полученные методом максимального правдоподобия и методом наименьших квадратов в случае нормально распределенной случайной составляющей, будут совпадать

- [redacted] если ковариационная матрица случайной составляющей пропорциональна единичной
- B никогда
- C если ковариационная матрица случайной составляющей нулевая
- D всегда
- E если ковариационная матрица случайной составляющей диагональна
- F *Нет верного ответа.*

Вопрос 1 : A B C D ■ F G

Вопрос 2 : A B C ■ E F

Вопрос 3 : A B C D ■

Вопрос 4 : A B ■ D E F

Вопрос 5 : A B C ■ E F

Вопрос 6 : A B C ■ E F

Вопрос 7 : ■ B C D E F

Вопрос 8 : A ■ C D E F G

Вопрос 9 : ■ B C D E F G

Вопрос 10 : ■ B C D E F

Часть 2. Задачи.

1. По данным для 39 районов Балтимора в 1970 г. были оценены уравнения

$$\ln \hat{Y}_i = \underset{t=54.7}{10.093} - \underset{t=-12.28}{0.239} X_i, \quad R^2 = 0.803$$

и

$$\frac{\ln \hat{Y}_i}{\sqrt{X_i}} = \underset{t=47.87}{9.093} \frac{1}{\sqrt{X_i}} - \underset{t=-15.10}{0.2258} \sqrt{X_i},$$

где Y_i — плотность населения района, X_i — расстояние до центрального делового квартала.

   а) С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок?

   б) Дайте интерпретацию полученным результатам.

2. Были обследованы 36 предприятий по трём показателям: K_i — основным фондам (млн. руб.), W_i — фонду оплаты труда (млн. руб.), R_i — расходам на НИОКР (млн. руб.). Получены оценки вектора средних $\hat{\mu} = (3, 4, 2)'$ и ковариационной матрицы $\hat{\Sigma} = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 10 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

   Найдите первую главную компоненту и определите долю суммарной дисперсии, которую она объясняет.

3. Мартовский Заяц и Безумный Шляпник почти все время пьют чай. Известно, что количество выпитого за день чая (в чашках) зависит от количества пирожных (в штуках) и печенья (в штуках). Алиса, гостившая у героев в течение 30 дней, заметила, что если оценить зависимость выпитого чая от закуски для Мартовского Зайца и Шляпника

$$Tea_i = \beta_1 + \beta_2 Biscuit_i + \beta_3 Cake_i + u_i,$$

   то получится регрессия с $RSS = 20$.

   Чтобы понять, удачную ли модель она построила, Алиса оценила еще одну регрессию

$$Tea_i = \beta_1 + \beta_2 Biscuit_i + \beta_3 Cake_i + +\gamma_2 \widehat{Tea_i^2} + \gamma_3 \widehat{Tea_i^3} + \gamma_4 \widehat{Tea_i^4} + \nu_i,$$

   с $RSS = 10$.

   Помогите Алисе понять, верную ли спецификацию модели она выбрала: сформулируйте основную и альтернативную гипотезы и проведите подходящий тест.

Несколько решений

1.

2.
```
eigen(matrix(c(2, 3, 3, 10), nrow = 2))
```

```
## $values
## [1] 11  1
##
## $vectors
##           [,1]       [,2]
## [1,] 0.3162278 -0.9486833
## [2,] 0.9486833  0.3162278
```

   Доля дисперсии: 0.6875

3.
$$F = \frac{(RSS_R - RSS_{UR})/3}{RSS_{UR}/(n - k_{UR})} = 9$$