Глава 1

Гетероскедастичность

Задача 1.1.

Что такое гетероскедастичность? Гомоскедастичность?

Задача 1.2.

В модели $y = \hat{\beta}_1 + \hat{\beta}_2 x + \varepsilon$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

Задача 1.3.

В модели $y = \hat{\beta}_1 + \hat{\beta}_2 x + \varepsilon$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \lambda |x_i|$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

Задача 1.4.

Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на x_i^2 гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

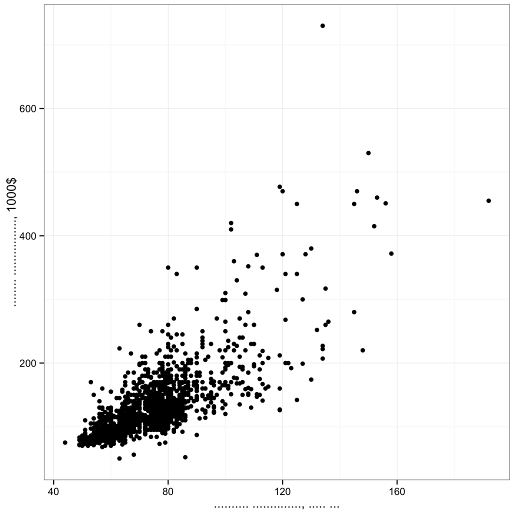
Задача 1.5.

Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на $\sqrt{x_i}$ гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

Задача 1.6.

Диаграмма рассеяния стоимости квартиры в Москве (в 1000\$) и общей площади квартиры имеет вид:

```
ggplot(flats,aes(x=totsp,y=price))+geom_point()+
    labs(x="Общая площадь, кв. м.",
    y="Цена квартиры, 1000$")
```



Какие подходы к оцениванию зависимости имеет смысл посоветовать исходя из данного графика?

Задача 1.7.

По наблюдениям $x = (1, 2, 3)'$, $y = (2, -1, 3)'$ оценивается модель $y = \beta_1 + \beta_2 x + \varepsilon$. Ошибки ε гетероскедастичны и известно, что $\mathrm{Var}(\varepsilon_i) = \sigma^2 \cdot x_i^2$.

1. Найдите оценки $\hat{\beta}_{ols}$ с помощью МНК и их ковариационную матрицу

2. Найдите оценки $\hat{\beta}_{gls}$ с помощью обобщенного МНК и их ковариационную матрицу

Задача 1.8.

Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 30$	1.21	1.89	2.74	48.69
$i = 1, \ldots, 11$	1.39	2.27	2.36	10.28
$i = 12, \ldots, 19$	0.75	2.23	3.19	5.31
$i = 20, \ldots, 30$	1.56	1.06	2.29	14.51

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

Задача 1.9.

Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 50$	1.16	1.99	2.97	174.69
$i = 1, \ldots, 21$	0.76	2.25	3.18	20.41
$i = 22, \ldots, 29$	0.85	1.81	3.32	3.95
$i = 30, \ldots, 50$	1.72	1.41	2.49	130.74

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 1%.

Задача 1.10.

Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 30$	0.96	2.25	3.44	52.70
$i = 1, \ldots, 11$	1.07	2.46	2.40	5.55
$i = 12, \ldots, 19$	1.32	1.01	2.88	11.69
$i = 20, \ldots, 30$	1.04	2.56	4.12	16.00

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

Задача 1.11.

Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 50$	0.93	2.02	3.38	145.85
$i = 1, \ldots, 21$	1.12	2.01	3.32	19.88
$i = 22, \ldots, 29$	0.29	2.07	2.24	1.94
$i = 30, \ldots, 50$	0.87	1.84	3.66	117.46

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

Задача 1.12.

Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$. При оценивании с помощью МНК были получены результаты: $\hat{\beta}_1 = 1.21$, $\hat{\beta}_2 = 1.11$, $\hat{\beta}_3 = 3.15$, $R^2 = 0.72$.

Оценена также вспомогательная регрессия: $\hat{\varepsilon}_i = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i$. Результаты оценивания следующие: $\hat{\delta}_1 = 1.50$, $\hat{\delta}_2 = -2.18$, $\hat{\delta}_3 = 0.23$, $\hat{\delta}_4 = 1.87$, $\hat{\delta}_5 = -0.56$, $\hat{\delta}_6 = -0.09$, $R^2_{aux} = 0.36$

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

Задача 1.13.

Объясните, с какой целью используются стандартные ошибки в форме Уайта. Приведите развернутый ответ. Верно ли, что стандартные ошибки в форме Уайта позволяют

1. устранить гетероскедастичность?

2. корректно тестировать гипотезы относительно коэффициентов регрессии в условиях гетероскедастичности?

Задача 1.14.

Объясните, с какой целью используются стандартные ошибки в форме Невье–Веста. Приведите развернутый ответ. Верно ли, что стандартные ошибки в форме Невье–Веста позволяют

1. устранить гетероскедастичность?
2. корректно тестировать гипотезы относительно коэффициентов регрессии в условиях гетероскедастичности?

Задача 1.15.

Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещенных оценок.

Задача 1.16.

Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещенных оценок.

Задача 1.17.

Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещенных оценок.

Задача 1.18.

Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещенных оценок.

Задача 1.19.

Докажите, что в условиях гетероскедастичности МНК- оценки остаются несмещенными.

Задача 1.20.

Оценка коэффициентов обобщенного МНК имеет вид $\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y$, где $V = \mathrm{Var}(\varepsilon)$. Совпадает ли оценка $\hat{\beta}_{GLS}$ с оценкой обычным МНК в условиях гомоскедастичности?

Задача 1.21.

Модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ оценивается по трём наблюдениям, $y = (9, 3, 6)$, $x = (1, 2, 4)$. Имеется гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$, ошибки ε_i нормально распределены.

1. Оцените $\hat{\beta}$ с помощью МНК проигнорировав гетероскедастичность. Постройте 95% доверительный интервал для каждого коэффициента, проигнорировав гетероскедастичность

2. Оцените $\hat{\beta}$ с помощью обобщенного МНК учтя гетероскедастичность. Постройте 95% доверительный интервал для каждого коэффициента с учётом гетероскедастичности

Задача 1.22.

Рассмотрим модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, где ошибки ε_i некоррелированы, $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$. Предлагается два способа оценить коэффициенты модели:

WLS. Взвешенный метод наименьших квадратов. Поделим каждое уравнение $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на σ_i. Затем обычным методом наименьших квадратов в преобразованной модели $y_i/\sigma_i = \beta_1 \cdot 1/\sigma_i + \beta_2 x_i/\sigma_i + \varepsilon_i/\sigma_i$ найдем оценки $\hat{\beta}_{WLS}$.

GLS. Обобщенный метод наименьших квадратов. Оценки $\hat{\beta}_{GLS}$ находим по формуле $\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y$, где

$$V = \mathrm{Var}(\varepsilon) = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{pmatrix}$$

1. Докажите, что в матричном виде преобразование взвешенного МНК записывается как $V^{-1/2}y = V^{-1/2}X\beta + V^{-1/2}\varepsilon$.

2. Верно ли, что $\hat{\beta}_{WLS} = \hat{\beta}_{GLS}$?

3. Найдите $\mathbb{E}(\hat{\beta}_{WLS})$, $\mathrm{Var}(\hat{\beta}_{WLGS})$

4. В явном виде выпишите $\hat{\beta}_{2,WLS}$

Задача 1.23.

Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, в которой ошибки ε_i независимы и имеют нормальное распределение $N(0, \sigma^2)$. Для $n = 200$ наблюдений найдите

1. вероятность того, что статистика Уайта окажется больше 10,
2. ожидаемое значение статистики Уайта,
3. дисперсию статистики Уайта.

Задача 1.24.

Найдите число коэффициентов во вспомогательной регрессии, необходимой для выполнения теста Уайта, если число коэффициентов в исходной регрессии равно k, включая свободный член.

Задача 1.25.

(Кирилл Фурманов) По 35 наблюдениям сотрудники НИИ оценили уравнение регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и рассчитали остатки ε_i. После того они приступили к диагностике возможных недостатков модели, обнаружили гетероскедастичность и решили её побороть.

(a) Самый младший научный сотрудник выдвинул предположение, что стандартное отклонение случайной составляющей может быть выражено так: $\sigma_{\varepsilon,i} = ax_i$, где a — неизвестный коэффициент. Каким образом нужно преобразовать исходное уравнение регрессии, чтобы избавиться от гетероскедастичности?

(b) Профессор решил перепроверить результаты и оценил регрессию:

$$\hat{e}_i^2 = -0.3 + 0.08x_i - 0.01x_i^2, R^2 = 0.15$$

Свидетельствует ли полученный профессором результат о наличии гетероскедастичности?

Задача 1.26.

Пусть $y_t = \beta x_t + \varepsilon$ где $\mathbb{E}(\varepsilon_t) = 0$ и известно, что оценки для параметров $\tilde{\beta} = \left(\sum_{t=1}^n y_t\right) / \left(\sum_{t=1}^n x_t\right)$ являются наилучшими (в смысле минимума дисперсии) среди линейных несмещенных оценок параметра β. Чему равна в этом случае матрица ковариаций вектора ε с точностью до пропорциональности?

Задача 1.27.

Для регрессии $y = X\beta + \varepsilon$ с $\mathbb{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \Sigma \neq \sigma^2 I$, оцененной с помощью обобщённого метода наименьших квадратов, найдите ковариационную матрицу $\mathrm{Cov}(\hat{\beta}_{GLS}, \varepsilon)$

Задача 1.28.

Найдите наиболее эффективную оценку коэффициента β_1 для модели $y_i = \beta_1 + \varepsilon$, $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma_\varepsilon^2 / x_i$, $x_i > 0$ в классе линейных несмещенных оценок

Задача 1.29.

Фурманов Кирилл. Исследователь оценил регрессионную модель $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$ и провёл диагностику различных проблемных явлений. Результаты его стараний приведены ниже:

Также $VIF_2 = 1.06$, $VIF_3 = 1.07$, $VIF_4 = 1.02$

(a) Определите, какие проблемные явления обнаружил исследователь. Обоснуйте свой ответ.

(b) Найдите коэффициент детерминации для регрессии: $x_{i2} = \gamma_1 + \gamma_2 x_{i3} + \gamma_3 x_{i4} + u_i$

Задача 1.30.

В модели $y_i = \beta x_i + \varepsilon_i$ предполагается гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \exp(\gamma_1 + \gamma_2 x_i)$.

1. Сформулируйте гипотезу о гомоскедастичности с помощью коэффициентов

2. Выведите в явном виде оценку максимального правдоподобия при предположении о гомоскедастичности

3. Выпишите условия первого порядка для оценки максимального правдоподобия без предположения о гомоскедастичности

4. Выведите в явном виде формулу для LM теста множителей Лагранжа