Эконометрика

с Монте-Карло и эконометрессами

в задачах и упражнениях

Дмитрий Борзых, Борис Демешев

30 ноября 2013 г.

Содержание

Todo list

1 МНК без матриц и вероятностей

Задача 1.1.
Верно ли, что для любых векторов $a = (a_1, \ldots, a_n)$ и $b = (b_1, \ldots, b_n)$ справедливы следующие равенства?

1. $\sum_{i=1}^{n}(a_i - \bar{a}) = 0$
2. $\sum_{i=1}^{n}(a_i - \bar{a})^2 = \sum_{i=1}^{n}(a_i - \bar{a})a_i$
3. $\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b}) = \sum_{i=1}^{n}(a_i - \bar{a})b_i$
4. $\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b}) = \sum_{i=1}^{n}a_i b_i$

Задача 1.2.
При помощи метода наименьших квадратов найдите оценку неизвестного параметра θ в следующих моделях:

1. $y_i = \theta + \theta x_i + \varepsilon_i$
2. $y_i = \theta - \theta x_i + \varepsilon_i$
3. $\ln y_i = \theta + \ln x_i + \varepsilon_i$
4. $y_i = \theta + x_i + \varepsilon_i$
5. $y_i = 1 + \theta x_i + \varepsilon_i$
6. $y_i = \theta / x_i + \varepsilon_i$
7. $y_i = \theta x_{i1} + (1-\theta)x_{i2} + \varepsilon_i$

Задача 1.3.
Покажите, что для моделей $y_i = \alpha + \beta x_i + \varepsilon_i$, $z_i = \gamma + \delta x_i + \upsilon_i$ и $y_i + z_i = \mu + \lambda x_i + \xi_i$ МНК-оценки связаны соотношениями $\hat{\mu} = \hat{\alpha} + \hat{\gamma}$ и $\hat{\lambda} = \hat{\beta} + \hat{\delta}$.

Задача 1.4.
Найдите МНК-оценки параметров α и β в модели $y_i = \alpha + \beta y_i + \varepsilon_i$.

Задача 1.5.
Рассмотрите модели $y_i = \alpha + \beta(y_i + z_i) + \varepsilon_i$, $z_i = \gamma + \delta(y_i + z_i) + \varepsilon_i$.

1. Как связаны между собой $\hat{\alpha}$ и $\hat{\gamma}$?
2. Как связаны между собой $\hat{\beta}$ и $\hat{\delta}$?

Задача 1.6.
Как связаны МНК-оценки параметров α, β и γ, δ в моделях $y_i = \alpha + \beta x_i + \varepsilon_i$ и $z_i = \gamma + \delta x_i + \upsilon_i$, если $z_i = 2y_i$.

Задача 1.7.
Для модели $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$ решите условную задачу о наименьших квадратах: $Q(\beta_1, \beta_2) := \sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2 \to \min\limits_{\beta_1 + \beta_2 = 1}$

Задача 1.8.
Даны n пар чисел: (x_1, y_1), …, (x_n, y_n). Мы прогнозируем y_i по формуле $\hat{y}_i = \hat{\beta} x_i$. Найдите $\hat{\beta}$ методом наименьших квадратов.

Задача 1.9.
Даны n чисел: $y_1, \ldots, y_n$. Мы прогнозируем y_i по формуле $\hat{y}_i = \hat{\beta}$. Найдите $\hat{\beta}$ методом наименьших квадратов.

Задача 1.10.
Даны n пар чисел: (x_1, y_1), …, (x_n, y_n). Мы прогнозируем y_i по формуле $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$. Найдите $\hat{\beta}_1$ и $\hat{\beta}_2$ методом наименьших квадратов.

Задача 1.11.
Даны n пар чисел: (x_1, y_1), …, (x_n, y_n). Мы прогнозируем y_i по формуле $\hat{y}_i = 1 + \hat{\beta} x_i$. Найдите $\hat{\beta}$ методом наименьших квадратов.

Задача 1.12.

Перед нами два золотых слитка и весы, производящие взвешивания с ошибками. Взвесив первый слиток, мы получили результат 300 грамм, взвесив второй слиток — 200 грамм, взвесив оба слитка — 400 грамм. Оцените вес каждого слитка методом наименьших квадратов.

Задача 1.13.

Аня и Настя утверждают, что лектор опоздал на 10 минут. Таня считает, что лектор опоздал на 3 минуты. С помощью мнк оцените на сколько опоздал лектор.

Задача 1.14.

Функция $f(x)$ дифференциируема на отрезке $[0;1]$. Найдите аналог МНК-оценок для регрессии без свободного члена в непрерывном случае. Более подробно: найдите минимум по $\hat{\beta}$ для функции

$$Q(\hat{\beta}) = \int_0^1 (f(x) - \hat{\beta}x)^2\, dx \tag{1}$$

Задача 1.15.

Есть двести наблюдений. Вовочка оценил модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ по первой сотне наблюдений. Петечка оценил модель $\hat{y} = \hat{\gamma}_1 + \hat{\gamma}_2 x$ по второй сотне наблюдений. Машенька оценила модель $\hat{y} = \hat{m}_1 + \hat{m}_2 x$ по всем наблюдениям.

1. Возможно ли, что $\hat{\beta}_2 > 0$, $\hat{\gamma}_2 > 0$, но $\hat{m}_2 < 0$?
2. Возможно ли, что $\hat{\beta}_1 > 0$, $\hat{\gamma}_1 > 0$, но $\hat{m}_1 < 0$?
3. Возможно ли одновременное выполнение всех упомянутых условий?

Задача 1.16.

Вася оценил модель $y = \beta_1 + \beta_2 d + \beta_3 x + \varepsilon$. Дамми-переменная d обозначает пол, 1 для мужчин и 0 для женщин. Оказалось, что $\hat{\beta}_2 > 0$. Означает ли это, что для мужчин $\bar{y}$ больше, чем $\bar{y}$ для женщин?

Задача 1.17.

Какие из указанные моделей можно представить в линейном виде?

1. $y_i = \beta_1 + \frac{\beta_2}{x_i} + \varepsilon_i$
2. $y_i = \exp(\beta_1 + \beta_2 x_i + \varepsilon_i)$
3. $y_i = 1 + \frac{1}{\exp(\beta_1 + \beta_2 x_i + \varepsilon_i)}$
4. $y_i = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i + \varepsilon_i)}$
5. $y_i = x_i^{\beta_2} e^{\beta_1 + \varepsilon_i}$

Задача 1.18.

У эконометриста Вовочки есть переменная 1_f, которая равна 1, если i-ый человек в выборке — женщина, и 0, если мужчина. Есть переменная 1_m, которая равна 1, если i-ый человек в выборке — мужчина, и 0, если женщина. Какие $\hat{y}$ получатся, если Вовочка попытается построить регрессии:

1. y на константу и 1_f
2. y на константу и 1_m
3. y на 1_f и 1_m без константы
4. y на константу, 1_f и 1_m

Задача 1.19.

У эконометриста Вовочки есть три переменных: r_i — доход i-го человека в выборке, m_i — пол (1 — мальчик, 0 — девочка) и f_i — пол (1 — девочка, 0 — мальчик). Вовочка оценил две модели

Модель A $m_i = \beta_1 + \beta_2 r_i + \varepsilon_i$

Модель B $f_i = \gamma_1 + \gamma_2 r_i + u_i$

1. Как связаны между собой оценки $\hat{\beta}_1$ и $\hat{\gamma}_1$?
2. Как связаны между собой оценки $\hat{\beta}_2$ и $\hat{\gamma}_2$?

Задача 1.20.

Эконометрист Вовочка оценил линейную регрессионную модель, где y измерялся в тугриках. Затем он оценил ту же модель, но измерял y в мунгу (1 тугрик = 100 мунгу). Как изменятся оценки коэффициентов?
Задача 1.21.
Возможно ли, что при оценке парной регрессии $y = \beta_1 + \beta_2 x + \varepsilon$ оказывается, что $\hat{\beta}_2 > 0$, а при оценке регрессии без константы, $y = \gamma x + \varepsilon$, оказывается, что $\hat{\gamma} < 0$?
Задача 1.22.
Эконометрист Вовочка оценил регрессию y только на константу. Какой коэффициент R^2 он получит?
Задача 1.23.
Эконометрист Вовочка оценил методом наименьших квадратов модель 1, $y = \beta_1 + \beta_2 x + \beta_3 z + \varepsilon$, а затем модель 2, $y = \beta_1 + \beta_2 x + \beta_3 z + \beta_4 w + \varepsilon$. Сравните полученные ESS, RSS, TSS и R^2.
Задача 1.24.
Создайте набор данных с тремя переменными y, x и z со следующими свойствами. При оценке модели $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ получается $\hat{\beta}_2 > 0$. При оценке модели $\hat{y} = \hat{\gamma}_1 + \hat{\gamma}_2 x + \hat{\gamma}_3 z$ получается $\hat{\gamma}_2 < 0$. Объясните принцип, руководствуясь которым легко создать такой набор данных.
Задача 1.25.
У меня есть набор данных с выборочным средним $\bar{y}$ и выборочной дисперсией s_y^2. Как нужно преобразовать данные, чтобы выборочное среднее равнялось 7, а выборочная дисперсия — 9?

2 Парный МНК без матриц

Задача 2.1.
Рассмотрим модель $y_t = \beta_1 + \beta_2 \cdot t + \varepsilon_t$, где ошибки ε_t независимы и равномерны на $[-1; 1]$. С помощью симуляций на компьютере оцените и постройте график функции плотности для $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{s}^2$, $\widehat{\mathrm{Var}}(\hat{\beta}_1)$, $\widehat{\mathrm{Var}}(\hat{\beta}_2)$ и $\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$.
Задача 2.2.
Пусть $y_i = \mu + \varepsilon_i$, где $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. Найдите:

1. $\mathbb{E}(\overline{y})$
2. $\mathrm{Var}(\overline{y})$
3. $\mathbb{E}(\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2)$
4. $\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2)$, если дополнительно известно, что ε_i нормально распределены

Задача 2.3.
Рассматривается модель $y_i = \beta x_i + \varepsilon_i$, $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. При каких значениях параметров c_i несмещённая оценка $\hat{\beta} = \frac{\sum_{i=1}^n c_i y_i}{\sum_{i=1}^n c_i x_i}$ имеет наименьшую дисперсию?
Задача 2.4.
Пусть $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и $i = 1, \ldots, 5$ — классическая регрессионная модель. Также имеются следующие данные: $\sum_{i=1}^5 y_i^2 = 55, \sum_{i=1}^5 x_i^2 = 3, \sum_{i=1}^5 x_i y_i = 12, \sum_{i=1}^5 y_i = 15, \sum_{i=1}^5 x_i = 3$. Используя их, найдите:

1. $\hat{\beta}_1$ и $\hat{\beta}_2$
2. $\mathrm{Corr}(\hat{\beta}_1, \hat{\beta}_2)$
3. TSS
4. ESS
5. RSS
6. R^2
7. $\hat{\sigma}^2$

Проверьте следующие гипотезы:

1. $\begin{cases} H_0 : \beta_2 = 2 \\ H_a : \beta_2 \neq 2 \end{cases}$

2. $\begin{cases} H_0 : \beta_1 + \beta_2 = 1 \\ H_a : \beta_1 + \beta_2 \neq 1 \end{cases}$

Задача 2.5.

Пусть $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и $i = 1, \ldots, 5$ — классическая регрессионная модель. Также имеются следующие данные: $\sum_{i=1}^5 y_i^2 = 55$, $\sum_{i=1}^5 x_i^2 = 2$, $\sum_{i=1}^5 x_i y_i = 9$, $\sum_{i=1}^5 y_i = 15$, $\sum_{i=1}^5 x_i = 2$. Используя их, найдите:

1. $\hat{\beta}_1$ и $\hat{\beta}_2$
2. $\mathrm{Corr}(\hat{\beta}_1, \hat{\beta}_2)$
3. TSS
4. ESS
5. RSS
6. R^2
7. $\hat{\sigma}^2$

Проверьте следующие гипотезы:

1. $\begin{cases} H_0 : \beta_2 = 2 \\ H_a : \beta_2 \neq 2 \end{cases}$
2. $\begin{cases} H_0 : \beta_1 + \beta_2 = 1 \\ H_a : \beta_1 + \beta_2 \neq 1 \end{cases}$

Задача 2.6.

Рассмотрите классическую линейную регрессионную модель $y_i = \beta x_i + \varepsilon_i$. Найдите $\mathbb{E}\hat{\beta}$. Какие из следующих оценок параметра β являются несмещенными:

1. $\hat{\beta} = \frac{y_1}{x_1}$
2. $\hat{\beta} = \frac{1}{2}\frac{y_1}{x_1} + \frac{1}{2}\frac{y_n}{x_n}$
3. $\hat{\beta} = \frac{1}{n}\frac{y_1}{x_1} + \ldots + \frac{y_n}{x_n}$
4. $\hat{\beta} = \frac{\overline{y}}{\overline{x}}$
5. $\hat{\beta} = \frac{y_n - y_1}{x_n - x_1}$
6. $\hat{\beta} = \frac{1}{2}\frac{y_2 - y_1}{x_2 - x_1} + \frac{1}{2}\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$
7. $\hat{\beta} = \frac{1}{n}\frac{y_2 - y_1}{x_2 - x_1} + \frac{1}{n}\frac{y_3 - y_2}{x_3 - x_2} + \ldots + \frac{1}{n}\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$
8. $\hat{\beta} = \frac{1}{n-1}\frac{y_2 - y_1}{x_2 - x_1} + \frac{y_3 - y_2}{x_3 - x_2} + \ldots + \frac{y_n - y_{n-1}}{x_n - x_{n-1}}$
9. $\hat{\beta} = \frac{x_1 y_1 + \ldots + x_n y_n}{x_1^2 + \ldots + x_n^2}$
10. $\hat{\beta} = \frac{1}{2}\frac{y_n - y_1}{x_n - x_1} + \frac{1}{2n}\frac{y_1}{x_1} + \ldots + \frac{y_n}{x_n}$
11. $\hat{\beta} = \frac{1}{2}\frac{y_n - y_1}{x_n - x_1} + \frac{1}{2}\frac{x_1 y_1 + \ldots + x_n y_n}{x_1^2 + \ldots + x_n^2}$
12. $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x^2})^2}$
13. $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(\overline{y} - y_i)}{\sum_{i=1}^n (x_i - \overline{x^2})^2}$
14. $\hat{\beta} = \frac{y_1 + 2y_2 + \ldots + n y_n}{x_1 + 2x_2 + \ldots + n x_n}$
15. $\hat{\beta} = \frac{\sum_{i=1}^n i(y_i - \overline{y})}{\sum_{i=1}^n i(x_i - \overline{x})}$
16. $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{y_i}{x_i}$
17. $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{y_i - \overline{y}}{x_i - \overline{x}}$

Задача 2.7.

Рассмотрите классическую линейную регрессионную модель $y_i = \beta x_i + \varepsilon_i$. Найдите $\mathrm{Var}(\hat{\beta})$.

1. $\hat{\beta} = \frac{y_1}{x_1}$
2. $\hat{\beta} = \frac{1}{2}\frac{y_1}{x_1} + \frac{1}{2}\frac{y_n}{x_n}$
3. $\hat{\beta} = \frac{1}{n}\frac{y_1}{x_1} + \ldots + \frac{y_n}{x_n}$
4. $\hat{\beta} = \frac{\overline{y}}{\overline{x}}$

5. $\hat{\beta} = \frac{y_n - y_1}{x_n - x_1}$
6. $\hat{\beta} = \frac{1}{2}\frac{y_2 - y_1}{x_2 - x_1} + \frac{1}{2}\frac{y_n - y_{n-1}}{x_n - x_{n-1}}$
7. $\hat{\beta} = \frac{x_1 y_1 + \ldots + x_n y_n}{x_1^2 + \ldots + x_n^2}$
8. $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x^2})^2}$
9. $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(\overline{y} - y_i)}{\sum_{i=1}^n (x_i - \overline{x^2})^2}$
10. $\hat{\beta} = \frac{y_1 + 2y_2 + \ldots + ny_n}{x_1 + 2x_2 + \ldots + nx_n}$
11. $\hat{\beta} = \frac{\sum_{i=1}^n i(y_i - \overline{y})}{\sum_{i=1}^n i(x_i - \overline{x})}$
12. $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{y_i}{x_i}$
13. $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n \frac{y_i - \overline{y}}{x_i - \overline{x}}$

Задача 2.8.
Рассмотрите классическую линейную регрессионную модель $y_i = \beta \cdot i + \varepsilon_i$, $i = 1, \ldots, n$. Какая из оценок $\hat{\beta}$ и $\tilde{\beta}$ является более эффективной?

1. $\hat{\beta} = y_1$ и $\tilde{\beta} = y_2/2$
2. $\hat{\beta} = y_1$ и $\tilde{\beta} = \frac{1}{2}y_1 + \frac{1}{2}\frac{y_2}{2}$
3. $\hat{\beta} = \frac{1}{n}\frac{y_1}{1} + \ldots + \frac{y_n}{n}$ и $\tilde{\beta} = \frac{1 \cdot y_1 + \ldots + n \cdot y_n}{1^2 + \ldots + n^2}$

Задача 2.9.
На основе 100 наблюдений была оценена функция спроса:

$$\underset{(s.e.)}{\widehat{\ln Q}} = \underset{(0.04)}{0.87} - \underset{(0.02)}{1.23} \ln P$$

Значимо ли коэффициент эластичности спроса по цене отличается от -1? Рассмотрите уровень значимости 5%.

Задача 2.10.
На основе 100 наблюдений была оценена функция спроса:

$$\underset{(s.e.)}{\widehat{\ln Q}} = \underset{(0.04)}{2.87} - \underset{(0.02)}{1.12} \ln P$$

На уровне значимости 5% проверьте гипотезу $H_0 : \beta_{\ln P} = -1$ против альтернативной $H_a : \beta_{\ln P} < -1$. Дайте экономическую интерпретацию проверяемой гипотезе и альтернативе.

Задача 2.11.
Используя годовые данные с 1960 по 2005 г., была построена кривая Филлипса, связывающая уровень инфляции Inf и уровень безработицы $Unem$:

$$\widehat{Inf} = 2.34 - 0.23 Unem$$

$$\sqrt{\widehat{Var}(\hat{\beta}_{Unem})} = 0.04, R^2 = 0.12$$

На уровне значимости 1% проверьте гипотезу $H_0 : \beta_{Unem} = 0$ против альтернативной $H_a : \beta_{Unem} \neq 0$.

Задача 2.12.
Пусть $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и $i = 1, \ldots, 18$ — классическая регрессионная модель, где $\mathbb{E}(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$. Также имеются следующие данные: $\sum_{i=1}^{18} y_i^2 = 4256, \sum_{i=1}^{18} x_i^2 = 185, \sum_{i=1}^{18} x_i y_i = 814.25, \sum_{i=1}^{18} y_i = 225, \sum_{i=1}^{18} x_i = 49.5$. Используя эти данные, оцените эту регрессию и на уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 = 3.5$ против альтернативной $H_a : \beta_1 > 3.5$:

1. Приведите формулу для тестовой статистики
2. Укажите распределение тестовой статистики
3. Вычислите наблюдаемое значение тестовой статистики
4. Укажите границы области, где основная гипотеза не отвергается

5. Сделайте статистический вывод

Задача 2.13.

Рассматривается модель $y_i = \mu + \varepsilon_i$, где $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ и $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$. При каких c_i несмещенная оцека

$$\hat{\mu} = \sum_{i=1}^{n} c_i y_i$$

имеет наименьшую дисперсию?

Задача 2.14.

Рассмотрим классическую линейную регрессионную модель, $y_t = \beta \cdot t + \varepsilon_t$. Какая из оценок, $\hat{\beta}$ или $\hat{\beta}'$ является более эффективной?

1. $\hat{\beta} = y_1$, $\hat{\beta}' = y_2/2$
2. $\hat{\beta} = y_1$, $\hat{\beta}' = 0.5y_1 + 0.5\frac{y_2}{2}$
3. $\hat{\beta} = \frac{1}{n}\left(y_1 + \frac{y_2}{2} + \frac{y_3}{3} + \ldots + \frac{y_n}{n}\right)$, $\hat{\beta}' = \frac{y_1 + 2y_2 + \ldots + ny_n}{1^2 + 2^2 + \ldots + n^2}$

Задача 2.15.

Ошибки регрессии ε_i независимы и равновероятно принимают значения $+1$ и -1. Также известно, что $y_i = \beta \cdot i + \varepsilon_i$. Модель оценивается всего по двум наблюдениям.

1. Найдите закон распределения $\hat{\beta}$, RSS, ESS, TSS, R^2
2. Найдите $\mathbb{E}(\hat{\beta})$, $\mathrm{Var}(\hat{\beta})$, $\mathbb{E}(RSS)$, $\mathbb{E}(ESS)$, $\mathbb{E}(R^2)$
3. При каком β величина $\mathbb{E}(R^2)$ достигает максимума?

Задача 2.16.

Рассмотрим модель с линейным трендом без свободного члена, $y_t = \beta t + \varepsilon_t + \varepsilon_t$.

1. Найдите МНК оценку коэффициента β
2. Рассчитайте $\mathbb{E}(\hat{\beta})$ и $\mathrm{Var}(\hat{\beta})$ в предположениях теоремы Гаусса-Маркова
3. Верно ли, что оценка $\hat{\beta}$ состоятельна?

Задача 2.17.

В модели $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, где $x_t = \begin{cases} 2,\ t = 1 \\ 1,\ t > 1 \end{cases}$:

1. Найдите мнк-оценку $\hat{\beta}_2$
2. Рассчитайте $\mathbb{E}(\hat{\beta}_2)$ и $\mathrm{Var}(\hat{\beta}_2)$ в предположениях теоремы Гаусса-Маркова
3. Верно ли, что оценка $\hat{\beta}_2$ состоятельна?

Задача 2.18.

В модели $y_t = \beta_1 + \beta_2 x_t$, где $x_t = \begin{cases} 1,\ t = 2k+1 \\ 0,\ t = 2k \end{cases}$:

1. Найдите мнк-оценку $\hat{\beta}_2$
2. Рассчитайте $\mathbb{E}(\hat{\beta}_2)$ и $\mathrm{Var}(\hat{\beta}_2)$ в предположениях теоремы Гаусса-Маркова
3. Верно ли, что оценка $\hat{\beta}_2$ состоятельна?

Задача 2.19.

Априори известно, что парная регрессия должна проходить через точку (x_0, y_0).

1. Выведите формулы МНК оценок;
2. В предположениях теоремы Гаусса-Маркова найдите дисперсии и средние оценок

Задача 2.20.

Мы предполагаем, что y_t растёт с линейным трендом, т.е. $y_t = \beta_1 + \beta_2 t + \varepsilon_t$. Все предпосылки теоремы Гаусса-Маркова выполнены. В качестве оценки $\hat{\beta}_2$ предлагается $\hat{\beta}_2 = \frac{Y_T - Y_1}{T-1}$, где T — общее количество наблюдений.

1. Найдите $\mathbb{E}(\hat{\beta}_2)$ и $\mathrm{Var}(\hat{\beta}_2)$
2. Совпадает ли оценка $\hat{\beta}_2$ с классической мнк-оценкой?
3. У какой оценки дисперсия выше, у $\hat{\beta}_2$ или классической мнк-оценки?

Задача 2.21.
Вася считает, что выборочная ковариация $\mathrm{sCov}(y,\hat{y}) = \frac{\sum(y_i-\bar{y})(\hat{y}_i-\bar{y})}{n-1}$ это неплохая оценка для $\mathrm{Cov}(y_i,\hat{y}_i)$. Прав ли он?
Задача 2.22.
В классической линейной регрессионной модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, дисперсия зависимой переменной не зависит от номера наблюдения, $\mathrm{Var}(y_i) = \sigma^2$. Почему для оценки σ^2 вместо известной из курса математической статистики формулы $\sum(y_i-\bar{y})^2/(n-1)$ используют $\sum\hat{\varepsilon}_i^2/(n-2)$?
Задача 2.23.
Оценка регрессии имеет вид $\hat{y}_i = 3-2x_i$. Выборочная дисперсия x равна 9, выборочная дисперсия y равна 40. Найдите R^2 и выборочные корреляции $\mathrm{sCorr}(x,y)$, $\mathrm{sCorr}(y,\hat{y})$.
Задача 2.24.
Слитки-вариант. Перед нами два золотых слитка и весы, производящие взвешивания с ошибками. Взвесив первый слиток, мы получили результат 300 грамм, взвесив второй слиток — 200 грамм, взвесив оба слитка — 400 грамм. Предположим, что ошибки взвешивания — независимые одинаково распределенные случайные величины с нулевым средним.

1. Найдите несмещеную оценку веса первого слитка, обладающую наименьшей дисперсией.
2. Как можно проинтерпретировать нулевое математическое ожидание ошибки взвешивания?

Задача 2.25.
Рассмотрим линейную модель $y_i = \beta_1+\beta_2 x_i+\varepsilon_i$, где ошибки ε_i нормальны $N(0;\sigma^2)$ и независимы.

1. Верно ли, что y_i одинаково распределены?
2. Верно ли, что $\bar{y}$ — это несмещенная оценка для $\mathbb{E}(y_i)$?
3. Верно ли, что $\sum(y_i-\bar{y})^2/(n-1)$ — несмещенная оценка для σ^2? Если да, то докажите, если нет, то определите величину смещения

Задача 2.26.
Модель регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, в которой ошибки ε_i независимы и нормальны $N(0;\sigma^2)$, оценивается по 22 наблюдениям. Найдите $\mathbb{E}(RSS)$, $\mathrm{Var}(RSS)$, $\mathbb{P}(10\sigma^2 < RSS < 30\sigma^2)$, $\mathbb{P}(10\hat{\sigma}^2 < RSS < 30\hat{\sigma}^2)$
Задача 2.27.
Модель регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, в которой ошибки ε_i независимы и нормальны $N(0;\sigma^2)$, оценивается по 12 наблюдениям. Найдите

1. $\mathbb{P}(\hat{\beta}_1 > \beta_1)$, $\mathbb{P}(\beta_1 > 0)$, $\mathbb{P}(|\hat{\beta}_1 - \beta_1| < se(\hat{\beta}_1))$, $\mathbb{P}(\hat{\beta}_2 > \beta_2 + se(\hat{\beta}_2))$, $\mathbb{P}(\hat{\beta}_2 > \beta_2 - se(\hat{\beta}_2))$
2. $\mathbb{E}(\hat{\beta}_1)$, $\mathbb{E}(\hat{\beta}_2)$, $\mathbb{E}(\beta_2)$
3. Закон распределения, математическое ожидание и дисперсию величин $\frac{\hat{\beta}_2-\beta_2}{\sqrt{\mathrm{Var}(\hat{\beta}_2)}}$, $\frac{\hat{\beta}_2-\beta_2}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_2)}}$, $\frac{\hat{\beta}_1+\hat{\beta}_2-\beta_1-\beta_2}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1+\hat{\beta}_2)}}$
4. $\mathbb{P}(\hat{s} > \sigma)$, $\mathbb{P}(\hat{s} > 2\sigma)$

Задача 2.28.
Для модели парной регрессии известны $y = (1,2,3,4,5)'$ и $\hat{y} = (2,2,2,4,5)'$. Найдите RSS, TSS, R^2, $\hat{s}^2$.
Задача 2.29.
В классической парной регрессионной модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ с нормально распределенными ошибками, оцениваемой по 30 наблюдениям, дополнительно известно, что $\mathrm{Var}(\varepsilon_7) = 9$. Найдите

1. $\mathbb{E}(\varepsilon_2)$, $\mathrm{Cov}(\varepsilon_1,\varepsilon_3)$, $\mathbb{E}(\varepsilon_3^5)$, $\mathbb{E}(e_5^3)$, $\mathrm{Var}(e_5)$, $\mathrm{Var}(y_3)$
2. $\mathbb{P}(e_2 > \varepsilon_3)$, $\mathbb{P}(e_1 > 0)$, $\mathbb{P}(e_1 > 3)$
3. $\mathbb{E}(RSS)$, $\mathrm{Var}(RSS)$, $\mathbb{P}(RSS > 200)$

Задача 2.30.
В модели парной регрессии придумайте такие наблюдения, чтобы:

- $R^2 = 0.9$
- $R^2 = 0.8$ и регрессия имела вид $\hat{y} = 2 + 3x$

3 Многомерный МНК без матриц

Задача 3.1.
Эконометрэсса Ширли зашла в пустую аудиторию, где царил приятный полумрак, и увидела на доске до боли знакомую надпись:

$$\hat{y} = \underset{(2.37)}{1.1} - \underset{(-0.4)}{0.7} \cdot x_2 + \underset{(3.15)}{0.9} \cdot x_3 - \underset{(-0.67)}{19} \cdot x_4$$

Помогите эконометрэссе Ширли определить, что находится в скобках

1. P-значения
2. t-статистики
3. стандартные ошибки коэффициентов
4. R^2 скорректированный на номер коэффициента
5. показатели VIF для каждого коэффициента

Задача 3.2.
Для нормальной регрессии с 5-ю факторами (включая свободный член) известны границы симметричного по вероятности 80% доверительного интервала для дисперсии σ^2_ε: $[45; 87.942]$.

1. Определите количество наблюдений в выборке
2. Вычислите $\hat{\sigma}^2_\varepsilon$

Задача 3.3.
Для коэффициентов регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$ даны 95%-ые доверительные интервалы: $\beta_2 \in (0.16; 0.66)$, $\beta_3 \in (-0.33; 0.93)$ и $\beta_4 \in (-1.01; 0.54)$.

1. Найдите $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$
2. Определите, какие из переменных в регрессии значимы на уровне значимости 5%.

Задача 3.4.
Для коэффициентов регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$ даны 95%-ые доверительные интервалы: $\beta_2 \in (-0.15; 1.65)$, $\beta_3 \in (0.32; 0.93)$ и $\beta_4 \in (0.14; 1.55)$.

1. Найдите $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$
2. Определите, какие из переменных в регрессии значимы на уровне значимости 5%.

Задача 3.5.
Эконометрэсса Мырли очень суеверна и поэтому оценила три модели:

M1 $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$ по всем наблюдениям.

M2 $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \beta_4 d_i + \varepsilon_i$ по всем наблюдениям, где d_i — дамми-переменная равная 1 для 13-го наблюдения и нулю иначе.

M3 $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$ по всем наблюдениям, кроме 13-го.

1. Сравните между собой RSS во всех трёх моделях
2. Есть ли совпадающие оценки коэффициентов в этих трёх моделях? Если есть, то какие?
3. Может ли Мырли не выполняя вычислений узнать ошибку прогноза для 13-го наблюдения при использовании третьей модели? Если да, то как?

Задача 3.6.
Рассмотрим модель $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \beta_4 z_i + \varepsilon_i$. При оценке модели по 24 наблюдениям оказалось, что $RSS = 15$, $\sum (y_i - \bar{y} - w_i + \bar{w})^2 = 20$. На уровне значимости 1% протестируйте гипотезу

$$H_0 : \begin{cases} \beta_2 + \beta_3 + \beta_4 = 1 \\ \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}$$

Задача 3.7.

Модель регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, в которой ошибки ε_i независимы и нормальны $N(0;\sigma^2)$, оценивается по 13 наблюдениям. Найдите $\mathbb{E}(RSS)$, $\mathrm{Var}(RSS)$, $\mathbb{P}(5\sigma^2 < RSS < 10\sigma^2)$, $\mathbb{P}(5\hat{\sigma}^2 < RSS < 10\hat{\sigma}^2)$

Задача 3.8.

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 $N(0,\sigma^2)$. Известно, что выборка в $n = 30$ наблюдений была разбита на три непересекающиеся подвыборки, содержащие $n_1 = 13$, $n_2 = 4$ и $n_3 = 13$ наблюдений. Пусть $\hat{s}_j^2$ — это оценка дисперсии случайных ошибок для регрессии, оцененной по j-ой подвыборке. Найдите

1. $\mathbb{P}(\hat{s}_3^2 > \hat{s}_1^2)$, $\mathbb{P}(\hat{s}_1^2 > 2\hat{s}_2^2)$
2. $\mathbb{E}(\hat{s}_2^2/\hat{s}_1^2)$, $\mathrm{Var}(\hat{s}_2^2/\hat{s}_1^2)$

Задача 3.9.

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 $n = 13$ наблюдения найдите уровень доверия следующих доверительных интервалов для неизвестного параметра σ^2:

1. $(0; RSS/4.865)$
2. $(RSS/18.307; RSS/3.940)$
3. $(RSS/15.987; \infty)$

4 МНК с матрицами и вероятностями

1. Пусть $y = X\beta + \varepsilon$ — регрессионная модель.
   (a) Сформулируйте теорему Гаусса-Маркова
   (b) Верно ли, что оценка $\hat{\beta} = (X'X)^{-1}X'y$ несмещённая?
   (c) В условиях теоремы Гаусса-Маркова найдите ковариационную матрицу $\hat{\beta}$
2. Пусть $y = X\beta + \varepsilon$ — регрессионная модель и $\tilde{\beta} = ((X'X)^{-1}X' + A)y$ — несмещённая оценка вектора неизвестных параметров β. Верно ли, что $AX = 0$?
3. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, $X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$, $\mathbb{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2 I$. Найдите коэффициент корреляции $\mathrm{Corr}(\hat{\beta}_1, \hat{\beta}_2)$.
4. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$. Пусть $Z = XD$, где $D = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Рассмотрите «новую» регрессионную модель $y = Z\alpha + u$, где $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$. Определите, как выражаются «новые» МНК-коэффициенты через «старые».
5. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$. Пусть $Z = XD$, где $D = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Рассмотрите «новую» регрессионную модель $y = Z\alpha + u$, где $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$. Определите, как выражаются «новые» МНК-коэффициенты через «старые».

6. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$. Пусть $Z = XD$, где $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Рассмотрите «новую» регрессионную модель $y = Z\alpha + u$, где $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$. Определите, как выражаются «новые» МНК-коэффициенты через «старые».
7. Пусть $y = X\beta + \varepsilon$ — регрессионная модель. Верно ли, что $\hat{\varepsilon}'\hat{y} = 0$ и $\hat{y}'\hat{\varepsilon} = 0$?
8. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $\mathbb{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2 I$. Пусть A — неслучайная матрица размера $k \times k$, $\det(A) \neq 0$. Совершается преобразование регрессоров по правилу $Z = XA$. В преобразованных регрессорах уравнение выглядит так: $y = Z\gamma + u$, где $\mathbb{E}(u) = 0$, $\mathrm{Var}(u) = \sigma_u^2 I$.
   (a) Как связаны между собой МНК-оценки $\hat{\beta}$ и $\hat{\gamma}$?
   (b) Как связаны между собой векторы остатков регрессий?
   (c) Как связаны между собой прогнозные значения, полученные по двум регрессиям?
9. Рассмотрим оценку вида $\tilde{\beta} = ((X'X)^{-1} + \gamma I)X'y$ для вектора коэффициентов регрессионного уравнения $y = X\beta + \varepsilon$, удовлетворяющего условиям классической регрессионной модели. Найдите $\mathbb{E}(\tilde{\beta})$ и $\mathrm{Var}(\tilde{\beta})$.
10. Верно ли, что при невырожденном преобразовании факторов R^2 не меняется? А именно, пусть заданы две регрессионные модели: $y = X\beta + \varepsilon$ и $y = Z\alpha + u$, где y — вектор размера $n \times 1$, X и Z — матрицы размера $n \times k$, β и α — вектора рамзера $k \times 1$, ε и u — вектора размера $n \times 1$, а также $Z = XD$, $\det(D) \neq 0$. Верно ли, что коэффициенты детерминации представленных выше моделей равны между собой?
11. Верно ли, что при невырожденном преобразовании факторов RSS не меняется. А именно, пусть заданы две регрессиионные модели: $y = X\beta + \varepsilon$ и $y = Z\alpha + u$, где y — вектор размера $n \times 1$, X и Z — матрицы размера $n \times k$, β и α — вектора рамзера $k \times 1$, ε и u — вектора размера $n \times 1$, а также $Z = XD$, $\det(D) \neq 0$. Верно ли, что сумма квадратов остатков в представленных выше моделях равны между собой?
12. Пусть регрессионная модель $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, $i = 1, \ldots, n$, задана в матричном виде при помощи уравнения $y = X\beta + \varepsilon$, где $\beta = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix}^T$. Известно, что $\mathbb{E}\varepsilon = 0$ и $\mathrm{Var}(\varepsilon) = 4 \cdot I$. Известно также, что:
$$y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$
Для удобства расчётов ниже приведены матрицы:
$$X^T X = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} \text{ и } (X^T X)^{-1} = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1.5 \end{pmatrix}.$$
Найдите:
   (a) $\mathrm{Var}(\varepsilon_1)$
   (b) $\mathrm{Var}(\beta_1)$
   (c) $\mathrm{Var}(\hat{\beta}_1)$
   (d) $\widehat{\mathrm{Var}}(\hat{\beta}_1)$
   (e) $\mathbb{E}(\hat{\beta}_1^2) - \beta_1^2$
   (f) $\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3)$

(g) $\widehat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3)$

(h) $\text{Var}(\hat{\beta}_2 - \hat{\beta}_3)$

(i) $\widehat{\text{Var}}(\hat{\beta}_2 - \hat{\beta}_3)$

(j) $\text{Var}(\beta_2 - \beta_3)$

(k) $\text{Corr}(\hat{\beta}_2, \hat{\beta}_3)$

(l) $\widehat{\text{Corr}}(\hat{\beta}_2, \hat{\beta}_3)$

(m) $\mathbb{E}(\hat{\sigma}^2)$

(n) $\hat{\sigma}^2$

13. Пусть $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \varepsilon_i$ — регрессионная модель, где $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$, ошибки ε_i независимы и нормально распределены с $\mathbb{E}(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 I$. Для удобства расчётов даны матрицы: $X'X = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ и $(X'X)' = \begin{pmatrix} 0.3333 & -0.3333 & 0.0000 \\ -0.3333 & 1.3333 & -1.0000 \\ 0.0000 & -1.0000 & 2.0000 \end{pmatrix}$

(a) Укажите число наблюдений

(b) Укажите число регрессоров в модели, учитывая свободный член

(c) Найдите $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$

(d) Найдите $RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

(e) Методом МНК найдите оценку для вектора неизвестных коэффициентов

(f) Чему равен R^2 в модели? Прокомментируйте полученное значение с точки зрения качества оценённого уравнения регрессии

(g) Сформулируйте основную и альтернативную гипотезы, которые соответствуют тесту на значимость переменной x_1 в уравнении регрессии

(h) Протестируйте на значимость переменную x_1 в уравнении регрессии на уровне значимости 10%:

  i. Приведите формулу для тестовой статистики

  ii. Укажите распределение тестовой статистики

  iii. Вычислите наблюдаемое значение тестовой статистики

  iv. Укажите границы области, где основная гипотеза не отвергается

  v. Сделайте статистический вывод о значимости переменной x_1

(i) Найдите $P-$значение, соответствующее наблюдаемому значению тестовой статистики (T_{obs}) из предыдущего пункта. На основе полученного $P-$значения сделайте вывод о значимости переменной x_1

(j) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 \neq 1$:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики
   iii. Вычислите наблюдаемое значение тестовой статистики
   iv. Укажите границы области, где основная гипотеза не отвергается
   v. Сделайте статистический вывод

(k) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 > 1$:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики
   iii. Вычислите наблюдаемое значение тестовой статистики
   iv. Укажите границы области, где основная гипотеза не отвергается
   v. Сделайте статистический вывод

(l) На уровне значимости 10% проверьте гипотезу $H_0 : \beta_1 = 1$ против альтернативной $H_a : \beta_1 < 1$:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики
   iii. Вычислите наблюдаемое значение тестовой статистики
   iv. Укажите границы области, где основная гипотеза не отвергается
   v. Сделайте статистический вывод

(m) Сформулируйте основную гипотезу, которая соответствует тесту на значимость регрессии «в целом»

(n) На уровне значимости 5% проверьте гипотезу о значимости регрессии «в целом»:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики
   iii. Вычислите наблюдаемое значение тестовой статистики
   iv. Укажите границы области, где основная гипотеза не отвергается
   v. Сделайте статистический вывод

(o) Найдите $P-$значение, соответствующее наблюдаемому значению тестовой статистики (T_{obs}) из предыдущего пункта. На основе полученного $P-$значения сделайте вывод о значимости регрессии «в целом»

(p) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 \neq 2$:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики
   iii. Вычислите наблюдаемое значение тестовой статистики
   iv. Укажите границы области, где основная гипотеза не отвергается
   v. Сделайте статистический вывод

(q) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 > 2$:

   i. Приведите формулу для тестовой статистики
   ii. Укажите распределение тестовой статистики

iii. Вычислите наблюдаемое значение тестовой статистики

iv. Укажите границы области, где основная гипотеза не отвергается

v. Сделайте статистический вывод

(г) На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 < 2$:

i. Приведите формулу для тестовой статистики

ii. Укажите распределение тестовой статистики

iii. Вычислите наблюдаемое значение тестовой статистики

iv. Укажите границы области, где основная гипотеза не отвергается

v. Сделайте статистический вывод

14. Пусть $y = X\beta + \varepsilon$ — регрессионная модель, где $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$,

$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$, $\mathbb{E}(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 I$.

На уровне значимости 5% проверьте гипотезу $H_0 : \beta_1 + \beta_2 = 2$ против альтернативной $H_a : \beta_1 + \beta_2 \neq 2$:

(a) Приведите формулу для тестовой статистики

(b) Укажите распределение тестовой статистики

(c) Вычислите наблюдаемое значение тестовой статистики

(d) Укажите границы области, где основная гипотеза не отвергается

(e) Сделайте статистический вывод

15. По 13 наблюдениям Вася оценил модель со свободным членом, пятью количественными регрессорами и двумя качественными. Качественные регрессоры Вася правильно закодировал с помощью дамми-переменных. Одна качественная переменная принимала четыре значения, другая — пять.

(a) Найдите SSR, R^2

(b) Как выглядит матрица $X(X'X)^{-1}X'$?

(c) Почему 13 — несчастливое число?

16. В рамках классической линейной модели найдите все математические ожидания и все ковариационные матрицы всех пар случайных векторов: ε, y, $\hat{y}$, $\hat{\varepsilon}$, $\hat{\beta}$. Т.е. найдите $\mathbb{E}(\varepsilon)$, $\mathbb{E}(y)$, ... и $\mathrm{Cov}(\varepsilon, y)$, $\mathrm{Cov}(\varepsilon, \hat{y})$, ...
17. Найдите $\mathbb{E}(\sum(\varepsilon_i - \bar{\varepsilon})^2)$, $\mathbb{E}(RSS)$
18. Используя матрицы $P = X(X'X)^{-1}X'$ и $\pi = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$ запишите RSS, TSS и ESS в матричной форме
19. $\mathbb{E}(TSS)$, $\mathbb{E}(ESS)$ — громоздкие
20. Вася строит регрессию y на некий набор объясняющих переменных и константу. А на самом деле $y_i = \beta_1 + \varepsilon_i$. Чему равно $\mathbb{E}(TSS)$, $\mathbb{E}(RSS)$, $\mathbb{E}(ESS)$ в этом случае?
21. Рассмотрим классическую линейную модель. Являются ли векторы $\hat{\varepsilon}$ и $\hat{y}$ перпендикулярными? Найдите $\mathrm{Cov}(\hat{\varepsilon}, \hat{y})$

22. Чему в классической модели регрессии равны $\mathbb{E}(\varepsilon)$, $\mathbb{E}(\hat{\varepsilon})$? Верно ли что $\sum \varepsilon_i$ равна 0? Верно ли что $\sum \hat{\varepsilon}_i$ равна 0?
23. Найдите на Картинке все перпендикулярные векторы. Найдите на Картинке все прямоугольные треугольники. Сформулируйте для них теоремы Пифагора.
24. Покажите на Картинке TSS, ESS, RSS, R^2, $\mathrm{sCov}(\hat{y}, y)$
25. Предложите аналог R^2 для случая, когда константа среди регрессоров отсутствует. Аналог должен быть всегда в диапазоне $[0;1]$, совпадать с обычным R^2, когда среди регрессоров есть константа, равняться единице в случае нулевого $\hat{\varepsilon}$.
26. Вася оценил регрессию y на константу, x и z. А затем, делать ему нечего, регрессию y на константу и полученный $\hat{y}$. Какие оценки коэффициентов у него получатся? Чему будет равна оценка дисперсии коэффицента при $\hat{y}$? Почему оценка коэффициента неслучайна, а оценка её дисперсии положительна?
27. При каких условиях $TSS = ESS + RSS$?
28. Истинная модель имеет вид $y = X\beta + \varepsilon$. Вася оценивает модель $\hat{y} = X\hat{\beta}$ по первой части выборки, получает $\hat{\beta}_a$, по второй части выборки — получает $\hat{\beta}_b$ и по всей выборке — $\hat{\beta}_{tot}$. Как связаны между собой $\hat{\beta}_a$, $\hat{\beta}_b$, $\hat{\beta}_{tot}$? Как связаны между собой ковариационные матрицы $\mathrm{Var}(\hat{\beta}_a)$, $\mathrm{Var}(\hat{\beta}_b)$ и $\mathrm{Var}(\hat{\beta}_{tot})$?
29. Модель линейной регрессии имеет вид $y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + u_i$. Сумма квадратов остатков имеет вид $Q\left(\hat{\beta}_1, \hat{\beta}_2\right) = \sum_{i=1}^n (y_1 - \hat{\beta}_1 x_{i,1} - \hat{\beta}_2 x_{i,2})^2$.
    (a) Выпишите необходимые условия минимума суммы квадратов остатков
    (b) Найдите матрицу $X'X$ и вектор $X'y$ если матрица X имеет вид $X = \begin{pmatrix} x_{1,1} & x_{1,2} \\ \vdots & \vdots \\ x_{n,1} & x_{n,2} \end{pmatrix}$, а вектор y имеет вид $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$
    (c) Докажите, что необходимые условия равносильны матричному уравнению $X'X\hat{\beta} = X'y$, где $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$
    (d) Предполагая, что матрица $X'X$ обратима, найдите $\hat{\beta}$
30. Вася оценил исходную модель:
$$y_i = \beta_1 + \beta_2 x_i + u_i$$
Для надежности Вася стандартизировал переменные, т.е. перешёл к $y_i^* = (y_i - \bar{y})/s_y$ и $x_i^* = (x_i - \bar{x})/s_x$. Затем Вася оценил ещё две модели:
$$y_i^* = \beta_1' + \beta_2' x_i^* + u_i'$$
и
$$y_i^* = \beta_2'' x_i^* + u_i''$$
В решении можно считать s_x и s_y известными.
    (a) Найдите $\hat{\beta}_1'$
    (b) Как связаны между собой $\hat{\beta}_2$, $\hat{\beta}_2'$ и $\hat{\beta}_2''$?
    (c) Как связаны между собой $\hat{u}_i$, $\hat{u}_i'$ и $\hat{u}_i''$?
    (d) Как связаны между собой $\widehat{\mathrm{Var}}\left(\hat{\beta}_2\right)$, $\widehat{\mathrm{Var}}\left(\hat{\beta}_2'\right)$ и $\widehat{\mathrm{Var}}\left(\hat{\beta}_2''\right)$?

(e) Как выглядит матрица $\widehat{\mathrm{Var}}\left(\hat{\beta}'\right)$?

(f) Как связаны между собой t-статистики $t_{\hat{\beta}_2}$, $t_{\hat{\beta}'_2}$ и $t_{\hat{\beta}''_2}$?

(g) Как связаны между собой R^2, $R^{2\prime}$ и $R^{2\prime\prime}$?

(h) В нескольких предложениях прокомментируйте последствия перехода к стандартизированным переменным

31. Регрессионная модель задана в матричном виде при помощи уравнения $y = X\beta + \varepsilon$, где $\beta = (\beta_1, \beta_2, \beta_3)'$. Известно, что $\mathbb{E}(\varepsilon) = 0$ и $\mathrm{Var}(\varepsilon) = \sigma^2 \cdot I$. Известно также, что

$$y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Для удобства расчетов приведены матрицы

$$X'X = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \text{ и } (X'X)^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -1 & 0 \\ -1 & 4 & -3 \\ 0 & -3 & 6 \end{pmatrix}.$$

(a) Укажите число наблюдений.

(b) Укажите число регрессоров с учетом свободного члена.

(c) Запишите модель в скалярном виде

(d) Рассчитайте $TSS = \sum(y_i - \bar{y})^2$, $RSS = \sum(y_i - \hat{y}_i)^2$ и $ESS = \sum(\hat{y}_i - \bar{y})^2$.

(e) Рассчитайте при помощи метода наименьших квадратов $\hat{\beta}$, оценку для вектора неизвестных коэффициентов.

(f) Чему равен $\hat{\varepsilon}_5$, МНК-остаток регрессии, соответствующий 5-ому наблюдению?

(g) Чему равен R^2 в модели? Прокомментируйте полученное значение с точки зрения качества оцененного уравнения регрессии.

(h) Используя приведенные выше данные, рассчитайте несмещенную оценку для неизвестного параметра σ^2 регрессионной модели.

(i) Рассчитайте $\widehat{\mathrm{Var}}(\hat{\beta})$, оценку для ковариационной матрицы вектора МНК-коэффициентов $\hat{\beta}$.

(j) Найдите $\widehat{\mathrm{Var}}(\hat{\beta}_1)$, несмещенную оценку дисперсии МНК-коэффициента $\hat{\beta}_1$.

(k) Найдите $\widehat{\mathrm{Var}}(\hat{\beta}_2)$, несмещенную оценку дисперсии МНК-коэффициента $\hat{\beta}_2$.

(l) Найдите $\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$, несмещенную оценку ковариации МНК-коэффициентов $\hat{\beta}_1$ и $\hat{\beta}_2$.

(m) Найдите $\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2)$, $\widehat{\mathrm{Var}}(\hat{\beta}_1 - \hat{\beta}_2)$, $\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)$, $\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2 - 2\hat{\beta}_3)$

(n) Найдите $\widehat{\mathrm{Corr}}(\hat{\beta}_1, \hat{\beta}_2)$, оценку коэффициента корреляции МНК-коэффициентов $\hat{\beta}_1$ и $\hat{\beta}_2$.

(o) Найдите $s_{\hat{\beta}_1}$, стандартную ошибку МНК-коэффициента $\hat{\beta}_1$.

(p) Рассчитайте выборочную ковариацию y и $\hat{y}$.

(q) Найдите выборочную дисперсию y, выборочную дисперсию $\hat{y}$.

32. Теорема Фриша-Вау. Регрессоры разбиты на две группы: матрицу X_1 размера $n \times k_1$ и матрицу X_2 размера $n \times k_2$. Рассмотрим две процедуры:

M1. Строим регрессия вектора y на все регрессоры, т.е. оцениваем модель:

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

M2. Процедура из двух шагов:

i. Строим регрессию вектора y на все регрессоры первой группы и получаем вектор остатков $M_1 y$, где $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$. Строим регрессию каждого регрессора из второй группы на все регрессоры первой группы и получаем в каждом случае вектор остатков. Эти остатки можно записать матрицей $M_1 X_2$.

ii. Строим регрессию вектора $M_1 y$ на остатки $M_1 X_2$.

Другими словами мы оцениваем модель:

$$M_1 y = M_1 X_2 \gamma_2 + u$$

(a) Верно ли, что МНК оценки коэффициентов $\hat{\beta}_2$ и $\hat{\gamma}_2$ совпадают?

(b) Верно ли, что остатки в обеих регрессиях совпадают?

33. Всего имеется 100 наблюдений. Для первых 50-ти наблюдений $X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}$, $X'y = \begin{pmatrix} 300 & 2000 \end{pmatrix}'$, $y'y = 2100$. По последним 50-ти наблюдениям: $X'X = \begin{pmatrix} 50 & 300 \\ 300 & 2100 \end{pmatrix}$, $X'y = \begin{pmatrix} 300 & 2200 \end{pmatrix}'$, $y'y = 2500$. По первым 50-ти наблюдениям оценивается модель $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, по последним 50-ти наблюдениям оценивается модель $y_i = \gamma_1 + \gamma_2 x_i + \varepsilon_i$. Предположеним, что во всех 100 наблюдениях ε_i независимы и нормальны $N(0; \sigma^2)$. На уровне значимости 5% проверьте гипотезу $H_0: \beta = \gamma$.

5 Метод максимального правдоподобия — общая теория

Пусть
$X = (X_1, \ldots, X_n)$ — случайная выборка
$x = (x_1, \ldots, x_n)$ — реализация данной случайной выборки
$f_{X_i}(x_i, \theta)$ — плотность распределения случайной величины X_i, $i = 1, \ldots, n$
$\theta = (\theta_1, \ldots, \theta_k)$ — вектор неизвестных параметров
$\Theta \subseteq \mathbb{R}^k$ — множество допустимых значений вектора неизвестных параметров
$\mathrm{L}(\theta) = \prod_{i=1}^n f_{X_i}(x_i, \theta)$ — функция правдоподобия
$l(\theta) := \ln \mathrm{L}(\theta)$ — логарифмическая функция правдоподобия
Пусть требуется протестировать систему (нелинейных) ограничений относительно вектора неизвестных параметров

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ g_2(\theta) = 0 \\ \ldots \\ g_r(\theta) = 0 \end{cases}$$

где $g_i(\theta)$ — функция, которая задаёт i-ое ограничение на вектор параметров θ, $i = 1, \ldots, r$.

$$\frac{\partial g}{\partial \theta'} = \begin{bmatrix} \partial g_1/\partial \theta' \\ \partial g_2/\partial \theta' \\ \vdots \\ \partial g_r/\partial \theta' \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \ldots & \frac{\partial g_1}{\partial \theta_k} \\ \frac{\partial g_2}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_2} & \ldots & \frac{\partial g_2}{\partial \theta_k} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial g_r}{\partial \theta_1} & \frac{\partial g_r}{\partial \theta_2} & \ldots & \frac{\partial g_r}{\partial \theta_k} \end{bmatrix}$$

$$\frac{\partial g'}{\partial \theta} = \begin{bmatrix} \frac{\partial g_1'}{\partial \theta} & \frac{\partial g_2'}{\partial \theta} & \ldots & \frac{\partial g_r'}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_1} & \ldots & \frac{\partial g_r}{\partial \theta_1} \\ \frac{\partial g_2}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_2} & \ldots & \frac{\partial g_r}{\partial \theta_2} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial g_1}{\partial \theta_k} & \frac{\partial g_2}{\partial \theta_k} & \ldots & \frac{\partial g_r}{\partial \theta_k} \end{bmatrix}$$

$I(\theta) = -\mathbb{E}\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}\right) = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 l}{\partial\theta_1\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 l}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_2\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_2\partial\theta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 l}{\partial\theta_k\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_k\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_k\partial\theta_k} \end{bmatrix}$ — информационная матрица Фишера

$\frac{\partial l}{\partial\theta} = \begin{bmatrix} \frac{\partial l}{\partial\theta_1} \\ \frac{\partial l}{\partial\theta_2} \\ \vdots \\ \frac{\partial l}{\partial\theta_k} \end{bmatrix}$

$\Theta_{UR} := \Theta$ — множество допустимых значений вектора неизвестных параметров без учёта ограничений

$\Theta_R := \{\theta \in \Theta : g(\theta) = 0\}$ — множество допустимых значений вектора неизвестных параметров с учётом ограничений

$\hat{\theta}_{UR} \in \Theta_{UR}$ — точка максимума функции l на множестве Θ_{UR}

$\hat{\theta}_R \in \Theta_R$ — точка максимума функции l на множестве Θ_R

Тогда для тестирования гипотезыH_0 можно воспользоваться одной из следующих ниже статистик.

$LR := -2(l(\hat{\theta}_R) - l) \overset{a}{\sim} \chi^2_r$ — статистика отношения правдоподобия

$W := g'(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial\theta'}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g'}{\partial\theta}(\hat{\theta}_{UR})\right]^{-1} g(\hat{\theta}_{UR}) \overset{a}{\sim} \chi^2_r$ — статистика Вальда

$LM := \left[\frac{\partial l}{\partial\theta}(\hat{\theta}_R)\right]' \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial l}{\partial\theta}(\hat{\theta}_R)\right] \overset{a}{\sim} \chi^2_r$ — статистика множителей Лагранжа

1. Эта задача дублирована дальше. Строка для сохранения нумерации.
2. Дядя Вова (Владимир Николаевич) и Скрипач (Гедеван) зарабатывают на Плюке чатлы, чтобы купить гравицапу. Число заработанных за i-ый день чатлов имеет пуассоновское распределение, заработки за разные дни независимы. За прошедшие 100 дней они заработали 250 чатлов.

   (a) Оцените параметр λ пуассоновского распределения методом максимального правдоподобия

   (b) Сколько дней им нужно давать концерты, чтобы оценка вероятности купить гравицапу составила 0.99? Гравицапа стоит пол кц или 2200 чатлов.

   (c) Постройте 95% доверительный интервал для λ

   (d) Проверьте гипотезу о том, что средний дневной заработок равен 2 чатла с помощью теста отношения правдоподобия, теста Вальда, теста множителей Лагранжа

3. Инопланетянин Капп совершил вынужденную посадку на Землю. Каждый день он выходит на связь со своей далёкой планетой. Продолжительность каждого сеанса связи имеет экспоненциальное распределение с параметром λ. Прошедшие 100 сеансов связи в сумме длились 11 часов.

   (a) Оцените параметр λ экспоненциального распределения методом максимального правдоподобия

   (b) Постройте 95% доверительный интервал для λ

   (c) Проверьте гипотезу о том, что средняя продолжительность сеанса связи равна 5 минутам с помощью теста отношения правдоподобия, теста Вальда, теста множителей Лагранжа

4. [R] По ссылке http://people.reed.edu/~jones/141/Coal.html скачайте данные о количестве крупных аварий на английских угольных шахтах.

   (a) Методом максимального правдоподобия оцените две модели:

i. Пуассоновская модель: количества аварий независимы и имеют Пуассоновское распределение с параметром λ.

ii. Модель с раздутым нулём «zero inflated poisson model»: количества аварий независимы, с вероятностью p аварий не происходит вообще, с вероятностью $(1-p)$ количество аварий имеет Пуассоновское распределение с параметром λ. Смысл этой модели в том, что по сравнению с Пуассоновским распределением у события $\{X_i = 0\}$ вероятность выше, а пропорции вероятностей положительных количеств аварий сохраняются. В модели с раздутым нулём дисперсия и среднее количества аварий отличаются. Чему в модели с раздутым нулём равна $\mathbb{P}(X_i = 0)$?

(b) С помощью тестов множителей Лагранжа, Вальда и отношения правдоподобия проверьте гипотезу H_0: верна пуассоновская модель против H_a: верна модель с раздутым нулём

(c) Постройте доверительные интервалы для оценённых параметров в обоих моделях

(d) Постройте доверительный интервал для вероятности полного отсутствия аварий по обеим моделям

5. Совместное распределение величин X и Y задано функцией

$$f(x,y) = \frac{\theta(\beta y)^x e^{-(\theta+\beta)y}}{x!}$$

Величина X принимает целые неотрицательные значения, а величина Y — действительные неотрицательные. Имеется случайная выборка (X_1, Y_1), $\ldots (X_n, Y_n)$.

(a) С помощью метода максимального правдоподобия оцените θ и β

(b) С помощью метода максимального правдоподобия оцените $a = \theta/(\beta+\theta)$

6. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из нормального распределения с математическим ожиданием μ и дисперсией ν; $\mu \in \mathbb{R}$ и $\nu > 0$ — неизвестные параметры. Реализация случайной выборки $x = (x_1, \ldots, x_n)$ приведена ниже:

-2.80	-1.12	-2.27	-1.31	-0.98	-2.15	-1.52	-2.82	-1.19	0.87

При помощи теста отношения правдоподобия, теста Вальда и теста множителей Лагранжа протестируйте гипотезу:

$$H_0 : \begin{cases} \mu = 0 \\ \nu = 1 \end{cases}$$

на уровне значимости 5%.

7. Пусть p — неизвестная вероятность выпадения орла при бросании монеты. Из 100 испытаний 42 раза выпал «Орел» и 58 — «Решка». Протестируйте на 5%-ом уровне значимости гипотезу о том, что монетка — «правильная» с помощью:

(a) теста отношения правдоподобия

(b) теста Вальда

(c) теста множителей Лагранжа

8. Пусть $x = (x_1, \ldots, x_n)$ — реализация случайной выборки из распределения Пуассона с неизвестным параметром $\lambda > 0$. Известно, что выборочное среднее $\overline{x}$ по 80 наблюдениям равно 1.7. Протестируйте на 5%-ом уровне значимости гипотезу $H_0 : \lambda = 2$ с помощью

(a) теста отношения правдоподобия

(b) теста Вальда

(c) теста множителей Лагранжа

6 Логит и пробит

1. Случайная величина X имеет логистическое распределение, если её функция плотности имеет вид $f(x) = e^{-x}/(1+e^{-x})^2$.

   (a) Является ли $f(x)$ чётной?

   (b) Постройте график $f(x)$

   (c) Найдите функцию распределения, $F(x)$

   (d) Найдите $\mathbb{E}(X)$, $\mathrm{Var}(X)$

   (e) На какое известный закон распределения похож логистический?

2. Логит модель часто формулируют в таком виде:

$$y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$$

   где ε_i имеет логистическое распределение, и

$$y_i = \begin{cases} 1, & y_i^* \geqslant 0 \\ 0, & y_i^* < 0 \end{cases}$$

   (a) Выразите $\mathbb{P}(y_i = 1)$ с помощью логистической функции распределения

   (b) Найдите $\ln\left(\frac{\mathbb{P}(y_i=1)}{\mathbb{P}(y_i=0)}\right)$

3. [R] Сравните на одном графике

   (a) Функции плотности логистической и нормальной $N(0, \pi^2/3)$ случайных величин

   (b) Функции распределения логистической и нормальной $N(0, \pi^2/3)$ случайных величин

4. Как известно, Фрекен Бок любит пить коньяк по утрам. За прошедшие 4 дня она записала, сколько рюмочек коньяка выпила утром, x_i, и видела ли она в этот день привидение, y_i,

y_i	1	0	1	0
x_i	2	1	3	0

   Зависимость между y_i и x_i описывается логит-моделью,

$$\ln\left(\frac{\mathbb{P}(y_i=1)}{\mathbb{P}(y_i=0)}\right) = \beta_1 + \beta_2 x_i$$

   (a) Выпишите в явном виде логарифмическую функцию максимального правдоподобия

   (b) [R] Найдите оценки параметров β_1 и β_2

5. При оценке логит модели

$$\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$$

   оказалось, что $\hat{\beta}_1 = 0.7$ и $\hat{\beta}_2 = 3$. Найдите максимальный предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$.

6. Винни-Пух знает, что мёд бывает правильный, $honey_i = 1$, и неправильный, $honey_i = 0$. Пчёлы также бывают правильные, $bee_i = 1$, и неправильные, $bee_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$honey_i = 1$	$honey_i = 0$
$bee_i = 1$	12	36
$bee_i = 0$	32	20

Используя метод максимального правдоподобия Винни-Пух хочет оценить логит-модель для прогнозирования правильности мёда с помощью правильности пчёл:

$$\ln\left(\frac{\mathbb{P}(honey_i = 1)}{\mathbb{P}(honey_i = 0)}\right) = \beta_1 + \beta_2 bee_i$$

(a) Выпишите функцию правдоподобия для оценки параметров β_1 и β_2

(b) Оцените неизвестные параметры

(c) С помощью теста отношения правдоподобия проверьте гипотезу о том, правильность пчёл не связана с правильностью мёда на уровне значимости 5%.

(d) Держась в небе за воздушный шарик, Винни-Пух неожиданно понял, что перед ним неправильные пчёлы. Помогите ему оценить вероятность того, что они делают неправильный мёд.

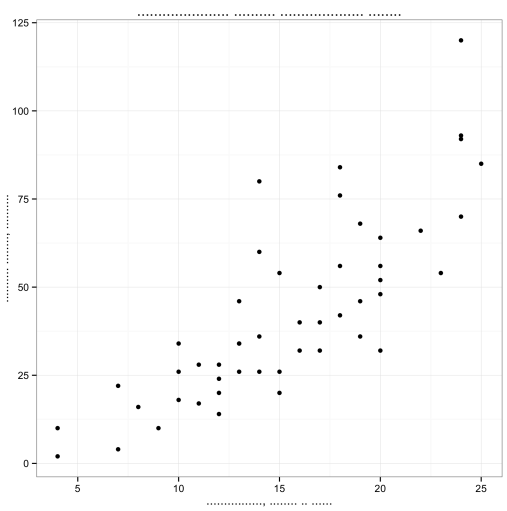
7 Мультиколлинеарность

Задача 7.1.
Сгенерируйте данные так, чтобы при оценке модели $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$ оказывалось, что по отдельности оценки коэффициентов $\hat{\beta}_2$ и $\hat{\beta}_3$ незначимы, но модель в целом — значима.
Задача 7.2.
В этом задании нужно сгенерировать зависимую переменную y и два регрессора x и z.

1. Сгенерируйте данные так, чтобы корреляция между регрессорами x и z была больше 0.9, и проблема мультиколлинеарности есть, т.е. по отдельности регрессоры не значимы, но регрессия в целом — значима.
2. А теперь сгенерируйте данные так, чтобы корреляция между регрессорами была по-прежнему больше 0.9, но проблемы мультиколлинеарности бы не было, т.е. все коэффициенты были бы значимы.
3. Есть несколько способов, как изменить генерации случайных величин, чтобы перейти от ситуации «a» к ситуации «b». Назовите хотя бы два.

Задача 7.3.
Исследуем зависимость длины тормозного пути автомобиля от скорости по историческим данным 1920-х годов.

```
h <- cars ggplot(h,aes(x=speed,y=dist))+geom_point()+labs(title=”,x=”, ,y=”, )
```



```
speed.mean <- mean(hspeed)
```

Построим результаты оценивания нецентрированной регрессии:

```
cars.model <- lm(dist speed+I(speed2)+I(speed3),data= h)cars.table< −as.table(coeftest(cars.m
```

с тремя переменными руками громоздко делать, а с двумя вроде не видно мультик.

```
xtable(cars.table)
```

	Estimate	Std. Error	t value	Pr(>|t|)
Константа	-19.51	28.41	-0.69	0.50
speed	6.80	6.80	1.00	0.32
speed¦2	-0.35	0.50	-0.70	0.49
speed¦3	0.01	0.01	0.91	0.37

Ковариационная матрица коэффициентов имеет вид:

```
cars.vcov <- vcov(cars.model) rownames(cars.vcov) <-c("Константа
,"speed
,"speed2,"speed3
) colnames(cars.vcov) <-c("Константа
,"speed
,"speed2,"speed3
) xtable(cars.vcov)
```

	Константа	speed	speed¦2	speed¦3
Константа	806.86	-186.20	12.88	-0.27
speed	-186.20	46.26	-3.35	0.07
speed¦2	12.88	-3.35	0.25	-0.01
speed¦3	-0.27	0.07	-0.01	0.00

1. Проверьте значимость всех коэффициентов и регрессии в целом
2. Постройте 95%-ый доверительный интервал для $\mathbb{E}(dist)$ при $speed = 10$
3. Постройте 95%-ый доверительный интервал для $\mathbb{E}(ddist/dspeed)$ при $speed = 10$
4. Как выглядит уравнение регрессии, если вместо $speed$ использовать центрированную скорость? Известно, что средняя скорость равна 15.4
5. С помощью регрессии с центрированной скоростью ответьте на предыдущие вопросы.

Задача 7.4.

Пионеры, Крокодил Гена и Чебурашка собирали металлолом несколько дней подряд. В распоряжение иностранной шпионки, гражданки Шапокляк, попали ежедневные данные по количеству собранного металлолома: вектор g — для Крокодила Гены, вектор h — для Чебурашки и вектор x — для Пионеров. Гена и Чебурашка собирали вместе, поэтому выборочная корреляция $\mathrm{sCorr}(g, h) = -0.9$. Гена и Чебурашка собирали независимо от Пионеров, поэтому выборочные корреляции $\mathrm{sCorr}(g, x) = 0$, $\mathrm{sCorr}(h, x) = 0$. Если регрессоры g, h и x центрировать и нормировать, то получится матрица $\tilde{X}$.

1. Найдите параметр обусловленности матрицы $(\tilde{X}'\tilde{X})$
2. Вычислите одну или две главные компоненты (выразите их через вектор-столбцы матрицы $\tilde{X}$), объясняющие не менее 70% общей выборочной дисперсии регрессоров
3. Шпионка Шапокляк пытается смоделировать ежедневный выпуск танков, y. Выразите коэффициенты регрессии $y = \beta_1 + \beta_2 g + \beta_3 h + \beta_4 x + \varepsilon$ через коэффициенты регрессии на главные компоненты, объясняющие не менее 70% общей выборочной дисперсии.

Задача 7.5.
Для модели $y_i = \beta x_i + \varepsilon$ рассмотрите модель Ridge regression с коэффициентом λ.

1. Выведите формулу для $\hat{\beta}_{RR}$
2. Найдите $\mathbb{E}(\hat{\beta}_{RR})$, смещение оценки $\hat{\beta}_{RR}$,
3. Найдите $\mathrm{Var}(\hat{\beta}_{RR})$, $MSE(\hat{\beta}_{RR})$
4. Всегда ли оценка $\hat{\beta}_{RR}$ смещена?
5. Всегда ли оценка $\hat{\beta}_{RR}$ имеет меньшую дисперсию, чем $\hat{\beta}_{ols}$?
6. Найдите такое λ, что $MSE(\hat{\beta}_{RR}) < MSE(\hat{\beta}_{ols})$

Задача 7.6.
Известно, что в модели $y = X\beta + \varepsilon$ все регрессоры ортогональны.

1. Как выглядит матрица $X'X$ в случае ортогональных регрессоров?
2. Выведите $\hat{\beta}_{rr}$ в явном виде
3. Как связаны между собой $\hat{\beta}_{rr}$ и $\hat{\beta}_{ols}$?

Задача 7.7.
Для модели $y_i = \beta x_i + \varepsilon_i$ выведите в явном виде $\hat{\beta}_{lasso}$.

Задача 7.8.
Предположим, что для модели $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$ выборочная корреляционная матрица регрессоров x_2, x_3, x_4 имеет вид

$$C = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}$$

1. Найдите такое значение $r^* \in (-1; 1)$ коэффициента корреляции, при котором $\det C = 0$.
2. Найдите собственные значения и собственные векторы матрицы C при корреляции равной найденному r^*.
3. Найдите число обусловленности матрицы C при корреляции равной найденному r^*.
4. Сделайте вывод о наличии мультиколлинеарности в модели при корреляции равной найденному r^*.

Задача 7.9.
Предположим, что для модели $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$ выборочная корреляционная матрица регрессоров x_2, x_3, x_4 и x_5 имеет вид

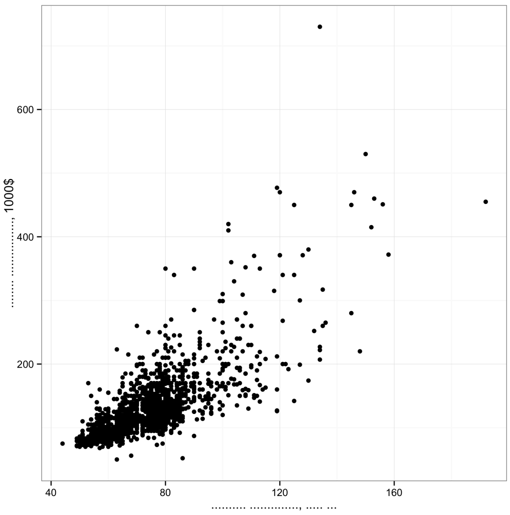
$$C = \begin{pmatrix} 1 & r & r & r \\ r & 1 & r & r \\ r & r & 1 & r \\ r & r & r & 1 \end{pmatrix}$$

1. Найдите такое значение $r^* \in (-1; 1)$ коэффициента корреляции, при котором $\det C = 0$.
2. Найдите собственные значения и собственные векторы матрицы C при корреляции равной найденному r^*.
3. Найдите число обусловленности матрицы C при корреляции равной найденному r^*.
4. Сделайте вывод о наличии мультиколлинеарности в модели при корреляции равной найденному r^*.

8 Гетероскедастичность

1. Что такое гетероскедастичность? Гомоскедастичность?
2. Диаграмма рассеяния стоимости квартиры в Москве (в 1000$) и общей площади квартиры имеет вид:

```r
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?

3. 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 $\text{Var}(\varepsilon_i) = \sigma^2 \cdot x_i^2$.

   (a) Найдите оценки $\hat{\beta}_{ols}$ с помощью МНК и их ковариационную матрицу

   (b) Найдите оценки $\hat{\beta}_{gls}$ с помощью обобщенного МНК и их ковариационную матрицу

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

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%.

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%.

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%.

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%.

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

   (a) устранить гетероскедастичность?

   (b) корректно тестировать гипотезы относительно коэффициентов регрессии в условиях гетероскедастичности?

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

   (a) устранить гетероскедастичность?

   (b) корректно тестировать гипотезы относительно коэффициентов регрессии в условиях гетероскедастичности?

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.

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.

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

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

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

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

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

    (b) Верно ли, что $\hat{\beta}_{WLS} = \hat{\beta}_{GLS}$?

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

    (d) В явном виде выпишите $\hat{\beta}_{2,WLS}$

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

    (a) вероятность того, что статистика Уайта окажется больше 10,

    (b) ожидаемое значение статистики Уайта,

    (c) дисперсию статистики Уайта.

9 Ошибки спецификации

1. По 25 наблюдениям при помощи метода наименьших квадратов оценена модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, для которой $RSS = 73$. При помощи вспомогательной регрессии $\hat{\hat{y}} = \hat{\gamma}_1 + \hat{\gamma}_2 x + \hat{\gamma}_3 z + \hat{\gamma}_4 \hat{y}^2$, для которой $RSS = 70$, выполните тест Рамсея на уровне значимости 5%.

2. По 20 наблюдениям при помощи метода наименьших квадратов оценена модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, для которой $R^2 = 0.7$. При помощи вспомогательной регрессии $\hat{\hat{y}} = \hat{\gamma}_1 + \hat{\gamma}_2 x + \hat{\gamma}_3 z + \hat{\gamma}_4 \hat{y}^2$, для которой $R^2 = 0.75$, выполните тест Рамсея на уровне значимости 5%.
3. По 30 наблюдениям при помощи метода наименьших квадратов оценена модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, для которой $RSS = 150$. При помощи вспомогательной регрессии $\hat{\hat{y}} = \hat{\gamma}_1 + \hat{\gamma}_2 x + \hat{\gamma}_3 z + \hat{\gamma}_4 \hat{y}^2 + \hat{\gamma}_5 \hat{y}^3$, для которой $RSS = 120$, выполните тест Рамсея на уровне значимости 5%.
4. По 35 наблюдениям при помощи метода наименьших квадратов оценена модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, для которой $R^2 = 0.7$. При помощи вспомогательной регрессии $\hat{\hat{y}} = \hat{\gamma}_1 + \hat{\gamma}_2 x + \hat{\gamma}_3 z + \hat{\gamma}_4 \hat{y}^2 + \hat{\gamma}_5 \hat{y}^3$, для которой $R^2 = 0.8$, выполните тест Рамсея на уровне значимости 5%.
5. Используя 80 наблюдений, исследователь оценил две конкурирующие модели: $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_1 = 36875$ и $\widehat{\ln y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_2 = 122$.
   Выполнив преобразование $y_i^* = y_i / \sqrt[n]{\prod y_i}$, исследователь также оценил две вспомогательные регрессии: $\hat{y}^* = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_1^* = 239$ и $\widehat{\ln y^*} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_2^* = 121$.
   Завершите тест Бокса-Кокса на уровне значимости 5%.
6. Используя 40 наблюдений, исследователь оценил две конкурирующие модели: $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_1 = 250$ и $\widehat{\ln y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_2 = 12$.
   Выполнив преобразование $y_i^* = y_i / \sqrt[n]{\prod y_i}$, исследователь также оценил две вспомогательные регрессии: $\hat{y}^* = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_1^* = 20$ и $\widehat{\ln y^*} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$, в которой $RSS_2^* = 25$.
   Завершите тест Бокса-Кокса на уровне значимости 5%.
7. Почему при реализации теста Бокса-Кокса на компьютере предпочтительнее использовать формулу $y_i^* = \exp(\ln y_i - \sum \ln y_i / n)$, а не формулу $y_i^* = y_i / \sqrt[n]{\prod y_i}$?
8. Обследовав выборку из 27 домохозяйств, исследователь оценил уравнение регрессии:
   $$\frac{\widehat{Exp_i}}{Size_i} = 926 + 235 \frac{1}{Size_i} + 0.3 \frac{Income_i}{Size_i}$$
   где Exp_i — месячные затраты i-го домохозяйства на питание в рублях, $Income_i$ — месячный доход домохозяйства (также в рублях), $Size_i$ — число членов домохозяйства. Известен коэффициент детерминации, $R^2 = 0.3$.
   (a) Каково, согласно оценённой модели, ожидаемое различие в затратах на питание между двумя домохозяйствами с одинаковым доходом, первое из которых больше второго на одного человека?
   (b) Известно, что нормировка переменных модели на размер семьи $Size_i$ была проведена с целью устранения гетероскедастичности в модели $Exp_i = \beta_1 + \beta_2 Size_i + \beta_3 Income_i + \varepsilon_i$. Какое предположение сделал исследователь о виде гетероскедастичности?
   (c) Для проверки правильности выбранной спецификации было оценено ещё одно уравнение:
   $$\frac{\widehat{\widehat{Exp_i}}}{Size_i} = 513 + 1499 \frac{1}{Size_i} + 0.5 \frac{Income_i}{Size_i} - 0.001 \left(\frac{\widehat{Exp_i}}{Size_i} \right)^2$$
   Известно, что $R^2 = 0.4$. Даёт ли эта проверка основание считать модель исследователя неверно специфицированной? Используйте уровень значимости 1%
9. Мартовский Заяц и Безумный Шляпник почти всё 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 25 дней, заметила, что если

оценить зависимость выпитого чая от закуски для Мартовского Зайца и Шляпника, то получится регрессия с $RSS = 11.5$:

$$\widehat{Tea}_i = 6 + 0.5 Biscuit_i + 1.5 Cake_i$$

Чтобы понять, удачную ли модель она построила, Алиса оценила ещё одну регрессию с $RSS = 9.5$:

$$\widehat{\widehat{Tea}}_i = 12.7 + 0.65 Biscuit_i - 0.8 Cake_i - 0.59 \widehat{Tea}_i^2 + 0.03 \widehat{Tea}_i^3$$

Помогите Алисе понять, верную ли спецификацию модели она выбрала

(a) Проведите подходящий тест

(b) Сформулируйте основную и альтернативную гипотезы

(c) Алиса решила проверить первоначальную короткую модель на наличие гетероскедастичности с помощью теста Уайта. Выпишите уравнение регрессии, которое она должна оценить.

10 Временные ряды

1. Что такое автокорреляция?
2. На графике представлены данные по уровню озера Гурóн в футах в 1875-1972 годах:

```
ggplot(df,aes(x=obs,y=level))+geom_line()+
    labs(x="Год",ylab="Уровень озера (футы)")
```

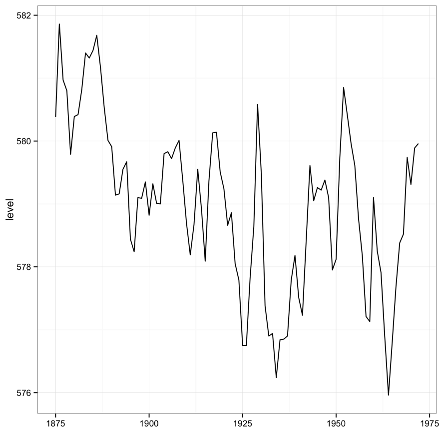
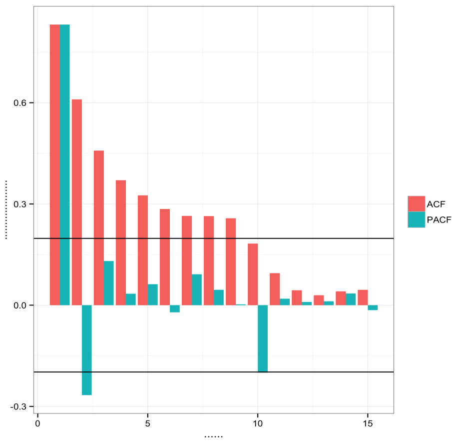


График автокорреляционной и частной автокорреляционной функций:

```
ggplot(acfs.df,aes(x=lag,y=acf,fill=acf.type))+
    geom_histogram(position="dodge",stat="identity")+
  xlab("Лаг")+ylab("Корреляция") +
  guides(fill=guide_legend(title=NULL))+
  geom_hline(yintercept=1.96/sqrt(nrow(df)))+
  geom_hline(yintercept=-1.96/sqrt(nrow(df)))
```



(a) Судя по графикам, какие модели класса ARMA или ARIMA имеет смысл оценить?

(b) По результатам оценки некоей модели ARMA с двумя параметрами, исследователь посчитал оценки автокорреляционной функции для остатков модели. Известно, что для остатков модели первые три выборочные автокорреляции равны соответственно 0.0047, -0.0129 и -0.063. С помощью подходящей статистики проверьте гипотезу о том, что первые три корреляции ошибок модели равны нулю.

3. Винни-Пух пытается выявить закономерность в количестве придумываемых им каждый день ворчалок. Винни-Пух решил разобраться, является ли оно стационарным процессом, для этого он оценил регрессию

$$\Delta\hat{y}_t = \underset{(0.5)}{4.5} - \underset{(0.1)}{0.4} y_{t-1} + \underset{(0.5)}{0.7} \Delta y_{t-1}$$

Из-за опилок в голове Винни-Пух забыл, какой тест ему нужно провести, то ли Доктора Ватсона, то ли Дикого Фуллера.

(a) Аккуратно сформулируйте основную и альтернативную гипотезы

(b) Проведите подходящий тест на уровне значимости 5%

(c) Сделайте вывод о стационарности ряда

(d) Почему Сова не советовала Винни-Пуху пользоваться широко применяемым в Лесу t-распределением?

4. Рассматривается модель $y_t = \beta_1 + \beta_2 x_{t1} + \ldots + \beta_k x_{tk} + \varepsilon_t$. Ошибки ε_t гомоскедастичны, но в них возможно присутствует автокорреляция первого порядка, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$. При известном числе наблюдений T на уровне значимости 5% сделайте статистический вывод о наличии автокорреляции.

(a) $T = 25$, $k = 2$, $DW = 0.8$

(b) $T = 30$, $k = 3$, $DW = 1.6$

(c) $T = 50$, $k = 4$, $DW = 1.8$

(d) $T = 100$, $k = 5$, $DW = 1.1$

5. По 100 наблюдениям была оценена модель линейной регрессии $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$. Оказалось, что $RSS = 120$, $\hat{\varepsilon}_1 = -1$, $\hat{\varepsilon}_{100} = 2$, $\sum_{t=2}^{100} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} = -50$. Найдите DW и ρ.

6. Применяется ли статистика Дарбина-Уотсона для выявления автокорреляции в следующих моделях

(a) $y_t = \beta_1 x_t + \varepsilon_t$

(b) $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$

(c) $y_t = \beta_1 + \beta_2 y_{t-1} + \varepsilon_t$

(d) $y_t = \beta_1 + \beta_2 t + \beta_3 y_{t-1} + \varepsilon_t$

(e) $y_t = \beta_1 t + \beta_2 x_t + \varepsilon_t$

(f) $y_t = \beta_1 + \beta_2 t + \beta_3 x_t + \beta_4 x_{t-1} + \varepsilon_t$

7. По 21 наблюдению была оценена модель линейной регрессии $\underset{(se)}{\hat{y}} = \underset{(0.3)}{1.2} + \underset{(0.18)}{0.9} \cdot y_{t-1} + \underset{(0.01)}{0.1} \cdot t$, $R^2 = 0.6$, $DW = 1.21$. Протестируйте гипотезу об отсутствии автокорреляции ошибок на уровне значимости 5%.
8. По 24 наблюдениям была оценена модель линейной регрессии $\underset{(se)}{\hat{y}} = \underset{(0.01)}{0.5} + \underset{(0.02)}{2} \cdot t$, $R^2 = 0.9$, $DW = 1.3$. Протестируйте гипотезу об отсутствии автокорреляции ошибок на уровне значимости 5%.
9. По 32 наблюдениям была оценена модель линейной регрессии $\underset{(se)}{\hat{y}} = \underset{(2.5)}{10} + \underset{(0.5)}{2.5} \cdot t - \underset{(0.01)}{0.1} \cdot t^2$, $R^2 = 0.75$, $DW = 1.75$. Протестируйте гипотезу об отсутствии автокорреляции ошибок на уровне значимости 5%.
10. Рассмотрим модель $y_t = \beta_1 + \beta_2 x_{t1} + \ldots + \beta_k x_{tk} + \varepsilon_t$, где ε_t подчиняются автокорреляционной схеме первого порядка, т.е.

    (a) $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, $-1 < \rho < 1$

    (b) $\mathrm{Var}(\varepsilon_t) = const$, $\mathbb{E}(\varepsilon_t) = const$

    (c) $\mathrm{Var}(u_t) = \sigma^2$, $\mathbb{E}(u_t) = 0$

    (d) Величины u_t независимы между собой

    (e) Величины u_t и ε_s независимы, если $t \geqslant s$

    Найдите:

    (a) $\mathbb{E}(\varepsilon_t)$, $\mathrm{Var}(\varepsilon_t)$

    (b) $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+h})$

    (c) $\mathrm{Corr}(\varepsilon_t, \varepsilon_{t+h})$

11. Ошибки в модели $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ являются автокоррелированными первого порядка, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$. Шаман-эконометрист Ойуун выполняет два камлания-преобразования. Поясните смысл камланий:

    (a) Камлание А, при $t \geqslant 2$, Ойуун преобразует уравнение к виду $y_t - \rho y_{t-1} = \beta_1(1-\rho) + \beta_2(x_t - \rho x_{t-1}) + \varepsilon_t - \rho\varepsilon_{t-1}$

    (b) Камлание Б, при $t = 1$, Ойуун преобразует уравнение к виду $\sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}\beta_1 + \sqrt{1-\rho^2}\beta_2 x_1 + \sqrt{1-\rho^2}\varepsilon_1$.

12. Пусть y_t — стационарный процесс. Верно ли, что стационарны:

    (a) $z_t = 2y_t$

    (b) $z_t = y_t + 1$

    (c) $z_t = \Delta y_t$

    (d) $z_t = 2y_t + 3y_{t-1}$

13. Известно, что временной ряд y_t порожден стационарным процессом, задаваемым соотношением $y_t = 1 + 0.5y_{t-1} + \varepsilon_t$. Имеется 1000 наблюдений. Вася построил регрессию y_t на константу и y_{t-1}. Петя построил регрессию на константу и y_{t+1}. Какие примерно оценки коэффициентов они получат?
14. Рассмотрим следующий AR(1)-ARCH(1) процесс, $y_t = 1 + 0.5y_{t-1} + \varepsilon_t$, $\varepsilon_t = \nu_t \cdot \sigma_t$
    ν_t независимые $N(0;1)$ величины.
    $\sigma_t^2 = 1 + 0.8\varepsilon_{t-1}^2$
    Также известно, что $y_{100} = 2$, $y_{99} = 1.7$

(a) Найдите $\mathbb{E}_{100}(\varepsilon_{101}^2)$, $\mathbb{E}_{100}(\varepsilon_{102}^2)$, $\mathbb{E}_{100}(\varepsilon_{103}^2)$, $\mathbb{E}(\varepsilon_t^2)$

(b) $\mathrm{Var}(y_t)$, $\mathrm{Var}(y_t|\mathcal{F}_{t-1})$

(c) Постройте доверительный интервал для y_{101}:

i. проигнорировав условную гетероскедастичность

ii. учтя условную гетерескедастичность

15. Пусть x_t, $t = 0, 1, 2, ...$ - случайный процесс и $y_t = (1 + \mathrm{L})^t x_t$. Выразите x_t с помощью y_t и оператора лага L.
16. Пусть F_n - последовательность чисел Фибоначчи. Упростите величину

$$F_1 + C_5^1 F_2 + C_5^2 F_3 + C_5^3 F_4 + C_5^4 F_5 + C_5^5 F_6$$

17. Пусть y_t, $t = ... - 2, -1, 0, 1, 2, ...$ - случайный процесс. И $y_t = x_{-t}$. Являются ли верными рассуждения?

(a) $\mathrm{L}y_t = \mathrm{L}x_{-t} = x_{-t-1}$

(b) $\mathrm{L}y_t = y_{t-1} = x_{-t+1}$

18. Представьте процесс AR(1), $y_t = 0.9y_{t-1} - 0.2y_{t-2} + \varepsilon_t$, ε ~WN(0;1) в виде модели состояние-наблюдение.

а) Выбрав в качестве состояний вектор $\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$

б) Выбрав в качестве состояний вектор $\begin{pmatrix} y_t \\ \hat{y}_{t,1} \end{pmatrix}$

Найдите дисперсии ошибок состояний

19. Представьте процесс MA(1), $y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$, ε ~WN(0;1) в виде модели состояние-наблюдение.

a) $\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix}$

b) $\begin{pmatrix} \varepsilon_t + 0.5\varepsilon_{t-1} \\ 0.5\varepsilon_t \end{pmatrix}$

20. Представьте процесс ARMA(1,1), $y_t = 0.5y_{t-1} + \varepsilon_t + \varepsilon_{t-1}$, ε ~WN(0;1) в виде модели состояние-наблюдение.

Вектор состояний имеет вид x_t, x_{t-1}, где $x_t = \frac{1}{1-0.5L}\varepsilon_t$

21. Рекурсивные коэффициенты

(a) Оцените модель вида $y_t = a + b_t x_t + \varepsilon_t$, где $b_t = b_{t-1}$.

(b) Сравните графики filtered state и smoothed state.

(c) Сравните финальное состояние b_T с коэффициентом в обычной модели линейной регрессии, $y_t = a + bx_t + \varepsilon_t$.

22. Пусть u_t — независимые нормальные случайные величины с математическим ожиданием 0 и дисперсией σ^2. Известно, что $\varepsilon_1 = u_1$, $\varepsilon_t = u_1 + u_2 + \ldots + u_t$. Рассмотрим модель $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$.

(a) Найдите $\mathrm{Var}(\varepsilon_t)$, $\mathrm{Cov}(\varepsilon_t, \varepsilon_s)$, $\mathrm{Var}(\varepsilon)$

(b) Являются ли ошибки ε_t гетероскедастичными?

(c) Являются ли ошибки ε_t автокоррелированными?

(d) Предложите более эффективную оценку вектора коэффициентов регрессии по сравнению МНК-оценкой.

(e) Результаты предыдущего пункта подтвердите симуляциями Монте-Карло на компьютере.

23. Найдите безусловная дисперсия GARCH-процессов
    (a) $\varepsilon_t = \sigma_t \cdot z_t$, $\sigma_t^2 = 0.1 + 0.8\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$
    (b) $\varepsilon_t = \sigma_t \cdot z_t$, $\sigma_t^2 = 0.4 + 0.7\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$
    (c) $\varepsilon_t = \sigma_t \cdot z_t$, $\sigma_t^2 = 0.2 + 0.8\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$
24. Являются ли верными следующие утверждения?
    (a) GARCH-процесс является процессом белого шума, условная дисперсия которого изменяется во времени
    (b) Модель GARCH(1,1) предназначена для прогнозирования меры изменчивости цены финансового инструмента, а не для прогнозирования самой цены инструмента
    (c) При помощи GARCH-процесса можно устранять гетероскедастичность
    (d) Безусловная дисперсия GARCH-процесса изменяется во времени
    (e) Модель GARCH(1,1) может быть использована для прогнозирования волатильности финансовых инструментов на несколько торговых недель вперёд
25. Рассмотрим GARCH-процесс $\varepsilon_t = \sigma_t \cdot z_t$, $\sigma_t^2 = k + g_1\sigma_{t-1}^2 + a_1\varepsilon_{t-1}^2$. Найдите
    (a) $\mathbb{E}(z_t)$, $\mathbb{E}(z_t^2)$, $\mathbb{E}(\varepsilon_t)$, $\mathbb{E}(\varepsilon_t^2)$
    (b) $\mathrm{Var}(z_t)$, $\mathrm{Var}(\varepsilon_t)$, $\mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1})$
    (c) $\mathbb{E}(\varepsilon_t \mid \mathcal{F}_{t-1})$, $\mathbb{E}(\varepsilon_t^2 \mid \mathcal{F}_{t-1})$, $\mathbb{E}(\sigma_t^2 \mid \mathcal{F}_{t-1})$
    (d) $\mathbb{E}(z_t z_{t-1})$, $\mathbb{E}(z_t^2 z_{t-1}^2)$, $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1})$, $\mathrm{Cov}(\varepsilon_t^2, \varepsilon_{t-1}^2)$
    (e) $\lim_{h\to\infty} \mathbb{E}(\sigma_{t+h}^2 \mid \mathcal{F}_t)$
26. Используя 500 наблюдений дневных логарифмических доходностей y_t , была оценена GARCH(1,1)-модель: $\hat{y}_t = -0.000708 + \hat{\varepsilon}_t$, $\varepsilon_t = \sigma_t \cdot z_t$, $\sigma_t^2 = 0.000455 + 0.6424\sigma_{t-1}^2 + 0.2509\varepsilon_{t-1}^2$. Также известно, что $\hat{\sigma}_{499}^2 = 0.002568$, $\hat{\varepsilon}_{499}^2 = 0.000014$, $\hat{\varepsilon}_{500}^2 = 0.002178$. Найдите
    (a) $\hat{\sigma}_{500}^2$, $\hat{\sigma}_{501}^2$, $\hat{\sigma}_{502}^2$
    (b) Волатильность в годовом выражении в процентах, соответствующую наблюдению с номером $t = 500$
27. Докажите, что в условиях автокорреляции МНК- оценки остаются несмещенными.
28. Продавец мороженного оценил динамическую модель объёмов продаж:

$$\ln \hat{Q}_t = 26.7 + 0.2 \ln \hat{Q}_{t-1} - 0.6 \ln P_t$$

    Здесь Q_t — число проданных в день t вафельных стаканчиков, а P_t — цена одного стаканчика в рублях. Продавец также рассчитал остатки $\hat{e}_t$.
    (a) Чему, согласно полученным оценкам, равна долгосрочная эластичность объёма продаж по цене?
    (b) Предположим, что продавец решил проверить наличие автокорреляции первого порядка с помощью теста Бройша-Годфри. Выпишите уравнение регрессии, которое он должен оценить.
29. Данные описываются моделью $y_t = \beta + \varepsilon_t$, где ε_t — стационарный AR(1) процесс, $\varepsilon_t = u_t + \rho\varepsilon_{t-1}$, $u \sim N(0; \sigma^2 \cdot I)$. Имеются наблюдения $y' = (1, 2, 0, 0, 1)$.
    (a) Выпишите функцию правдоподобия для оценивания параметров β, σ^2, ρ
    (b) Найдите оценки неизвестных параметров

11 SVM

1. Имеются три наблюдения A, B и C:

	x	y
A	1	-2
B	2	1
C	3	0

   (a) Найдите расстояние AB и косинус угла ABC

   (b) Найдите расстояние AB и косинус угла ABC в расширенном пространстве с помощью гауссовского ядра с $\sigma = 1$.

   (c) Найдите расстояние AB и косинус угла ABC в расширенном пространстве с помощью полиномиального ядра второй степени

2. Переход из двумерного пространства в расширяющее задан функцией

$$f : (x_1, x_2) \to (1, x_1, x_2, 3x_1x2, 2x_1^2, 4x_2^2)$$

   Найдите соответствующую ядерную функцию

3. Ядерная функция имеет вид

$$K(x, y) = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

   Как может выглядеть функция $f : \mathbb{R}^2 \to \mathbb{R}^3$ переводящие исходные векторы в расширенное пространство?

4. Дана плоскость. На ней точки. Симметрично ох. Найдите разделяющую гиперплоскость при разных C.

12 Деревья и Random Forest

1. Для случайных величин X и Y найдите индекс Джини и энтропию

X	0	1
$\mathbb{P}()$	0.2	0.8

,

Y	0	1	5
$\mathbb{P}()$	0.2	0.3	0.5

2. Случайная величина X принимает значение 1 с вероятностью p и значение 0 с вероятностью $1 - p$.

   (a) Постройте график зависимости индекса Джини и энтропии от p

   (b) При каком p энтропия и индекс Джини будут максимальны?

3. табличка с тремя признаками...

   (a) Какой фактор нужно использовать при прогнозировании y, чтобы минимизировать энтропию?

   (b) Какой фактор нужно использовать при прогнозировании y, чтобы минимизировать индекс Джини?

13 Линейная алгебра

1. Найдите каждую из следующих матриц в каждой из следующих степеней $\frac{1}{2}$, $\frac{1}{3}$, $-\frac{1}{2}$, $-\frac{1}{3}$, -1, 100.

   (a) $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$

2. Найдите ортогональную проекцию и ортогональную составляющую (перпендикуляр) вектора u_1 на линейное подпространство $L = \mathcal{L}(u_2)$, порождённое вектором u_2, если

   (a) $u_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}, u_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}$

   (b) $u_1 = \begin{pmatrix} 2 & 2 & 2 & 2 \end{pmatrix}, u_2 = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}$

   (c) $u_1 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}, u_2 = \begin{pmatrix} 7 & 0 & 0 & 7 \end{pmatrix}$

3. Найдите обратные матрицы ко всем матрицам, представленным ниже.

   (a) $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$

   (c) $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

   (d) $\begin{pmatrix} 0 & 0 & a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

4. Найдите ранг следующих матриц в зависимости от значений параметра λ.

   (a) $\begin{pmatrix} \lambda & 1 & 1 \\ 1 & \lambda & 1 \\ 1 & 1 & \lambda \end{pmatrix}$

   (b) $\begin{pmatrix} 1-\lambda & 1-2\lambda \\ 1+\lambda & 1+3\lambda \end{pmatrix}$

   (c) $\begin{pmatrix} 1 & \lambda & -1 & 2 \\ 2 & -1 & \lambda & 5 \\ 1 & 10 & -6 & 1 \end{pmatrix}$

   (d) $\begin{pmatrix} \lambda & 1 & -1 & -1 \\ 1 & \lambda & -1 & -1 \\ 1 & 1 & -\lambda & -1 \\ 1 & 1 & -1 & -\lambda \end{pmatrix}$

5. Пусть $i = (1, \ldots, 1)'$ — вектор из n единиц и $\pi = i(i'i)^{-1}i'$. Найдите:

   (a) $\mathrm{tr}(\pi)$ и $\mathrm{rk}(\pi)$

   (b) $\mathrm{tr}(I - \pi)$ и $\mathrm{rk}(I - \pi)$

6. Пусть X — матрица размера $n \times k$, где $n > k$, и пусть $\mathrm{rk}(X) = k$. Верно ли, что матрица $P = X(X'X)^{-1}X'$ симметрична и идемпотентна?

7. Пусть X — матрица размера $n \times k$, где $n > k$, и пусть $\mathrm{rk}(X) = k$. Верно ли, что каждый столбец матрицы $P = X(X'X)^{-1}X'$ является собственным вектором матрицы P, отвечающим собственному значению 1?

8. Пусть X — матрица размера $n \times k$, где $n > k$, пусть $\mathrm{rk}(X) = k$ и $P = X(X'X)^{-1}X'$. Верно ли, что каждый вектор-столбец u, такой что $X'u = 0$, является собственным вектором матрицы P, отвечающим собственному значению 0?

9. Верно ли, что для любых матриц A размера $m \times n$ и матриц B размера $n \times m$ выполняется равенство $\mathrm{tr}(AB) = \mathrm{tr}(BA)$?
10. Верно ли, что собственные значения симметричной и идемпотентной матрицы могут быть только нулями и единицами?
11. Пусть P — матрица размера $n \times n$, $P' = P$, $P^2 = P$. Верно ли, что $\mathrm{rk}(P) = \mathrm{tr}(P)$?
12. Верно ли, что для симметричной матрицы собственные векторы, отвечающие различным собственным значениям, ортогональны?
13. Найдите собственные значения и собственные векторы матрицы $P = X(X'X)^{-1}X'$, если

    (a) $X = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$

    (b) $X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$

    (c) $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

    (d) $X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$

14. Задача 13.1.
    Приведите пример таких A и B, что $\det(AB) \neq \det(BA)$.
15. Для матриц-проекторов $\pi = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$ и $P = X(X'X)^{-1}X'$ найдите $\mathrm{tr}(\pi)$, $\mathrm{tr}(P)$, $\mathrm{tr}(I - \pi)$, $\mathrm{tr}(I - P)$.
16. Выпишите в явном виде матрицы $X'X$, $(X'X)^{-1}$ и $X'y$, если
    $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ и $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$
17. Выпишите в явном виде матрицы π, πy, $\pi\varepsilon$, $I - \pi$, если $\pi = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$.
18. Формула Фробениуса. Матрицу A размера $(n+m) \times (n+m)$ разрезали на 4 части: $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. Кусок A_{11} имеет размер $n \times n$ и обратим, кусок A_{22} имеет размер $m \times m$. Известно, что A — обратима и $A^{-1} = B$. На аналогичные по размеру и расположению части разрезали матрицу $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$.

    (a) Каковы размеры кусков A_{12} и A_{21}?

    (b) Чему равно $B_{22}(A_{22} - A_{21}A_{11}^{-1}A_{12})$?

19. Спектральное разложение. Симметричная матрица A размера $n \times n$ имеет n собственных чисел $\lambda_1, \ldots, \lambda_n$ с собственными векторами $u_1, \ldots, u_n$. Докажите, что A можно представить в виде $A = \sum \lambda_i u_i u_i'$.
20. Найдите определитель, собственные значения, собственные векторы и число обусловленности матрицы A. Также найдите A^{-1}, $A^{-1/2}$ и $A^{1/2}$.

(a) $A = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix}$

(b) $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

(c) $A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$

(d) $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

(e) $A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$

(f) $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

(g) $A = \begin{pmatrix} 4 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 4 \end{pmatrix}$

(h) $A = \begin{pmatrix} 6 & -2 & 2 \\ -2 & 5 & 0 \\ 2 & 0 & 7 \end{pmatrix}$

14 Случайные векторы

1. Пусть $y = (y_1, y_2, y_3, y_4, y_5)'$ — случайный вектор доходностей пяти ценных бумаг. Известно, что $\mathbb{E}(y') = (5, 10, 20, 30, 40)$, $\operatorname{Var}(y_1) = 0$, $\operatorname{Var}(y_2) = 10$, $\operatorname{Var}(y_3) = 20$, $\operatorname{Var}(y_4) = 40$, $\operatorname{Var}(y_5) = 40$ и

$$\operatorname{Corr}(y) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.3 & -0.2 & 0.1 \\ 0 & 0.3 & 1 & 0.3 & -0.2 \\ 0 & -0.2 & 0.3 & 1 & 0.3 \\ 0 & 0.1 & -0.2 & 0.3 & 1 \end{pmatrix}$$

   С помощью компьютера найдите ответы на вопросы:

   (a) Какая ценная бумага является безрисковой?

   (b) Найдите ковариационную матрицу $\operatorname{Var}(y)$

   (c) Найдите ожидаемую доходность и дисперсию доходности портфеля, доли ценных бумаг в котором равны соответственно:

      i. $\alpha = (0.2, 0.2, 0.2, 0.2, 0.2)'$

      ii. $\alpha = (0.0, 0.1, 0.2, 0.3, 0.4)'$

      iii. $\alpha = (0.0, 0.4, 0.3, 0.2, 0.1)'$

   (d) Составьте из данных бумаг пять некоррелированных портфелей

2. Пусть $i = (1, \ldots, 1)'$ — вектор из n единиц, $\pi = i(i'i)^{-1}i'$ и $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \sim N(0, I)$.

   (a) Найдите $\mathbb{E}(\varepsilon'\pi\varepsilon)$, $\mathbb{E}(\varepsilon'(I - \pi)\varepsilon)$ и $\mathbb{E}(\varepsilon\varepsilon')$

   (b) Как распределены случайные величины $\varepsilon'\pi\varepsilon$ и $\varepsilon'(I - \pi)\varepsilon$?

   (c) Запишите выражения $\varepsilon'\pi\varepsilon$ и $\varepsilon'(I - \pi)\varepsilon$, используя знак суммы

3. Пусть $X=\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$, $P=X(X'X)^{-1}X'$, случайные величины $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$ независимы и одинаково распределены $\sim N(0,1)$.
   (a) Найдите распределение случайной величины $\varepsilon' P\varepsilon$, где $\varepsilon=\begin{pmatrix}\varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4\end{pmatrix}'$
   (b) Найдите $\mathbb{E}(\varepsilon' P\varepsilon)$
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon' P\varepsilon>q)=0.1$
4. Пусть $X=\begin{pmatrix}1&1\\1&2\\1&3\\1&4\end{pmatrix}$, $P=X(X'X)^{-1}X'$, случайные величины $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$ независимы и одинаково распределены $\sim N(0,1)$.
   (a) Найдите распределение случайной величины $\varepsilon' P\varepsilon$, где $\varepsilon=\begin{pmatrix}\varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4\end{pmatrix}'$
   (b) Найдите $\mathbb{E}(\varepsilon' P\varepsilon)$
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon' P\varepsilon>q)=0.1$
5. Пусть $X=\begin{pmatrix}1&0&0\\1&0&0\\1&1&0\\1&1&1\end{pmatrix}$, $P=X(X'X)^{-1}X'$, случайные величины $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$ независимы и одинаково распределены $\sim N(0,1)$.
   (a) Найдите распределение случайной величины $\varepsilon' P\varepsilon$, где $\varepsilon=\begin{pmatrix}\varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4\end{pmatrix}'$.
   (b) Найдите $\mathbb{E}(\varepsilon' P\varepsilon)$.
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon' P\varepsilon>q)=0.1$.
6. Пусть $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, $\mathbb{E}(x)=\begin{pmatrix}1\\2\end{pmatrix}$, $\operatorname{Var}(x)=\begin{pmatrix}2&1\\1&2\end{pmatrix}$. Найдите $\mathbb{E}(y)$, $\operatorname{Var}(y)$ и $\mathbb{E}(z)$, если
   (a) $y=x-\mathbb{E}(x)$
   (b) $y=\operatorname{Var}(x)x$
   (c) $y=\operatorname{Var}(x)(x-\mathbb{E}(x))$
   (d) $y=\operatorname{Var}(x)^{-1}(x-\mathbb{E}(x))$
   (e) $y=\operatorname{Var}(x)^{-1/2}(x-\mathbb{E}(x))$
   (f) $z=(x-\mathbb{E}(x))'\operatorname{Var}(x)(x-\mathbb{E}(x))$
   (g) $z=(x-\mathbb{E}(x))'\operatorname{Var}(x)^{-1}(x-\mathbb{E}(x))$
   (h) $z=x'\operatorname{Var}(x)x$
   (i) $z=x'\operatorname{Var}(x)^{-1}x$
7. Пусть $x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, $\mathbb{E}(x)=\begin{pmatrix}1\\4\end{pmatrix}$, $\operatorname{Var}(x)=\begin{pmatrix}4&1\\1&4\end{pmatrix}$. Найдите $\mathbb{E}(y)$, $\operatorname{Var}(y)$ и $\mathbb{E}(z)$, если
   (a) $y=x-\mathbb{E}(x)$
   (b) $y=\operatorname{Var}(x)x$
   (c) $y=\operatorname{Var}(x)(x-\mathbb{E}(x))$
   (d) $y=\operatorname{Var}(x)^{-1}(x-\mathbb{E}(x))$
   (e) $y=\operatorname{Var}(x)^{-1/2}(x-\mathbb{E}(x))$

(f) $z = (x - \mathbb{E}(x))' \operatorname{Var}(x)(x - \mathbb{E}(x))$

(g) $z = (x - \mathbb{E}(x))' \operatorname{Var}(x)^{-1}(x - \mathbb{E}(x))$

(h) $z = x' \operatorname{Var}(x) x$

(i) $z = x' \operatorname{Var}(x)^{-1} x$

8. Известно, что случайные величины x_1, x_2 и x_3 имеют следующие характеристики:

   (a) $\mathbb{E}(x_1) = 5$, $\mathbb{E}(x_2) = 10$, $\mathbb{E}(x_3) = 8$

   (b) $\operatorname{Var}(x_1) = 6$, $\operatorname{Var}(x_2) = 14$, $\operatorname{Var}(x_3) = 1$

   (c) $\operatorname{Cov}(x_1, x_2) = 3$, $\operatorname{Cov}(x_1, x_3) = 1$, $\operatorname{Cov}(x_2, x_3) = 0$

   Пусть случайные величины y_1, y_2 и y_3, представляют собой линейные комбинации случайных величин X_1, X_2 и X_3:

$$y_1 = x_1 + 3x_2 - 2x_3$$

$$y_2 = 7x_1 - 4x_2 + x_3$$

$$y_3 = -2x_1 - x_2 + 4x_3$$

   (a) Выпишите математическое ожидание и ковариационную матрицу случайного вектора $x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}^T$

   (b) Напишите матрицу A, которая позволяет перейти от случайного вектора $x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}^T$ к случайному вектору $y = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}^T$

   (c) С помощью матрицы A найдите математическое ожидание и ковариационную матрицу случайного вектора $y = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}^T$

9. Пусть ξ_1, ξ_2, ξ_3 — случайные величины, такие что $\operatorname{Var}(\xi_1) = 2$, $\operatorname{Var}(\xi_2) = 3$, $\operatorname{Var}(\xi_3) = 4$, $\operatorname{Cov}(\xi_1, \xi_2) = 1$, $\operatorname{Cov}(\xi_1, \xi_3) = -1$, $\operatorname{Cov}(\xi_2, \xi_3) = 0$. Пусть $\xi = \begin{pmatrix} \xi_1 & \xi_2 & \xi_3 \end{pmatrix}^T$. Найдите $\operatorname{Var}(\xi)$ и $\operatorname{Var}(\xi_1 + \xi_2 + \xi_3)$.

10. Пусть $h = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$; $\mathbb{E}(h) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\operatorname{Var}(h) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$; $z_1 = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$. Найдите $\mathbb{E}(z_1)$ и $\operatorname{Var}(z_1)$.

11. Пусть $h = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$; $\mathbb{E}(h) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\operatorname{Var}(h) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$; $z_2 = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Найдите $\mathbb{E}(z_2)$ и $\operatorname{Var}(z_2)$

12. Пусть $h = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$; $\mathbb{E}(h) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\operatorname{Var}(h) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$; $z_3 = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} - \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix}$. Найдите $\mathbb{E}(z_3)$ и $\operatorname{Var}(z_3)$

13. Пусть r_1, r_2 и r_3 — годовые доходности трёх рисковых финансовых инструментов. Пусть α_1, α_2 и α_3 — доли, с которыми данные инструменты входят в портфель инвестора. Считаем, что $\sum_{i=1}^{3} \alpha_i = 1$ и $\alpha_i \geqslant 0$ для всех $i = 1, 2, 3$. Пусть $r = \begin{pmatrix} r_1 & r_2 & r_3 \end{pmatrix}^T$, $\mathbb{E}(r) = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}^T$, $\operatorname{Var}(r) = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$. Параметры $\{a_i\}$ и $\{c_i\}$ известны.

    (a) Найдите годовую доходность портфеля Π инвестора

    (b) Докажите, что дисперсия доходности портфеля Π равна $\sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i c_{ij} \alpha_j$

    (c) Для случая $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, $\alpha 3 = 0.4$, $\mathbb{E}(r) = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}^T = \begin{pmatrix} 0.10 & 0.06 & 0.05 \end{pmatrix}^T$, $\operatorname{Var}(r) = \begin{pmatrix} 0.04 & 0 & -0.005 \\ 0 & 0.01 & 0 \\ -0.005 & 0 & 0.0025 \end{pmatrix}$ найдите $\mathbb{E}(\Pi)$ и $\operatorname{Var}(\Pi)$

14. Пусть $h = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$; $\mathbb{E}(h) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\mathrm{Var}(h) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$; $z_3 = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} - \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix}$; $z_4 = \mathrm{Var}(h)^{-1/2} z_3$. Найдите $\mathbb{E}(z_4)$ и $\mathrm{Var}(z_4)$
15. Пусть $h = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$; $\mathbb{E}(h) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $\mathrm{Var}(h) = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$; $z_3 = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} - \begin{pmatrix} \mathbb{E}\xi_1 \\ \mathbb{E}\xi_2 \end{pmatrix}$; $z_4 = \mathrm{Var}(h)^{-1/2} z_3$. Найдите $\mathbb{E}(z_4)$ и $\mathrm{Var}(z_4)$
16. Случайные величины w_1 и w_2 независимы с нулевым ожиданием и единичной дисперсией. Из них составлено два вектора, $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ и $z = \begin{pmatrix} -w_2 \\ w_1 \end{pmatrix}$
    (a) Являются ли векторы w и z перпендикулярными?
    (b) Найдите $\mathbb{E}(w)$, $\mathbb{E}(z)$
    (c) Найдите $\mathrm{Var}(w)$, $\mathrm{Var}(z)$, $\mathrm{Cov}(w, z)$
17. Есть случайный вектор $w = (w_1, w_2, \ldots, w_n)'$.
    (a) Возможно ли, что $E(w) = 0$ и $\sum w_i = 0$?
    (b) Возможно ли, что $E(w) \neq 0$ и $\sum w_i = 0$?
    (c) Возможно ли, что $E(w) = 0$ и $\sum w_i \neq 0$?
    (d) Возможно ли, что $E(w) =\neq$ и $\sum w_i \neq 0$?
18. Известна ковариационная матрица вектора $\varepsilon = (\varepsilon_1, \varepsilon_2)$,

$$\mathrm{Var}(\varepsilon) = \begin{pmatrix} 9 & -1 \\ -1 & 9 \end{pmatrix}$$

    Найдите четыре различных матрицы A, таких что вектор $v = A\varepsilon$ имеет некоррелированные компоненты с единичной дисперсией, то есть $\mathrm{Var}(A\varepsilon) = I$.

15 Многомерное нормальное и квадратичные формы

1. Пусть $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, I)$ и матрица A представлена ниже. Найдите $\mathbb{E}(\varepsilon' A \varepsilon)$ и распределение случайной величины $\varepsilon' A \varepsilon$.
    (a) $\begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix}$
    (b) $\begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$
    (c) $\begin{pmatrix} 1/3 & 1/3 & -1/3 \\ 1/3 & 1/3 & -1/3 \\ -1/3 & -1/3 & 1/3 \end{pmatrix}$
    (d) $\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$
    (e) $\begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$

(f) $\begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix}$

(g) $\begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(h) $\begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(i) $\begin{pmatrix} 0.8 & 0.4 & 0 \\ 0.4 & 0.2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(j) $\begin{pmatrix} 0.2 & -0.4 & 0 \\ -0.4 & 0.8 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

2. Пусть $i = (1, \ldots, 1)'$ — вектор из n единиц, $\pi = i(i'i)^{-1}i'$ и $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \sim N(0, I)$.
   (a) Найдите $\mathbb{E}(\varepsilon'\pi\varepsilon)$, $\mathbb{E}(\varepsilon'(I-\pi)\varepsilon)$ и $\mathbb{E}(\varepsilon\varepsilon')$
   (b) Как распределены случайные величины $\varepsilon'\pi\varepsilon$ и $\varepsilon'(I-\pi)\varepsilon$?
   (c) Запишите выражения $\varepsilon'\pi\varepsilon$ и $\varepsilon'(I-\pi)\varepsilon$, используя знак суммы

3. Пусть $X = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$, $P = X(X'X)^{-1}X'$, случайные величины $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ независимы и одинаково распределены $\sim N(0, 1)$.
   (a) Найдите распределение случайной величины $\varepsilon'P\varepsilon$, где $\varepsilon = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 \end{pmatrix}'$
   (b) Найдите $\mathbb{E}(\varepsilon'P\varepsilon)$
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon'P\varepsilon > q) = 0.1$

4. Пусть $X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$, $P = X(X'X)^{-1}X'$, случайные величины $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ независимы и одинаково распределены $\sim N(0, 1)$.
   (a) Найдите распределение случайной величины $\varepsilon'P\varepsilon$, где $\varepsilon = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 \end{pmatrix}'$
   (b) Найдите $\mathbb{E}(\varepsilon'P\varepsilon)$
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon'P\varepsilon > q) = 0.1$

5. Пусть $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $P = X(X'X)^{-1}X'$, случайные величины $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ независимы и одинаково распределены $\sim N(0, 1)$.
   (a) Найдите распределение случайной величины $\varepsilon'P\varepsilon$, где $\varepsilon = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 \end{pmatrix}'$.
   (b) Найдите $\mathbb{E}(\varepsilon'P\varepsilon)$.
   (c) При помощи таблиц найдите такое число q, что $\mathbb{P}(\varepsilon'P\varepsilon > q) = 0.1$.

6. Пусть $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, I)$. Найдите $\mathbb{E}(\varepsilon' P \varepsilon)$ и распределение случайной величины $\varepsilon' P \varepsilon$, если $P = X(X'X)^{-1}X'$ и матрица X' представлена ниже.
   (a) $\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$
   (b) $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$
   (c) $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$
   (d) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$
   (e) $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$
7. Пусть $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \sim N(0, \sigma^2 I)$, $i = (1, \ldots, 1)'$ — вектор из n единиц, $\pi = i(i'i)^{-1}i'$, X — матрица размера $n \times k$, $P = X(X'X)^{-1}X'$. Найдите:
   (a) $\mathbb{E}(\varepsilon'(P - \pi)\varepsilon)$
   (b) $\mathbb{E}(\varepsilon'(I - \pi)\varepsilon)$
   (c) $\mathbb{E}(\varepsilon' P \varepsilon)$
   (d) $\mathbb{E}(\sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2)$
8. Пусть $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, 4I)$, $A = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Найдите:
   (a) $\mathbb{E}(\varepsilon' A \varepsilon)$
   (b) $\mathbb{E}(\varepsilon'(I - A)\varepsilon)$
9. Пусть $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ — случайный вектор, имеющий двумерное нормальное распределение с математическим ожиданием $\mu = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$ и ковариационной матрицей $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.
   (a) Найдите Σ^{-1}
   (b) Найдите $\Sigma^{-1/2}$
   (c) Найдите математическое ожидание и ковариационную матрицу случайного вектора $y = \Sigma^{-1/2} \cdot (x - \mu)$
   (d) Какое распределение имеет вектор y из предыдущего пункта?
   (e) Найдите распределение случайной величины $q = (x - \mu)^T \cdot \Sigma^{-1} \cdot (x - \mu)$
10. Пусть $z = \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}^T \sim N(0, I_{3x3})$, $b = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$,
    $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, $K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$.
    (a) Найдите $\mathbb{E}x$ и $\mathrm{Var}(x)$ случайного вектора $x = A \cdot z + b$
    (b) Найдите распределение случайного вектора x
    (c) Найдите $\mathbb{E}q$ случайной величины $q = z^T \cdot K \cdot z$
    (d) Найдите распределение случайной величины q

11. Известно, что $\varepsilon \sim N(0, I)$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)'$. Матрица $A = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$.

    (a) Найдите $\mathbb{E}(\varepsilon' A \varepsilon)$

    (b) Как распределена случайная величина $\varepsilon' A \varepsilon$?

12. Известно, что $\varepsilon \sim N(0, A)$, $\varepsilon = (\varepsilon_1, \varepsilon_2)'$. Матрица $A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$, матрица $B = \begin{pmatrix} -1 & 3 \\ 2 & 1 \end{pmatrix}$

    (a) Как распределен вектор $h = B\varepsilon$?

    (b) Найдите $A^{-1/2}$

    (c) Как распределен вектор $u = A^{-1/2}\varepsilon$?

16 Задачи по программированию

Все наборы данных доступны по ссылке https://github.com/bdemeshev/em301/wiki/Datasets.

1. Начиная с какого знака в числе $\pi = 3.1415\ldots$ можно обнаружить твой номер телефона? Первый 10 миллионов знаков числа π можно найти на сайте http://code.google.com/p/pc2012-grupo-18-turma-b/downloads/list. Если не хватает, то миллиард знаков, файл размера примерно в 1 гигабайт, доступен по ссылке http://stuff.mit.edu/afs/sipb/contrib/pi/. Настоящие челябинцы рассчитывают π самостоятельно. Краткая история о том, как маньяки считали π до 10 миллиардов знаков и потеряли полгода из-за сбоев компьютерного железа, http://www.numberworld.org/misc_runs/pi-10t/details.html.
2. Отряд Иосифа Флавия из 40 воинов, защищающий город Йодфат, блокирован в пещере превосходящими силам римлян. Чтобы не сдаться врагу, воины стали по кругу и договорились, что сами будут убивать каждого третьего, пока не погибнут все. При этом двое воинов, оставшихся последними в живых, должны были убить друг друга. Хитренький Иосиф Флавий, командующий этим отрядом, хочет определить, где нужно встать ему и его товарищу, чтобы остаться последними. Не для того, чтобы убить друг друга, а чтобы сдать крепость римлянам. Напишите программу, которая для n воинов вставших в круг определяет, какие двое останутся последними, если будут убивать каждого k-го.
3. Напишите программу, которая печатает сама себя.
4. Задача Макар-Лиманова. У торговца 55 пустых стаканчиков, разложенных в несколько стопок. Пока нет покупателей он развлекается: берет верхний стаканчик из каждой стопки и формирует из них новую стопку. Потом снова берет верхний стаканчик из каждой стопки и формирует из них новую стопку и т.д.

    (a) Напишите функцию ‘makar_step‘. На вход функции подаётся вектор количества стаканчиков в каждой стопке до перекладывания. На выходе функция возвращает количества стаканчиков в каждой стопке после одного перекладывания.

    (b) Изначально стаканчики были разложены в две стопки, из 25 и 30 стаканчиков. Как разложатся стаканчики если покупателей не будет достаточно долго?

5. Напишите программу, которая находит сумму элементов побочной диагонали квадратной матрицы.
6. Напишите функцию, которая по матрице X и вектору y для модели $Y = X\beta + \varepsilon$ вычисляет значение статистики Дарбина-Уотсона.
7. Напишите функцию, которая по матрице X и вектору y для модели $Y = X\beta + \varepsilon$ вычисляет оценки дисперсии коэффициентов, скорректированные на гетероскедастичность по формуле Уайта

$$\widehat{\mathrm{Var}}_{White}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{u}_{ij}^2}{RSS_j},$$

где $\hat{u}_{ij}$ — остатки в линейной регрессии фактора x_j на остальные регрессоры, а RSS_j — сумма квадратов остатков в этой регрессионной модели.

8. Напишите функцию, которая по матрице X и вектору y для модели $Y = X\beta + \varepsilon$ вычисляет оценки ковариационной матрицы коэффициентов, скорректированную на гетероскедастичность по формуле Уайта:

$$\widehat{\mathrm{Var}}_{White}(\hat{\beta}_{OLS}) = (X'X)^{-1} \left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 X'_{i\cdot} X_{i\cdot} \right) (X'X)^{-1},$$

где $X_{i\cdot}$ — i-ая строка матрицы X.

9. Напишите программу, которая по заданной матрице регрессоров X возвращает матрицу Z , столбцами которой являются все столбцы матрицы X , «квадраты» столбцов матрицы X, а также перекрестные «произведения» столбцов матрицы X.
10. Напишите функцию, которая по матрице X и вектору y возвращает значение статистики Уайта.
11. Напишите функцию, которая по матрице X, вектору y и уровню значимости реализует тест Уайта.
12. Напишите функцию, которая по матрице X, вектору y и количеству лагов L находит оценку ковариационной матрицу коэффициентов, скорректированную на гетероскедастичность и автокорреляцию по формуле Невье-Веста:

$$\widehat{\mathrm{Var}}_{NW}(\hat{\beta}_{OLS}) = (X'X)^{-1} \hat{S} (X'X)^{-1},$$

где

$$\hat{S} = \sum_{t=1}^{n} \hat{\varepsilon}_t^2 X'_{t\cdot} X_{t\cdot} + \sum_{j=1}^{L} w_j \left(\sum_{t=j+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} (X'_{t\cdot} X_{t-j\cdot} + X'_{t-j\cdot} X_{t\cdot}) \right),$$

где ε_t — остатки в регрессии $y = X\beta + \varepsilon$, а $X_{t\cdot}$ — строка номер t матрицы X. Напишите две версии данной функции, для разных способов рассчета весов w_j:

(a) $w_j = 1 - j/L$

(b)

$$w_j = \begin{cases} 1 - \frac{6j^2}{(1+L)^2} + \frac{6j^3}{(1+L)^3}, \text{ если } j \leqslant (1+L)/2 \\ 2\left(1 - \frac{j}{1+L}\right)^2, \text{ если } j > (1+L)/2 \end{cases}$$

17 Устав проверки гипотез

1. Условия применимости теста
2. Формулировка H_0, H_a и уровня значимости α
3. Формула расчета и наблюдаемое значения статистики, S_{obs}
4. Закон распределения S_{obs} при верной H_0
5. Область в которой H_0 не отвергается
6. Точное P-значение
7. Статистический вывод о том, отвергается ли H_0 или нет.

В качестве статистического вывода допускается только одна из двух фраз:

- Гипотеза H_0 отвергается
- Гипотеза H_0 не отвергается

Остальные фразы считаются неуставными