Econometrics: class notes

https://github.com/bdemeshev/metrics_pro

March 20, 2025

Contents

1 OLS is...

True model. For example, $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$.
Forecasting formula. For example, $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i$.
Ordinary least squares, $\sum (y_i - \hat{y}_i)^2 \to \min$.

1.1 Everyday Maria eats candies and solves econometrics problems. Let x_i be the number of problems solved and y_i be the number of candies eaten.

x_i	y_i
1	1
2	2
2	4

Consider the model $y_i = \beta x_i + u_i$.

a) Find the OLS estimate $\hat{\beta}$ for the toy dataset.
b) Draw original points and estimated regression line on a graph.
c) Derive the general formula for $\hat{\beta}$ in the case of n observations.

Now consider the model $y_i = \beta_1 + \beta_2 x_i + u_i$.

d) Find the OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ for the toy dataset.
e) Draw original points and estimated regression line on a graph.
f) Derive the general formula for $\hat{\beta}$ in the case of n observations.

1.2 Prove the following formulas or provide a counter-example:

a) $\sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \bar{x}) x_i$;
b) $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i (x_i - \bar{x}) y_i$;
c) $\sum_{ij} (x_i - x_j)^2 = 2n \sum_i (x_i - \bar{x})^2$;
d) $\sum_{ij} (x_i - x_j)(y_i - y_j) = 2 \sum_i (x_i - \bar{x})(y_i - \bar{y})$;
e) $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i$;

1.3 Simplify the following expressions:

a) $n\bar{x} - \sum x_i$;
b) $\sum (x_i - \bar{x})\bar{x}$;
c) $\sum (x_i - \bar{x})\bar{z}$;
d) $\sum (x_i - \bar{x})^2 + n\bar{x}^2$.

1.4 Using OLS find the estimator θ in the following models:

a) $y_i = \theta + \theta x_i + \varepsilon_i$;
b) $y_i = 1 + \theta x_i + u_i$;
c) $y_i = \theta / x_i + u_i$;
d) $y_i = \theta x_i + (1 - \theta) z_i + u_i$.

1.5 Find the OLS estimators α and β for the model $y_i = \alpha + \beta y_i + u_i$.

1.6 Рассмотрите модели $y_i = \alpha + \beta(y_i + z_i) + u_i$, $z_i = \gamma + \delta(y_i + z_i) + u_i$.

a) Как связаны между собой $\hat{\alpha}$ и $\hat{\gamma}$?
b) Как связаны между собой $\hat{\beta}$ и $\hat{\delta}$?

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

1.8 Для модели $y_i = \beta_1 x_i + \beta_2 z_i + u_i$ решите условную задачу о наименьших квадратах:

$$Q(\beta_1, \beta_2) := \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i)^2 \to \min_{\hat{\beta}_1 + \hat{\beta}_2 = 1}.$$

1.9 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.

1.10 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.

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

a) Возможно ли, что $\hat{\beta}_2 > 0$, $\hat{\gamma}_2 > 0$, но $\hat{\phi}_2 < 0$?
b) Возможно ли, что $\hat{\beta}_1 > 0$, $\hat{\gamma}_1 > 0$, но $\hat{\phi}_1 < 0$?
c) Возможно ли одновременное выполнение всех упомянутых условий?
d) Возможно ли одновременное выполнение всех упомянутых условий, если в каждой сотне наблюдений $\sum x_i > 0$?

1.12 Эконометрист Вовочка собрал интересный набор данных по студентам третьего курса:

- переменная y_i — количество пирожков, съеденных i-ым студентом за прошлый год;
- переменная f_i, которая равна 1, если i-ый человек в выборке — женщина, и 0, если мужчина.
- переменная[1] m_i, которая равна 1, если i-ый человек в выборке — мужчина, и 0, если женщина.

Вовочка оценил 4 регрессии:

A: y на константу и f, $\hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 f_i$;
B: y на константу и m, $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 m_i$;
C: y на f и m без константы, $\hat{y}_i = \hat{\gamma}_1 f_i + \hat{\gamma}_2 m_i$;
D: y на константу, f и m, $\hat{y}_i = \hat{\delta}_1 + \hat{\delta}_2 f_i + \hat{\delta}_3 m_i$;

a) Какой смысл будут иметь оцениваемые коэффициенты?
b) Как связаны между собой оценки коэффициентов этих регрессий?

1.13 Эконометрист Вовочка оценил методом наименьших квадратов модель 1, $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$, а затем модель 2, $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + u$. Сравните полученные ESS, $\mathrm{SS^{res}}$, TSS и R^2.

1.14 Что происходит с TSS, $\mathrm{SS^{res}}$, ESS, R^2 при добавлении нового наблюдения? Если величина может изменяться только в одну сторону, то докажите это. Если возможны и рост, и падение, то приведите пример.

1.15 Эконометресса Аглая подглядела, что у эконометрессы Жозефины получился R^2 равный 0.99 по 300 наблюдениям. От чёрной зависти Аглая не может ни есть, ни спать.

a) Аглая добавила в набор данных Жозефины ещё 300 наблюдений с такими же регрессорами, но противоположными по знаку игреками, чем были у Жозефины. Как изменится R^2?

[1] Это нетолерантная задача и здесь либо f равно 1, либо m

b) Жозефина заметила, что Аглая добавила 300 наблюдений и вычеркнула их, вернув набор данных в исходное состояние. Хитрая Аглая решила тогда добавить всего одно наблюдение так, чтобы R^2 упал до нуля. Удастся ли ей это сделать?

1.16 Foma estimated simple regression model using three observations and calculated the forecasts $\hat{y}_i$. Unfortunately he remembers only this table:

y_i	$\hat{y}_i$
0	1
6	?
6	?

He recalls that $\hat{y}_3 > \hat{y}_2$.

Help Foma to reconstruct the missing entries in the table.

1.17 Вся выборка поделена на две части. Возможны ли такие ситуации:

a) Выборочная корреляция между y и x примерно равна нулю в каждой части, а по всей выборке примерно равна единице;
b) Выборочная корреляция между y и x примерно равна единице в каждой части, а по всей выборке примерно равна нулю?

1.18 Бесстрашный исследователь Ипполит оценил парную регрессию. При этом оказалось, каждый $x_i > 0$ и обе оценки коэффициентов $\hat{\beta}_1$ и $\hat{\beta}_2$ также положительны.

a) Возможно ли, что среди $\hat{y}_i$ есть отрицательные? Среди y_i есть отрицательные?
b) Возможно ли, что сумма $\sum \hat{y}_i$ отрицательна? Сумма $\sum y_i$ отрицательна?
c) Как изменятся ответы, если известно, что $\sum x_i > 0$?

1.19 Propose a way to define a correlation

a) between binary and factor variables;
b) between quantitative and factor variables.

1.20 Начинающий художник Франческо нарисовал облако из 1000 точек в осях (x, y). Разные наблюдения независимы, для каждой пары точек $\text{Corr}(x_i, y_i) = 0.5$. По отдельности величины x_i и y_i имеют стандартное нормальное распределение.

a) Правда ли, что облако точек примерно симметрично относительно оси $x = y$?
b) Воспроизведите примерно шедевр начинающего мастера.
c) Добавьте к облаку точек линии парных регрессий y от x и x от y.

1.21 Подающий надежды молодой художник Франческо увлекся минимализмом. Он нарисовал три точки на плоскости (x, y) симметрично линии $x = y$ так, что выборочные средние $\bar{x} = \bar{y} = 0$, выборочные дисперсии равны 1. А выборочная корреляция равна 0.5.

a) Воспроизведите рисунок начинающего маэстро.
b) Добавьте на рисунок линии парных регрессий y от x и x от y.

1.22 Усердный муравей Виталий вместо того, чтобы построить одну регрессию по n точкам, построил все возможные парные регрессии для каждой пары точек (i, j). При этом Виталий получил множество оценок коэффициентов $\hat{\beta}_{1ij}$ и $\hat{\beta}_{2ij}$.

Сами наблюдения не сохранились, однако для каждой пары точек помимо двух оценок также сохранился квадрат расстояния между ними по горизонтали $q_{ij} = (x_i - x_j)^2$.

a) Помогите Виталию восстановить классическую мнк-оценку $\hat{\beta}_2$.
b) Помогите Виталию восстановить классическую мнк-оценку $\hat{\beta}_1$.

2 Fun with differential

- $d(A+B) = dA + dB$;
- If A is a constant matrix, then $dA = 0$;
- $d(AB) = dA \cdot B + A \cdot dB$; If A is a constant matrix, then $d(AB) = A \cdot dB$ and $d(BA) = dB \cdot A$;

Transpose of a matrix is often denoted by prime, $A' = A^T$.

2.1 Вспомним дифференциал :)

a) Известно, что $f(x) = x^2 + 3x$. Найдите $f'(x)$ и df. Чему равен df в точке $x = 5$ при $dx = 0.1$?

b) Известно, что $f(x_1, x_2) = x_1^2 + 3x_1x_2^3$. Найдите df. Чему равен df в точке $x_1 = -2$, $x_2 = 1$ при $dx_1 = 0.1$ и $dx_2 = -0.1$?

c) Известно, что $F = \begin{pmatrix} 5 & 6x_1 \\ x_1x_2 & x_1^2x_2 \end{pmatrix}$. Найдите dF.

d) Известно, что $F = \begin{pmatrix} 7 & 8 & 9 \\ 2 & -1 & -2 \end{pmatrix}$. Найдите dF.

e) Матрица F имеет размер 2×2, в строке i столбце j у неё находится элемент f_{ij}. Выпишите выражение $\operatorname{trace}(F'dF)$ в явном виде без матриц.

2.2 Let A and B be constant matrices, R be a matrix of variables and r be a column vector of variables.

a) Find $d(ARB)$;

b) Find $d(r'r)$;

c) Find $d(r'Ar)$. Simplify the answer for symmetric matrix A.

d) Find $d(R^{-1})$. Hint: $R^{-1} \cdot R = I$;

e) Find $d\cos(r'r)$;

f) Find $d(r'Ar/r'r)$. Simplify the answer for symmetric matrix A.

2.3 В методе наименьших квадратов минимизируется функция

$$Q(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}).$$

a) Найдите $dQ(\hat{\beta})$ и $d^2Q(\hat{\beta})$;

b) Выпишите условия первого порядка для задачи МНК;

c) Выразите $\hat{\beta}$ предполагая, что $X'X$ обратима.

2.4 В гребневой регрессии (ridge regression) минимизируется функция

$$Q(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) + \lambda\hat{\beta}'\hat{\beta},$$

где λ — положительный параметр, штрафующий функцию за слишком большие значения $\hat{\beta}$.

a) Найдите $dQ(\hat{\beta})$ и $d^2Q(\hat{\beta})$;

b) Выпишите условия первого порядка для задачи гребневой регрессии;

c) Выразите $\hat{\beta}$.

2.5 Исследователь Никодим поймал 100 морских ежей и у каждого измерил длину, a_i, и вес b_i. Вектор измерений, относящихся к одному ежу обозначим $y_i = \begin{pmatrix} a_i \\ b_i \end{pmatrix}$. Никодим считает, что ежи независимы друг от друга, а длина и вес имеют совместное нормальное распределение

$$y_i = \begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim \mathcal{N}(\mu, C)$$

a) Выпишите логарифмическую функцию правдоподобия, $\ell(\mu, C)$;

b) Предполагая ковариационную матрицу известной, $C = \begin{pmatrix} 9 & 4 \\ 4 & 6 \end{pmatrix}$, найдите $d\ell$ и оценку $\hat{\mu}$ методом максимального правдоподобия.

c) Предполагая, вектор ожиданий известным, $\mu = \begin{pmatrix} 10 \\ 5 \end{pmatrix}$, найдите $d\ell$ и оценку $\hat{C}$ методом максимального правдоподобия.

d) Найдите $d\ell(\mu, C)$ и оценки для параметров μ и C, в случае, когда μ и C неизвестны.

3 OLS in matrix form and geometry

3.1 Рассмотрим регрессию $\hat{y}_i = \hat{\beta}_1 z_i + \hat{\beta}_2 x_i$. Все исходные данные поместим в матрицу X и вектор y:

$$X = \begin{pmatrix} z_1 & x_1 \\ \vdots & \vdots \\ z_n & x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

a) Выпишите явно матрицы X', $X'y$, $X'X$, $y'X$, $y'z$ и укажите их размер.

b) Выпишите условия первого порядка для оценок $\hat{\beta}_1$ и $\hat{\beta}_2$ по методу наименьших квадратов.

c) Запишите эти же условия в виде линейной системы

$$\begin{cases} \hat{\beta}_1 \cdot \ldots + \hat{\beta}_2 \cdot \ldots = \ldots \\ \hat{\beta}_1 \cdot \ldots + \hat{\beta}_2 \cdot \ldots = \ldots \end{cases}$$

d) Как упростится данная система для регресии $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$?

e) Запишите систему условий первого порядка с помощью матрицы X и вектора y;

3.2 Consider the model $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$, where

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix}$$

Random errors u_i are independen identically distributed with $\mathbb{E}(u \mid X) = 0$, $\operatorname{Var}(u \mid X) = \sigma^2 I$.

For simplicity $X'X$ and $(X'X)^{-1}$ matrices are provided:

$$X'X = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, (X'X)^{-1} = \begin{pmatrix} 1/3 & -1/3 & 0 \\ -1/3 & 4/3 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

a) Find the number of observations.

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

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

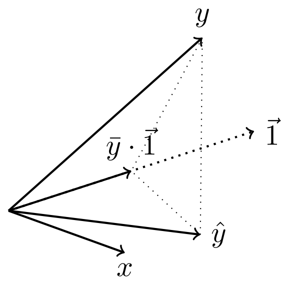
d) Методом МНК найдите оценку для вектора неизвестных коэффициентов.

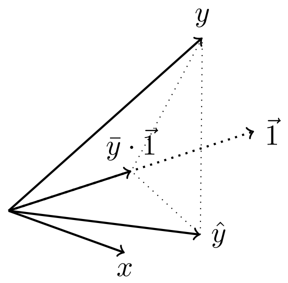
e) Найдите вектор прогнозов $\hat{y}$.

f) Найдите $\mathrm{SS}^{\mathrm{res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

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

3.3 Find all orthogonal vectors. Find all right triangles. State Pythagorean theorem for all right triangles.



3.4 

3.5 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 $\hat{u}$.

3.6 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?

3.7 Under which conditions $TSS = ESS + \mathrm{SS}^{\mathrm{res}}$?

3.8 Вася построил парную регрессию y на x и получил коэффициент наклона 1.4. Построил парную регрессию x на y и получил коэффициент наклона 0.6. Известно, что $y = x + z$.

a) Найдите выборочные корреляции между x и y, y и z, x и z;

b) В какой пропорции соотносятся выборочные дисперсии x, y и z?

3.9 Which matrices are positive definite?

a) $X'X$;

b) XX';

c) $H = X(X'X)^{-1}X'$;

d) $I - H$;

e) $A'(I - H)A$;

f) $A'A - G(G'A^{-1}(A')^{-1}G)^{-1}G'$

3.10 We have n observations. For each case find the hat-matrix H that projects vectors onto the column space of regressor matrix X, vector $\hat{y}$, $\mathrm{SS}^{\mathrm{res}}$, ESS, TSS:

a) $\hat{y}_i = \hat{\beta}$, $n = 10$;

b) $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$, $n = 3$, $x_i = i$, $y_i = i^2$;

c) $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$, $n = 15$, $y_i = i$, first five x_i are zero, other are equal to one;

d) $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$, $n = 15$, $y_i = i$, first five x_i are one, other are equal to zero;

4 Projection and distribution laws

4.1 Рассмотрим пространство $\mathbb{R}^3$ и два подпространства в нём, $W = \{(x_1, x_2, x_3) \mid x_1 + 2x_2 + x_3 = 0\}$ и $V = \mathrm{Lin}((1, 2, 3)^T)$.

a) Найдите $\dim V$, $\dim W$, $\dim V \cap W$, $\dim V^{\perp}$, $\dim W^{\perp}$.

b) Найдите проекцию произвольного вектора u на V, W, $V \cap W$, $V^{\perp}$, $W^{\perp}$. Найдите квадрат длины каждой проекции.

c) Как распределён квадрат длины проекции в каждом случае, если дополнительно известно, что вектор u имеет многомерное стандартное нормальное распределение?

4.2 Рассмотрим пространство $\mathbb{R}^n$, где $n > 2$, и два подпространства в нём, $V = \mathrm{Lin}((1, 1, \ldots, 1)^T)$ и $W = \{x \mid x_1 = x_2 + x_3 + \ldots + x_n\}$.

a) Найдите $\dim V$, $\dim W$, $\dim V \cap W$, $\dim V^{\perp}$, $\dim W^{\perp}$, $\dim V \cap W^{\perp}$, $\dim V^{\perp} \cap W$.

b) Найдите проекцию произвольного вектора u на каждое упомянутое подпространство. Найдите квадрат длины каждой проекции.

c) Как распределён квадрат длины проекции в каждом случае, если дополнительно известно, что вектор u имеет многомерное стандартное нормальное распределение?

4.3 Храбрая исследовательница Евлампия оценивает модель множественной регрессии $\hat{y} = X\hat{\beta}$. Однако на самом деле β_0, и $y = u$, где u_i независимы и нормальны $u_i \sim \mathcal{N}(0; \sigma^2)$.

Какое распределение в регрессии Евлампии имеют $\bar{y}$, $\sum y_i^2/\sigma^2$, $\sum \hat{y}_i^2/\sigma^2$, $n\bar{y}^2/\sigma^2$, TSS/σ^2, $\mathrm{SS}^{\mathrm{res}}/\sigma^2$, ESS/σ^2?

4.4 Компоненты вектора $x = (x_1, x_2)'$ независимы и имеют стандартное нормальное распределение. Вектор y задан формулой $y = (2x_1 + x_2 + 2, x_1 - x_2 - 1)$.

a) Выпишите совместную функцию плотности вектора x;

b) Нарисуйте на плоскости линии уровня функции плотности вектора x;

c) Выпишите совместную функцию плотности вектора y;

d) Найдите собственные векторы и собственные числа ковариационной матрицы вектора y;

e) Нарисуйте на плоскости линии уровня функции плотности вектора y.

4.5 Компоненты вектора $x = (x_1, x_2, x_3)'$ независимы и имеют стандартное нормальное распределение.

a) Как выглядят в пространстве поверхности уровня совместной функции плотности?

b) Рассмотрим три апельсина с кожурой одинаковой очень маленькой толщины: бэби-апельсин радиуса 0.1, стандартный апельсин радиуса 1 и гранд-апельсин радиуса 10. В кожуру какого апельсина вектор x попадает с наибольшей вероятностью?

c) Мы проецируем случайный вектор на x на плоскость $2x_1 + 3x_2 - 7x_3 = 0$. Какое распределение имеет квадрат длины проекции?

d) Введём вектор y независимый от x и имеющий такое же распределение. Спроецируем вектор x на плоскость проходящую через начало координат и перпендикулярную вектору y. Какое распределение имеет квадрат длины проекции?

5 OLS and statistical assumptions

5.1 Исследовательница Мишель собрала данные по 20 студентам. Переменная y_i — количество решённых задач по эконометрике i-ым студентом, а x_i — количество просмотренных серий любимого сериала за прошедший год. Оказалось, что $\sum y_i = 10$, $\sum x_i = 0$, $\sum x_i^2 = 40$, $\sum y_i^2 = 50$, $\sum x_i y_i = 60$.

a) Найдите МНК оценки коэффициентов парной регрессии.

b) В рамках предположения $\mathbb{E}(u_i \mid X) = 0$ найдите $\mathbb{E}(y_i \mid X)$, $\mathbb{E}(\hat{\beta}_j \mid X)$, $\mathbb{E}(\hat{u}_i \mid X)$, $\mathbb{E}(\hat{y}_i \mid X)$.

c) Предположим дополнительно, что $\mathrm{Var}(u_i \mid X) = \sigma^2$ и u_i при фиксированных X независимы. Найдите $\mathrm{Var}(y_i \mid X)$, $\mathrm{Var}(y_i(x_i - \bar{x}) \mid X)$, $\mathrm{Var}(\sum y_i(x_i - \bar{x}) \mid X)$, $\mathrm{Var}(\hat{\beta}_2 \mid X)$.

5.2 Consider classic regression model $y = X\beta + u$ with Gauss — Markov assumptions: $\mathbb{E}(u \mid X) = 0$ and $\mathrm{Var}(u \mid X) = \sigma^2 I$. For all random vectors $(y, \hat{y}, \hat{\beta}, u, \hat{u}, \bar{y})$ find all possible expected values and covariance matrices $\mathbb{E}(\cdot)$, $\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot, \cdot)$.

5.3 Consider the model $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$, where

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix}.$$

Random errors u_i are independen identically distributed with $\mathbb{E}(u \mid X) = 0$ and $\mathrm{Var}(u \mid X) = \sigma^2 I$.

For simplicity $X'X$ and $(X'X)^{-1}$ matrices are provided:

$$X'X = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1.5 \end{pmatrix}.$$

a) Find $\mathbb{E}(\hat{\sigma}^2 \mid X)$, $\hat{\sigma}^2$.

b) Find $\mathrm{Var}(u_1)$, $\mathrm{Var}(\beta_1)$, $\mathrm{Var}(\hat{\beta}_1 \mid X)$, $\widehat{\mathrm{Var}}(\hat{\beta}_1 \mid X)$, $\mathbb{E}(\hat{\beta}_1^2 \mid X) - \beta_1^2$;

c) Find $\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\widehat{\mathrm{Cov}}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\mathrm{Var}(\hat{\beta}_2 - \hat{\beta}_3 \mid X)$, $\widehat{\mathrm{Var}}(\hat{\beta}_2 - \hat{\beta}_3 \mid X)$;

d) Find $\mathrm{Var}(\beta_2 - \beta_3)$, $\mathrm{Corr}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\widehat{\mathrm{Corr}}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$;

Let's additionally suppose that random errors are normally distributed conditionally on X.

e) Construct 95% confidence interval for β_2.

f) Test the null-hypothesis H_0: $\beta_2 = 0$ against H_a: $\beta_2 \neq 0$;

g) Test the null-hypothesis H_0: $\beta_2 = \beta_3$ against H_a: $\beta_2 \neq \beta_3$;

5.4 Начинающий социолог Аполлон опросил 40 человек. В результате у него получилась одна количественная зависимая переменная и 15 регрессоров. Аполлон говорит «хочу построить модель».

Что можно посоветовать Аполлону?

5.5 Consider the model $y_i = \beta x_i + u_i$ with two observations, $x_1 = 1$, $x_2 = 2$. Random errors u_1 and u_2 are independent with $\mathbb{P}(u_i = 1 \mid x) = \mathbb{P}(u_i = -1 \mid x) = 1/2$.

a) Find the estimator $\hat{\beta}_{\text{ols}}$ for β using least squares.

b) Find the variance $\mathrm{Var}(\hat{\beta}_{\text{ols}} \mid x)$ and the expected value $\mathbb{E}(\hat{\beta}_{\text{ols}} \mid x)$.

c) Construct an unbiased estimator $\hat{\beta}_{\text{best}|x}$ with lowest possible variance.

d) What is the variance $\mathrm{Var}(\hat{\beta}_{\text{best}} \mid x)$?

e) What about Gauss — Markov theorem? Why is it possible to create an unbiased estimator with lower variance compared to OLS estimator?

5.6 Troye and Charli use different approaches to test the hypothesis H_0 that $\beta_a = 0$ and $\beta_b = 0$ in the model $y_i = \beta_1 + \beta_a a_i + \beta_b b_i + u_i$ against H_1 that at least one of β_a or β_b is not equal to zero.

They estimated three regressions:

$$(A): \hat{y}_i = \hat{\beta}_1, \quad (B): \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_a a_i, \quad (C): \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_a a_i + \hat{\beta}_b b_i.$$

Troye calculates statistic

$$F_T = \frac{(\mathrm{SS}_A^{\mathrm{res}} - \mathrm{SS}_C^{\mathrm{res}})/2}{\mathrm{SS}_C^{\mathrm{res}}/(n-3)}$$

and compares it with 5% right quantile of the $F_{2,n-3}$ distribution. If statistic is higher than the critical value she rejects H_0.

Charli calculates statistic

$$F_C = \frac{\mathrm{SS}_A^{\mathrm{res}} - \mathrm{SS}_B^{\mathrm{res}}}{\mathrm{SS}_B^{\mathrm{res}}/(n-2)}$$

and compares it with 5% right quantile of the $F_{2,n-3}$ distribution. If statistic is higher than the critical value she rejects H_0.

a) Find the probability of the first type error for Troye.

b) Find the probability of the first type error for Charli.

Assume that the alternative is true, $y_i = 2 + b_i + u_i$, there are $n = 100$ observations. Regressors a_i, b_i and random errors u_i are independent and have standard normal distribution.

c) Estimate the probability of the second type error for Troye, using $B = 10000$ simulations.

d) Estimate the probability of the second type error for Charli, using $B = 10000$ simulations.

Assume that the alternative is true, $y_i = 2 + a_i + u_i$, there are $n = 100$ observations. Regressors a_i, b_i and random errors u_i are independent and have standard normal distribution.

e) Estimate the probability of the second type error for Troye, using $B = 10000$ simulations.

f) Estimate the probability of the second type error for Charli, using $B = 10000$ simulations.

g) Which approach, Troye or Charli, is better and why?

5.7 Consider the model $y_i = \beta_1 + \beta_x x_i + u_i$. Lefty, Middly and Righty test the same H_0: $\beta_1 = 0$ against different alternatives, H_1: $\beta_x < 0$, H_1: $\beta_x \neq 0$ and H_1: $\beta_x > 0$. All three of them use the significance level $\alpha = 0.05$.

a) Find the probability of the first type error for Lefty, Middly and Righty.

Assume that alternative hypothesis is true, $y_i = 2 + x_i + u_i$, there are $n = 100$ observations. Regressor x_i and random errors u_i are independent and have standard normal distribution.

b) Estimate the probability of the second type error for Lefty, Middly and Righty using $B = 10000$ simulations.

5.8 Consider the simple regression model $y_i = \beta_1 + \beta_2 x_i + u_i$, $\mathbb{E}(u \mid X) = 0$, $\mathrm{Var}(u \mid X) = \sigma^2 I$, observations are independent and identically distributed.

The number of observations is even. Consider the estimator

$$\hat{\beta}_2^A = \frac{2}{n}\left(\frac{y_2 - y_1}{x_2 - x_1} + \frac{y_4 - y_3}{x_4 - x_3} + \cdots + \frac{y_n - y_{n-1}}{x_n - x_{n-1}}\right).$$

a) Is $\hat{\beta}_2^A$ unbiased for β_2? Consistent?

Now consider the estimator

$$\hat{\beta}_2^B = \frac{1}{n-1}\left(\frac{y_2 - y_1}{x_2 - x_1} + \frac{y_3 - y_1}{x_3 - x_1} + \cdots + \frac{y_n - y_1}{x_n - x_1}\right).$$

b) Is $\hat{\beta}_2^B$ unbiased for β_2? Consistent?

5.9 Sheep Dolly had a dataset of n observations with Gauss — Markov assumptions satisfied. She cloned every observation once and added each cloned observation after original observation.

a) Describe the error covariance matrix of the augmented dataset.

b) How the answer in (a) will change if Dolly will clone only the last observation n times?

6 F-test

6.1 Consider the standard F-test

$$F = \frac{(SSRes_R - SSRes_{UR})/q}{SSRes_{UR}/(n-k)}$$

for the unrestricted model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{k-1} x_{i,k-1} + u_i.$$

Consider the null-hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_{k-1} = 0$.

Express the F-statistic as a function of R^2_{UR} in the unrestricted regression.

6.2 Consider the standard F-test

$$F = \frac{(SSRes_R - SSRes_{UR})/q}{SSRes_{UR}/(n-k)}$$

for the unrestricted model

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Consider the null-hypothesis H_0: $\beta_1 = 0$.

Express the F-statistic as a function of the corresponding t-statistic in the unrestrited regression.

6.3

6.4

7 Gamma and beta distributions

7.1 Вася делает эксперименты без устали со скоростью d экспериментов в минуту. Каждый эксперимент независимо от других может окончится успехом с вероятностью p или неудачей.

Пусть X — количество успехов за первую минуту, а Y — номер опыта, в котором произошёл первый успех, Z — время, когда случился первый успех.

a) Найдите $\mathbb{P}(X = k)$, $\mathbb{E}(X)$, $\operatorname{Var}(X)$. Как называется закон распределения X?

b) Найдите $\mathbb{P}(Y = k)$, $\mathbb{E}(Y)$, $\operatorname{Var}(Y)$. Как называется закон распределения Y?

c) Найдите $\mathbb{P}(Z \leqslant t)$, $\mathbb{E}(Z)$, $\operatorname{Var}(Z)$.

Теперь Вася ускоряется и устремляет d в бесконечность. Из-за того, что он торопится, p начинает стремится к нулю :) Причем ожидаемое количество успехов за минуту оказывается постоянно и равно λ.

d) Выразите p через λ и d.

e) Найдите предел $\mathbb{P}(Z \leqslant t)$. Является ли предельная функция $\mathbb{P}(Z \leqslant t)$ непрерывной? Какая в предельном случае получается функция плотности у величины Z? Как называется этот закон распределения Z? Чему равен предел $\mathbb{E}(Z)$ и $\mathrm{Var}(Z)$?

f) Найдите предел вероятности $\mathbb{P}(X = k)$ и пределы $\mathbb{E}(X)$ и $\mathrm{Var}(X)$. Как называется предельный закон распределения X?

7.2 Энтомолог Джон Поллак ловит бабочек. На поимку i-ой бабочки у него уходит Y_i минут, величины Y_i независимы. Каждая Y_i имеет экспоненциальное распределение с интенсивностью λ бабочек в минуту. Всего он решил поймать n бабочек. Рассмотрим величины $S = Y_1 + \ldots + Y_n$, $X_1 = Y_1/S$, $X_2 = Y_2/S$, ..., $X_{n-1} = Y_{n-1}/S$.

a) Выпишите совместную функцию плотности Y_1, ..., Y_n;

b) Найдите совместную функцию плотности X_1, X_2, X_3, ..., X_{n-1}, S.

c) Зависит ли величина S и вектор X_1, X_2, ..., X_{n-1}?

d) С точностью до сомножителя выпишите функцию плотности S. Как называется закон распределения S?

e) С точностью до сомножителя выпишите совместную функцию плотности для X_1, ..., X_{n-1}.

Рассмотрим также величины $Z_1 = Y_1/(Y_1 + Y_2)$, $Z_2 = (Y_1 + Y_2)/(Y_1 + Y_2 + Y_3)$, ..., $Z_{n-1} = (Y_1 + \ldots + Y_{n-1})/(Y_1 + \ldots + Y_n)$.

f) Найдите совместную функцию плотности Z_1, Z_2, ..., Z_{n-1}, S.

g) Зависимы ли величины Z_1, Z_2, ..., Z_{n-1}, S?

h) С точностью до константы найдите частную функцию плотности S и каждого Z_i в отдельности;

7.3 Быстрый исследователь Вася снова проводит независимые идентичные опыты с очень высокой скоростью. В среднем λ опытов в минуту оказываются успешными. Поэтому время до очередного успеха можно считать экспоненциально распределённым, а время от начала до k-го успеха — имеющим гамма-распределение $Gamma(k, \lambda)$. На этот раз Вася решил дождаться k_1 успеха, затем быстренько пообедать, а затем дождаться ещё k_2 успехов. Пусть X_1 — время от начала наблюдения до обеда, а X_2 — время от обеда до конца опытов. Также введём $S = X_1 + X_2$ и $Z = X_1/S$ — долю времени до обеда от общего времени набопытов.

a) Найдите совместную функцию плотности S и Z с точностью до константы.

b) Являются ли S и Z независимыми случайными величинами?

c) Найдите частные функции плотности S и Z.

d) Как называется закон распределения S?

e) Как называется закон распределения Z?

f) Какой закон распределения имеет величина $W = 1 - Z$?

7.4 Вася оценивает регрессию y на регрессоры X, включающие константу, а на самом деле все коэффициенты β_j кроме константы равны нулю. Ошибки u_i распределены нормально $\mathcal{N}(0; \sigma^2)$. Какое распределение имеет R^2?

8 Block matrices

8.1 Найдите матрицу M и укажите размеры всех блоков

a) Блок C имеет размер $p \times p$, блок F — размер $q \times q$

$$M = \begin{pmatrix} A & B \end{pmatrix} \cdot \begin{pmatrix} C & D \\ E & F \end{pmatrix}$$

b) Блок C имеет размер $p \times p$, блок F — размер $q \times q$

$$M = \begin{pmatrix} C & D \\ E & F \end{pmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix}$$

c) Блок C имеет размер $p \times p$, блок B — размер $q \times q$

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^T$$

8.2 Найдите обратную матрицу M^{-1} для каждого из случаев

a) Блоки $A_{p\times p}$ и $B_{q\times q}$ обратимы,

$$M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

b) Блоки $A_{p\times p}$ и $B_{q\times q}$ обратимы,

$$M = \begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}$$

c) Блоки $A_{p\times p}$ и $B_{q\times q}$ обратимы,

$$M = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

d) Блоки $A_{p\times p}$ и $B_{q\times q}$ обратимы,

$$M = \begin{pmatrix} A & 0 \\ C & B \end{pmatrix}$$

8.3 Блоки $A_{p\times p}$ и $B_{q\times q}$ обратимы, матрица M имеет вид

$$M = \begin{pmatrix} A & C \\ D & B \end{pmatrix}$$

Рассмотрим обратную матрицу M^{-1}

$$M^{-1} = \begin{pmatrix} X & Z \\ Y & W \end{pmatrix}$$

a) Найдите блок X с помощью процедуры Гаусса;

b) Найдите блок X решив систему двух уравнений на блоки X и Y;

c) Докажите тождество Вудберри

$$(A - CB^{-1}D)^{-1} = A^{-1} + A^{-1}C(B - DA^{-1}C)^{-1}DA^{-1}$$

8.4 Матрица регрессоров X состоит из двух блоков, $X = \begin{pmatrix} L & R \end{pmatrix}$.

a) Из каких блоков состоит матрица $X'X$?

b) Из каких блоков состоит матрица $(X'X)^{-1}$?

9 Regularization

For standardized regressors X and centered y ridge regression minimizes the loss function

$$\mathrm{loss}(\hat\beta) = (y - \hat y)^T(y - \hat y) + \lambda \hat\beta^T \hat\beta, \quad \hat y = X\hat\beta.$$

The matrix $(X^TX)^{-1}$ can be written as

$$(X^TX)^{-1} = \begin{pmatrix} 1/\mathrm{SS}_1^{\mathrm{res}} & -\hat\beta_{12}/\mathrm{SS}_1^{\mathrm{res}} & -\hat\beta_{13}/\mathrm{SS}_1^{\mathrm{res}} \\ -\hat\beta_{21}/\mathrm{SS}_2^{\mathrm{res}} & 1/\mathrm{SS}_2^{\mathrm{res}} & -\hat\beta_{23}/\mathrm{SS}_2^{\mathrm{res}} \\ -\hat\beta_{31}/\mathrm{SS}_3^{\mathrm{res}} & -\hat\beta_{32}/\mathrm{SS}_3^{\mathrm{res}} & 1/\mathrm{SS}_3^{\mathrm{res}} \end{pmatrix},$$

where $\mathrm{SS}_1^{\mathrm{res}}$ is the residual sum of squares in the regression of x_1 onto all other regressors:

$$\hat{x}_1 = \hat{\beta}_{12}x_2 + \hat{\beta}_{13}x_3.$$

9.1 All regressors X are standardized, dependent variable y is centered.

In the ridge regression one minimizes the goal function

$$\mathrm{loss}(\hat{\beta}) = (y - \hat{y})^T(y - \hat{y}) + \lambda\hat{\beta}^T\hat{\beta}, \quad \hat{y} = X\hat{\beta}.$$

a) Find the optimal $\hat{\beta}$ for fixed λ.
b) What happens to the estimates when $\lambda \to \infty$?
c) What happens to the estimates when $\lambda \to 0$?

True model is $y = X\beta + u$ with $\mathbb{E}(u \mid X) = 0$, $\mathrm{Var}(u \mid X) = \sigma^2 I$.

d) Find $\mathbb{E}(\hat{\beta} \mid X)$.
e) Find $\mathrm{Var}(\hat{\beta} \mid X)$.
f) What happens to the $\mathbb{E}(\hat{\beta} \mid X)$ and $\mathrm{Var}(\hat{\beta} \mid X)$ when $\lambda \to \infty$?
g) What happens to the $\mathbb{E}(\hat{\beta} \mid X)$ and $\mathrm{Var}(\hat{\beta} \mid X)$ when $\lambda \to 0$?

9.2 All regressors X are standardized, dependent variable y is centered.

Bob minimizes the goal function of the ridge regression:

$$\mathrm{loss}(\hat{\beta}) = (y - \hat{y})^T(y - \hat{y}) + \lambda\hat{\beta}^T\hat{\beta}, \quad \hat{y} = X\hat{\beta}.$$

Alice adds some artificial observations to the original dataset and than estimates a linear regresion on the augmented dataset y^a, X^a.

Which estimates should she add to obtain the same estimates $\hat{\beta}$ as Bob?

9.3 Matrix X has 100 rows and 4 columns: column of ones, columns a, b and c.

$$1000 \cdot (X^TX)^{-1} = \begin{pmatrix} 58.2 & 1.6 & 0.7 & 15.2 \\ & 16.5 & -7.7 & -0.7 \\ & & 23.6 & -13.1 \\ & & & 13.1 \end{pmatrix}$$

$$\sum(a_i - \bar{a})^2 = 100, \quad \sum(b_i - \bar{b})^2 = 160, \quad \sum(c_i - \bar{c})^2 = 240.$$

Regression with OLS gives us

$$\hat{y}_i = 2 + 3a_i - 5b_i + 6c_i.$$

Sum of squared residuals is $\mathrm{SS}^{\mathrm{res}} = 80$, total sum of squares is $\mathrm{SST} = 100$

a) Find the 95% confidence intervals for the coefficients $\hat{\beta}_a$, $\hat{\beta}_b$ and $\hat{\beta}_c$ under classic Gauss – Markov assumptions and normal random errors.
b) Find OLS estimates of coefficents in regression $\hat{a}_i = \hat{\gamma}_1 + \hat{\gamma}_b b_i + \hat{\gamma}_c c_i$ and $\mathrm{SS}_a^{\mathrm{res}}$, R_a^2 for this regression.
c) Find OLS estimates of coefficents in regression $\hat{b}_i = \hat{\nu}_1 + \hat{\nu}_a a_i + \hat{\nu}_c c_i$ and $\mathrm{SS}_b^{\mathrm{res}}$, R_b^2 for this regression.
d) Find OLS estimates of coefficents in regression $\hat{c}_i = \hat{\delta}_1 + \hat{\delta}_a a_i + \hat{\delta}_b b_i$ and $\mathrm{SS}_c^{\mathrm{res}}$, R_c^2 for this regression.
e) Find the variance inflation factors VIF_a, VIF_b, VIF_c.

9.4 Consider the model $y = X\beta + u$. There are two centered regressors in the matrix X, variable a and variable b, $\mathrm{VIF}_a = 10$, $\mathrm{SST}_a = 125$, $\mathrm{SST}_b = 5$.

a) Find possible values of sample correlation between a and b.

b) Find $(X^T X)^{-1}$.

9.5 We have two absolutely identical preliminary standardized regressors x and x. The dependent variable y is centered.

In the ridge regression one minimizes the goal function

$$\text{loss}(\hat{\beta}) = (y - \hat{y})^T (y - \hat{y}) + \lambda \hat{\beta}^T \hat{\beta}, \quad \hat{y} = \hat{\beta}_1 x + \hat{\beta}_2 x.$$

a) Find the optimal $\hat{\beta}_1$ and $\hat{\beta}_2$ for fixed λ.

b) What happens to the estimates when $\lambda \to \infty$?

c) What happens to the sum $\hat{\beta}_1 + \hat{\beta}_2$ when $\lambda \to 0$?

9.6

9.7

9.8

10 SVD

10.1 You know the SVD of $X = UDV^T$.

Find the SVD of $X^T X$, $(X^T X)^{-1}$, $(X^T X)^{-1} X^T$, $X(X^T X)^{-1} X^T$.

10.2 The matrix X has size $n \times 3$. All columns of X are standardized, ie they have zero mean and unit sample variance. You know the SVD of $X = UDV^T$ with $\text{diag}(D) = (3, 2, 1)$.

Find the SVD of sample correlation matrix of columns of X, its eigenvalues and eigenvectors.

10.3 We know that $U^T U = I$, where I is an identity matrix. Prove the following statements or provide a counter-example.

a) $UU^T = I$;

b) The lenght of vectors x and Ux is always the same.

c) Scalar products are equal, $\langle x, y \rangle = \langle Ux, Uy \rangle$.

d) The cosine of the angle between x and y is equal to the cosine of angle between Ux and Uy.

10.4 Find the SVD-decomposition of matrices

a) $X = \begin{pmatrix} 0 & -7 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$;

b) $X = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$;

c) $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$;

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

e) $\begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$;

f) $X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{pmatrix}$.

10.5 The $k \times k$ matrix A is constant symmetric positive definite and its highest eigenvalues are 5 and 2.

a) Find $d(v^T Av)$ and $d(v^T Av/v^T v)$.
b) Solve the optimization problem $\max_v v^T Av$ where v is $k \times 1$ vector with $\|v\| = 1$
   Hint: you may use the differential from (a) or Lagrange multiplier method.
c) Solve the optimization problem $\max_w w^T Aw$ where w is $k \times 1$ vector with $\|w\| = 1$ and $w \perp v$.

10.6 The matrix X has SVD $X = UDV^T$. The columns of U are u_1, u_2, ..., u_n. Singular values of the diagonal of D are sorted from highest to lowest by absolute value, $d_{11} = 5$, $d_{22} = 3$. The columns of V are v_1, v_2, ..., v_k.

a) Find $X \cdot v_1$, $X \cdot v_2$.
b) Solve the optimization problem $\max_w \|Xw\|$ where w is $k \times 1$ vector with $\|w\| = 1$.
c) Solve the optimization problem $\max_z \|Xz\|$ where z is $k \times 1$ vector with $\|z\| = 1$ and $z \perp w$.

10.7

10.8

10.9

10.10

10.11

11 Приятно-следственные связи, CUPED

11.1 Предположим, что w_i не зависим с гипотетическими значениями $y_i(0)$ и $y_i(1)$.

Верно ли, что $ATE = \mathbb{E}(y_i(1) - y_i(0))$ можно записать как $\mathbb{E}(y_i \mid w_i = 1) - \mathbb{E}(y_i \mid w_i = 0)$?

11.2 Предположим, что $y_i = \beta_1 + \delta w_i + \beta_2 x_i + \beta_3 x_i^2 + u_i$, выполнены классические предпосылки на u_i: $\mathbb{E}(u_i \mid X) = 0$, $\operatorname{Var}(u_i \mid X) = \sigma^2$, $\operatorname{Cov}(u_i, u_j \mid X) = 0$. Наблюдения представляют собой случайную выборку, четвёртые моменты всех упомянутых случайных величин конечны.

Здесь $w_i \in \{0, 1\}$ — индикатор того, что индивид получил воздействие в рамках рандомизированного эксперимента, x_i — характеристика индивида, а y_i — интересующая переменная.

a) Чему равны $\operatorname{Cov}(w_i, x_i \mid X)$ и $\operatorname{Cov}(w_i, u_i \mid X)$, если рандомизированные эксперимент корректно проведён?
b) Агнесса использует простую регрессию $\hat{y}_i = \hat{alpha} + \hat{\delta}_a w_i$. Верно ли, что оценка Агнессы $\hat{\delta}_a$ будет несмещённой и состоятельной?
c) Бриджит использует разницу средних

$$\hat{\delta}_b = \frac{\sum_{w_i=1} y_i}{n_1} - \frac{\sum_{w_i=1} y_i}{n_0}.$$

   Верно ли, что оценка Бриджит $\hat{\delta}_b$ будет несмещённой и состоятельной?
d) Василиса использует множественную регрессию

$$\hat{y}_i = \hat{alpha} + \hat{\delta}_c w_i + \hat{\gamma} x_i.$$

   Верно ли, что оценка Василисы $\hat{\delta}_c$ будет несмещённой и состоятельной?

11.3 Предположим, что $y_i = \beta_1 + \delta w_i + \beta_2 x_i + \beta_3 x_i^2 + u_i$, выполнены классические предпосылки на u_i: $\mathbb{E}(u_i \mid X) = 0$, $\mathrm{Var}(u_i \mid X) = \sigma^2$, $\mathrm{Cov}(u_i, u_j \mid X) = 0$. Наблюдения представляют собой случайную выборку, четвёртые моменты всех упомянутых случайных величин конечны.

Здесь $w_i \in \{0, 1\}$ — индикатор того, что индивид получил воздействие в рамках рандомизированного эксперимента, x_i — характеристика индивида, а y_i — интересующая переменная.

Василиса использует множественную регрессию

$$\hat{y}_i = \hat{\alpha} + \hat{\delta}_c w_i + \hat{\gamma} x_i.$$

Галатея использует CUPED. Для этого на первом шаге она оценивает множественную регрессию Василисы. Затем Галатея создаёт новую переменную, очищая зависимую переменную y_i от эффекта x_i, $r_i = y_i - \hat{\gamma} x_i$.

На втором шаге Галатея оценивает парную регрессию $\hat{r}_i = \hat{\mu} + \hat{\delta}_d w_i$.

a) Верно ли, что оценки $\hat{\mu}$ и $\hat{\alpha}$ совпадают?
b) Верно ли, что оценки $\hat{\delta}_c$ и $\hat{\delta}_d$ совпадают?
c) Верно ли, что совпадают суммы квадратов остатков в двух регрессиях Галатеи?
d) Верно ли, что совпадают классические стандартные ошибки $se(\hat{\delta}_c)$ и $se(\hat{\delta}_d)$?

12 Maximal likelihood

12.1 Рассмотрим модель регрессии с одним параметром, $y_i = \beta x_i + u_i$, где x_i неслучайны и не все равны нулю, а u_i нормальны $\mathcal{N}(0; \sigma^2)$ и независимы.

Рассмотрим варианты предпосылок:

a) Величина σ^2 известна, и исследователь хочет проверить гипотезу H_0: $\beta = 7$ против H_a: $\beta \neq 7$.
b) Величина β известна, и исследователь хочет проверить гипотезу H_0: $\sigma^2 = 1$ против H_a: $\sigma^2 \neq 1$.
c) Величины σ^2 и β неизвестны, и исследователь хочет проверить гипотезу H_0: $\beta = 7$ против H_a: $\beta \neq 7$.
d) Величины σ^2 и β неизвестны, и исследователь хочет проверить гипотезу H_0: $\beta = 7$, $\sigma^2 = 1$ против H_a: $\beta \neq 7$ или $\sigma^2 \neq 1$.

Для каждого варианта предпосылок:

a) Найдите функцию правдоподобия $\ell(\theta)$, её градиент $s(\theta)$, матрицу Гессе $H(\theta)$, теоретическую информацию Фишера $I(\theta)$.
b) Найдите $\hat{\theta}_{UR}$, $\hat{\theta}_R$, $\ell(\hat{\theta}_{UR})$, $\ell(\hat{\theta}_R)$, $s(\hat{\theta}_{UR})$, $s(\hat{\theta}_R)$.
c) Выпишите формулу для $\mathrm{SS}_R^{\mathrm{res}}$ и $\mathrm{SS}_{UR}^{\mathrm{res}}$.
d) Найдите оценку информации Фишера $\hat{I}_R$, $\hat{I}_{UR}$ подставив в теоретическую информацию Фишера оценённые параметры.
e) Выведите формулы для LR, LM и W статистики. Можно выражать их через $\mathrm{SS}_R^{\mathrm{res}}$ и $\mathrm{SS}_{UR}^{\mathrm{res}}$.
f) Упорядочьте статистики по возрастанию.

12.2 Величины y_1, y_2, ..., y_n независимы и экспоненциально распределены с параметром λ. По выборке из 100 наблюдений оказалось, что $\sum y_i = 200$. Исследователь Андреас хочет проверить гипотезу H_0: $\mathbb{E}(y_i) = 1$ против альтернативной $\mathbb{E}(y_i) \neq 1$.

a) Выпишите логарифмическую функцию правдоподобия $\ell(\lambda)$;
b) Найдите оценку $\hat{\lambda}$ методом максимального правдоподобия в общем виде и для имеющейся выборки;

c) Найдите теоретическую информацию Фишера $I(\lambda)$ для n наблюдений;

d) Выведите формулы для статистик отношения правдоподобия, множителей Лагранжа и Вальда в общем виде;

e) Найдите значения статистик отношения правдоподобия, множителей Лагранжа и Вальда для имеющейся выборки;

f) Проверьте гипотезу H_0 с помощью трёх статистик.

12.3 Рассмотрим модель простой регрессии $y_i = \beta x_i + u_i$, где ошибки u_i независимы и имеют стандартное нормальное распределение, $u_i \sim \mathcal{N}(0; 1)$. По выборке из 100 наблюдений оказалось, что $\sum x_i^2 = 100$, $\sum y_i^2 = 900$, а $\sum y_i x_i = 250$. Исследователь Рамирес хочет проверить H_0: $\beta = 0$.

a) Выпишите логарифмическую функцию правдоподобия $\ell(\beta)$;

b) Найдите оценку $\hat{\beta}$ методом максимального правдоподобия в общем виде и для имеющейся выборки;

c) Найдите теоретическую информацию Фишера $I(\beta)$ для n наблюдений;

d) Выведите формулы для статистик отношения правдоподобия, множителей Лагранжа и Вальда в общем виде;

e) Найдите значения статистик отношения правдоподобия, множителей Лагранжа и Вальда для имеющейся выборки;

f) Проверьте гипотезу H_0 с помощью трёх статистик.

12.4 Исследовательница Геральдина заглядывает n раз в случайные аудитории бывшей шпульно-катушечной фабрики. В каждой аудитории независимо от других идёт семинар по теории вероятностей, эконометрике, микро или макро. Пусть p_1, p_2, p_3 — это вероятности семинаров по теории вероятностей, эконометрике и микро. Вероятность семинара по макро мы отдельным параметром не вводим, так как иначе параметры будут зависимы и нужно будет искать ограниченный экстремум правдоподобия. Пусть y_1, y_2, y_3 — количество попаданий Геральдины на теорию вероятностей, эконометрику и микро.

По выборки из 100 наблюдений оказалось, что $y_1 = 20$, $y_2 = 30$, $y_3 = 20$. Геральдина предполагает, что все четыре дисциплины равновероятны.

a) Выпишите логарифмическую функцию правдоподобия $\ell(p)$;

b) Найдите оценку $\hat{p}$ методом максимального правдоподобия в общем виде и для имеющейся выборки;

c) Найдите теоретическую информацию Фишера $I(p)$ для n наблюдений;

d) Найдите явно $I^{-1}(p)$;

e) Выведите формулы для статистик отношения правдоподобия, множителей Лагранжа и Вальда в общем виде;

f) Найдите значения статистик отношения правдоподобия, множителей Лагранжа и Вальда для имеющейся выборки;

g) Проверьте гипотезу H_0 с помощью трёх статистик на уровне значимости 5%.

h) (*) Обобщиет формулы трёх статистик на случай произвольного количества дисциплин и произвольной гипотезы H_0: $p = p^0$.

12.5 Логарифмическая функция правдоподобия имеет вид

$$\ell(\theta) = a - \frac{1}{2}(\theta - h(y))' Q(\theta - h(y)),$$

где Q — постоянная симметричная матрица, а $h(y)$ — функция от выборки. Вектор параметров θ состоит из двух блоков, а матрица Q — из четырёх блоков

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad Q = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

Настырный исследователь Никанор хочет проверить гипотезу H_0: $\theta_1 = \theta_1^0$ про часть параметров, входящих в вектор θ;

a) Найдите неограниченную оценку метода максимального правдоподобия $\hat{\theta}^{UR}$;

b) Найдите ограниченную оценку метода максимального правдоподобия $\hat{\theta}^{R}$;

c) Выведите формулу для LR статистики;

d) Выведите формулу для LM статистики;

e) Выведите формулу для W статистики;

f) Какие из указанных формул равны?

12.6 Рассмотрим модель множественной регрессии $y = X\beta + u$, где регрессоры детерминистические, ошибки имеют многомерное нормальное распределение, а ковариационная матрица $\mathrm{Var}(u)$ единичная. Разобьём вектор коэффициентов β на две части

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

a) Докажите, что логарифмическая функция правдоподобия представима в виде

$$\ell(\theta) = a - \frac{1}{2}(\theta - h(y))' Q(\theta - h(y)),$$

b) Явно найдите матрицу Q и функцию $h(y)$;

c) Выведите формулу для LR, LM и теста Вальда для проверки гипотезы H_0: $\beta_1 = \beta_1^0$;

d) Как найденная формула отличается от обычной F статистики?

e) Как найденная формула упрощается для случая проверки гипотезы о незначимости регрессии в целом?

12.7 Рассмотрим модель множественной регрессии $y = X\beta + u$, где регрессоры детерминистические, ошибки имеют многомерное нормальное распределение, а ковариационная матрица $\mathrm{Var}(u) = \sigma^2 I$. Разобьём вектор коэффициентов β на две части

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

a) Докажите, что логарифмическая функция правдоподобия представима в виде

$$\ell(\theta) = a + (\theta - h(y))' Q(\theta - h(y)),$$

b) Явно найдите матрицу Q и функцию $h(y)$;

c) Выведите формулы для LR, LM и теста Вальда для проверки гипотезы H_0: $\beta_1 = \beta_1^0$;

d) Как найденные формулы отличается от обычной F статистики?

e) Как найденные формулы упрощается для случая проверки гипотезы о незначимости регрессии в целом?

12.8 Рассмотрим LR, LM и W статистики в задаче оценки параметров модели $y = X\beta + u$, с нормальными ошибками $u \sim \mathcal{N}(0; \sigma^2 \cdot I)$ и неизвестной σ^2.

Известно, что при проверке гипотезы о линейных ограничениях на β оказывается, что

$$\begin{cases} LR = n \ln s \\ W = n(s-1) \\ LM = n(s-1)/s \end{cases},$$

где $s = SS^{res}_R / SS^{res}_{UR}$.

Докажите, что $LM \leqslant LR \leqslant W$.

12.9 Рассмотрим LR, LM и W статистики в задаче оценки параметров модели $y = X\beta + u$, с нормальными ошибками $u \sim \mathcal{N}(0; \sigma^2 \cdot I)$ и неизвестной σ^2.

Известно, что при проверке гипотезы о линейных ограничениях на β оказывается, что

$$\begin{cases} LR = n \ln s \\ W = n(s-1) \\ LM = n(s-1)/s \end{cases},$$

где $s = SS^{res}_R / SS^{res}_{UR}$.

Эконометрэсса Фиалка по 60 наблюдениям проверяет гипотезу о равенстве пяти параметров нулю в регрессии с десятью параметрами β при 5%-м уровне значимости.

Найдите точные критические значения для LR, LM и W и сравните их с асимптотическими.

13 Heteroskedasticity

13.1 Имеется три наблюдения

x_i	1	2	2
y_i	1	2	3

Экономэтр Антоний хочет оценить зависимость $y_i = \beta x_i + u_i$.

a) Найдите оценку $\hat{\beta}$ с помощью МНК;
b) Найдите стандартную ошибку $se(\hat{\beta})$ предполагая гомоскедастичность;
c) Найдите робастные к гетероскедастичности стандартные ошибки $se_{HC0}(\hat{\beta})$ и $se_{HC3}(\hat{\beta})$;
d) Найдите эффективную оценку $\hat{\beta}$, если дополнительно известно, что $\mathrm{Var}(u_i \mid x_i) = \sigma^2(3x_i - 2)$;
e) Найдите эффективную оценку $\hat{\beta}$, если дополнительно известно, что

$$\mathrm{Var}(u \mid X) = \begin{pmatrix} 4\sigma^2 & -\sigma^2 & 0 \\ -\sigma^2 & 9\sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$$

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

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

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	SS^{res}
$i = 1, \ldots, 30$	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%.

13.4 Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$ по 50 наблюдениям. 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: $\hat{u}_i^2 = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i$. 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%.

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

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

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

b) ожидаемое значение статистики Уайта;

c) дисперсию статистики Уайта.

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

13.8 Экономэтр Антоний исследует зависимость надоя коров в литрах в год, y_i, от дамми-переменной x_i, отвечающей за прослушивание коровами ежедневно Девятой симфонии, $y_i = \beta_1 + \beta_2 x_i + u_i$. Антоний раздобыл следующие данные:

Подвыборка	Размер	$\sum y_i$	$\sum y_i^2$
$x_i = 0$	$n_0 = 100$	200	4000
$x_i = 1$	$n_1 = 100$	300	5000

a) Найдите МНК-оценки β_1 и β_2;

b) Постройте 95%-ый доверительный интервал для β_2 предполагая гомоскедастичность u_i;

c) Найдите робастную к гетероскедастичности оценку $\widehat{\operatorname{Var}}_{HC0}(\hat{\beta})$;

d) Найдите робастную к гетероскедастичности оценку $\widehat{\operatorname{Var}}_{HC3}(\hat{\beta})$;

e) Постройте 95%-ый доверительный интервал для β_2 с помощью скорректированной $se_{HC0}(\hat{\beta}_2)$;

f) Дополнительно предположив, что $\operatorname{Var}(u_i \mid x_i) = \sigma^2(1 + 3x_i)$, найдите эффективную оценку $\hat{\beta}_2$ и постройте доверительный интервал для неё.

13.9 Эконометресса Прасковья использует традиционную оценку ковариационной матрицы, а эконометресса Мелони — оценку Уайта.

Какие оценки дисперсии $\hat{\beta}_1$ и формулы для t-статистики получат Прасковья и Мелони в модели $y_i = \beta_1 + u_i$?

13.10 Эконометресса Прасковья использует традиционную оценку ковариационной матрицы, а эконометресса Мелони — оценку Уайта.

Обе эконометрессы оценивают модель $y_i = \beta_1 + \beta_2 d_i + u_i$, где d_i — дамми-переменная, равна 0 или 1. Дамми-переменная делит выборку на две части. Обозначим количество наблюдений в «нулевой» части как n_0, среднее — как $\bar{y}_0$, и общую сумму квадратов — как TSS_0. Аналогичные величины для «единичной» части выборки — n_1, $\bar{y}_1$ и TSS_1. И для всей выборки — n, $\bar{y}$, TSS.

a) Найдите оценки $\hat\beta_1$ и $\hat\beta_2$.

b) Найдите оценки $\widehat{\mathrm{Var}}(\hat\beta_1)$ и $\widehat{\mathrm{Var}}_W(\hat\beta_1)$. Верно ли, что $\widehat{\mathrm{Var}}_W(\hat\beta_1) \geqslant \widehat{\mathrm{Var}}(\hat\beta_1)$?

c) Найдите оценки $\widehat{\mathrm{Var}}(\hat\beta_2)$ и $\widehat{\mathrm{Var}}_W(\hat\beta_2)$. Верно ли, что $\widehat{\mathrm{Var}}_W(\hat\beta_2) \geqslant \widehat{\mathrm{Var}}(\hat\beta_2)$?

d) Найдите оценки $\widehat{\mathrm{Cov}}(\hat\beta_1, \hat\beta_2)$ и $\widehat{\mathrm{Cov}}_W(\hat\beta_1, \hat\beta_2)$.

e) Найдите оценки $\widehat{\mathrm{Var}}(\hat\beta_1 + \hat\beta_2)$ и $\widehat{\mathrm{Var}}_W(\hat\beta_1 + \hat\beta_2)$.

13.11 В модели $y_i = \beta x_i + u_i$ предполагается гетероскедастичность вида $\mathrm{Var}(u_i) = \exp(\gamma_1 + \gamma_2 x_i)$ и нормальность ошибок.

a) Сформулируйте гипотезу о гомоскедастичности с помощью коэффициентов.

b) Выведите в явном виде оценку максимального правдоподобия при предположении о гомоскедастичност

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

d) Выведите в явном виде формулу для LM теста множителей Лагранжа.

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

13.13 В оригинальном тесте Бройша-Пагана на гетероскедастичность два шага. Сначала строится основная регрессия y_i на некоторые регрессоры и получаются остатки $\hat u_i$. На втором шаге строится регрессия квадрата остатков $\hat u_i$ на переменные, от которых потенциально зависит условная дисперсия $\mathrm{Var}(u_i \mid Z)$. Статистика Бройша-Пагана считается как $BP = ESS/2$, где ESS — объяснённая сумма квадратов регрессии второго шага. Оригинальный тест Уайта считается как $W = nR^2$, где R^2 — коэффициент детерминации регрессии второго шага.

a) Найдите отношение $\frac{nR^2}{ESS/2}$;

b) Найдите предел по вероятности $\operatorname{plim} \frac{nR^2}{ESS/2}$;

c) Какое распределение имеют статистики BP и W?

d) Какой вид имеет статистика множителей Лагранжа?

распотрошить статью BP на задачу, статья о похожести BP и W, отдельно Коэнкера про студентизированную версию

14 Логит и пробит!

14.1 Бандерлог из Лога оценил логистическую регрессию по четырём наблюдениям и одному признаку с константой, получил $b_i = \hat{\mathbb{P}}(y_i = 1 \mid x_i)$, но потерял последнее наблюдение:

y_i	b_i
1	0.7
-1	0.2
-1	0.3
?	?

a) Выпишите функцию правдоподобия для задачи логистической регрессии.

b) Выпишите условие первого порядка по коэффициенту перед константой.

c) Помогите бандерлогу восстановить пропущенные значения!

14.2 Рассмотрим логистическую функцию $\Lambda(w) = e^w/(1 + e^w)$.

a) Как связаны между собой $\Lambda(w)$ и $\Lambda(-w)$?

b) Как связаны между собой $\Lambda'(w)$ и $\Lambda'(-w)$?

c) Постройте графики функций $\Lambda(w)$ и $\Lambda'(w)$.

d) Найдите $\Lambda(0)$, $\Lambda'(0)$, $\ln \Lambda(0)$.

e) Найдите обратную функцию $\Lambda^{-1}(p)$.

f) Как связаны между собой $\frac{d \ln \Lambda(w)}{dw}$ и $\Lambda(-w)$?

g) Как связаны между собой $\frac{d \ln \Lambda(-w)}{dw}$ и $\Lambda(w)$?

h) Разложите $h(\beta_1, \beta_2) = \ln \Lambda(y_i(\beta_1 + \beta_2 x_i))$ в ряд Тейлора до второго порядка в окрестности точки $\beta_1 = 0$, $\beta_2 = 0$.

14.3 Винни-Пух знает, что мёд бывает правильный, $\text{honey}_i = 1$, и неправильный, $\text{honey}_i = 0$. Пчёлы также бывают правильные, $\text{bee}_i = 1$, и неправильные, $\text{bee}_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$\text{honey}_i = 1$	$\text{honey}_i = 0$
$\text{bee}_i = 1$	12	36
$\text{bee}_i = 0$	32	20

Винни-Пух использует логистическую регрессию с константой для прогнозирования правильности мёда с помощью правильности пчёл.

a) Какие оценки коэффициентов получит Винни-Пух?

b) Какой прогноз вероятности правильности мёда при встрече с неправильными пчёлами даёт логистическая модель? Как это число можно посчитать без рассчитывания коэффициентов?

c) Проверьте гипотезу о том, что правильность пчёл не оказывает влияние на правильность мёда с помощью тестов LR, LM и W.

14.4 Винни-Пух оценил логистическую регрессию для прогнозирования правильности мёда от высоты дерева (м) x_i и удалённости от дома (км) z_i: $\ln odds_i = 2 + 0.3x_i - 0.5z_i$.

a) Оцените вероятность того, что $y_i = 1$ для $x = 15$, $z = 3.5$.

b) Оцените предельный эффект увеличения x на единицу на вероятность того, что $y_i = 1$ для $x = 15$, $z = 3.5$.

c) При каком значении x предельный эффект увеличения x на единицу в точке $z = 3.5$ будет максимальным?

14.5 Придумайте такие три наблюдения для парной логистической регрессии, чтобы все x_i были разными, не все y_i были одинаковые, а оценки логит-модели не существовали.

Какое решение задачи этой проблемы разумно предложить при большом количестве наблюдений?

14.6 При оценке логит модели $\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$ по 500 наблюдениям оказалось, что $\hat\beta_1 = 0.7$ и $\hat\beta_2 = 3$. Оценка ковариационной матрицы коэффициентов имеет вид

$$\begin{pmatrix} 0.04 & 0.01 \\ 0.01 & 0.09 \end{pmatrix}$$

a) Проверьте гипотезу о незначимости коэффициента $\hat\beta_2$.

b) Найдите предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$ при $x_i = -0.5$.

c) Найдите максимальный предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$.

d) Постройте точечный прогноз вероятности $\mathbb{P}(y_i = 1)$ если $x_i = -0.5$.

e) Найдите стандартную ошибку построенного прогноза.

f) Постройте 95%-ый доверительный интервал для $\mathbb{P}(y_i = 1)$ двумя способами (через преобразование интервала для $\hat{y}_i^*$ и через дельта-метод).

14.7 Почему в пробит-модели предполагается, что $u_i \sim \mathcal{N}(0; 1)$, а не $u_i \sim \mathcal{N}(0; \sigma^2)$ как в линейной регрессии?

14.8 Что произойдёт с оценками логит-модели $\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$, их стандартными ошибками, если у зависимой переменной поменять 0 и 1 местами?

14.9 Исследователь Матвей оценил логит-модель по 10 тысячам наблюдений, $\hat{\mathbb{P}}(y_i = 1) = \Lambda(-0.5 + 1.2x_i)$. Переменная x_i — бинарная, 4 тысячи единиц и 6 тысяч нулей.

a) Сколько наблюдений с $y_i = 1$?

b) Сколько наблюдений с $y_i = 1$ и $x_i = 0$?

c) Сколько наблюдений с $y_i = 0$ и $x_i = 1$?

14.10 Если выбрать покемона наугад, то с вероятностью p покемон окажется ядовитым, а с вероятностью $(1-p)$ — не ядовитым. Вес ядовитых покемонов распределен нормально, $\mathcal{N}(\mu_1, \sigma^2)$, а вес неядовитых — нормально с другим ожиданием и той же дисперсией, $\mathcal{N}(\mu_2, \sigma^2)$.

Охотник Джон только что поймал покемона и взвесил его.

a) Найдите условную вероятность того, что покемон ядовит, если его вес равен x.

b) Запишите найденную условную вероятность в виде

$$\mathbb{P}(\text{покемон ядовит} \mid x) = F(\beta_1 + \beta_2 x),$$

где F — некая функция распределения. Выразите β_1 и β_2 через исходные параметры. Как называется функция F?

14.11 Три богатыря в дальнем походе раздобыли набор данных в котором и зависимая переменная и предикторы — бинарные. Илья Муромец оценивает логит-модель. Добрыня Никитич использует обычный МНК. Алёша Попович строит классификационное дерево максимально возможной длины, используя энтропию для деления узла на два. Змей Горыныч, раздобывший тот же набор данных, использует пробит модель.

Как между собой соотносятся прогнозы вероятностей у трёх богатырей и Змея?

15 Hobbit

15.1 The random variable u has standard normal distribution $\mathcal{N}(0, 1)$ and $w = \max\{u + s, 0\}$, where s is a constant.

a) Write the function $\lambda(s) = \mathbb{E}(u \mid u + s > 0)$ in terms of standard normal density or distribution function.

b) Write $\mathbb{P}(w > 0)$, $\mathbb{E}(w)$ and $\mathbb{E}(w \mid w > 0)$ in terms of standard normal density or distribution function.

The function $\lambda(s)$ is called inverse Mills ratio.

15.2 Consider the tobit-I model:

$$\begin{cases} y_i^* = x_i^T \beta + u_i, \quad y^* = X^T \beta + u \\ (u \mid X) \sim \mathcal{N}(0, \sigma^2 I) \\ y_i = \max\{y_i^*, 0\} \\ (x_i, y_i) \text{ are observable for all } i \end{cases}$$

Find $\mathbb{E}(y_i^* \mid x_i)$, $\mathbb{E}(y_i \mid x_i)$, $\mathbb{E}(y_i \mid x_i, y_i^* > 0)$.

15.3 Consider a logit model with a constant but without other regressors,

$$\mathbb{P}(y_i^* = 1) = \Lambda(\beta).$$

Observations are independent.

a) Estimate β for the sample $y^* = (1, 1, 0, 1, 0)$.

Dark Fairy removes all observations with $y_i^* = 0$.

b) Now we have two unknowns: n and β. Estimate both unknown parameters for the truncated sample $y = (1, 1, 1)$.

15.4 Random variables y_i^* are independnent and exponentially distributed with unknown rate λ.

Dark Fairy replaces all observations greater than 10 by 10.

a) Estimate λ by maximal likelihood for the sample $y = (5, 10, 6, 10)$.
b) Find the estimator of λ using maximum likelihood for large arbitrary sample and construct the 95% confidence interval.

15.5 Random variables y_i^* are independnent and exponentially distributed with unknown rate λ.

Dark Fairy removes all observations lower than 10.

a) Estimate λ by maximal likelihood for the sample $y = (12, 13, 14)$.
b) Find the estimator of λ using maximum likelihood for large arbitrary sample and construct the 95% confidence interval.
c) What happens if one tries to estimate original sample size n using truncated sample?

15.6 Random variables y_i^* are independent and uniformly distributed on $[0, a]$ with unknown $a > 2$.

Dark Fairy arrives and makes her black magic!

a) If possible estimate a using maximum likelihood for the sample $y = (1.7, 2, 0.8, 2, 2, 0.9)$, if Dark Fairy replaces all observations larger than 2 by 2.
b) If possible estimate a using maximum likelihood for the sample $y = (1.7, 0.8, 0.9)$, if Dark Fairy removes all observations larger than 2.
c) If possible estimate a using maximum likelihood for the sample $y = (2.7, 2, 3.8, 2, 2, 4.9)$, if Dark Fairy replaces all observations less than 2 by 2.
d) If possible estimate a using maximum likelihood for the sample $y = (2.7, 3.8, 4.9)$, if Dark Fairy removes all observations less than 2.

15.7 The joint distribution of s_i and y_i^* is described by

	$y_i^* = 0$	$y_i^* = 1$
$s_i = 0$	$2a$	$1 - 6a$
$s_i = 1$	$3a$	a

Observations are independent.

Dark Fairy arrives and makes her black magic! Sleeping Beauty awakens and observes the modified sample $y = (1, 1, 0, 1, 1, 0)$.

a) If possible estimate a by maximum likelihood, if Dark Fairy replaces all y_i^* by 0 where $s_i = 1$.
b) If possible estimate a by maximum likelihood, if Dark Fairy removes all observations with $s_i = 1$.
c) If possible estimate a by maximum likelihood, if Dark Fairy does nothing.

15.8

16 Endogeneity

A good source on instrumental variables is [Sch].

16.1 Величины x_i, z_i и u_i имеют совместное распределение, задаваемое табличкой:

x_i	0	1	0	1
z_i	0	1	0	0
u_i	-1	-1	1	1
Вероятность	0.2	0.3	0.3	0.2

Рассмотрим модель $y_i = \beta x_i + u_i$.

a) Найдите $\operatorname{plim} \hat{\beta}_{OLS}$;

b) Найдите $\operatorname{plim} \hat{\beta}_{IV}$, если в качестве инструмента для x_i используется z_i;

16.2 Рассмотрим три вектора: y, x и z. Проведем гипер-плоскость ортогональную z через конец вектора y. Эта гипер-плоскость пересекает прямую порождаемую вектором x в точке $\hat{\beta}_{IV} x$.

Исходя из данного геометрического определения $\hat{\beta}_{IV}$:

a) Выведите алгоритм двухшагового МНК;

b) Выведите явную формулу для $\hat{\beta}_{IV}$;

c) Докажите, что оценка $\hat{\beta}_{IV}$ показывает, насколько в среднем растёт y при таком росте z, при котором x в среднем растёт на единицу.

сказать смысл IV попроще? на две фразы?

16.3 Рассмотрим модель $y_i = \beta x_i + u_i$. Исследовательница Мишель строит оценку $\hat{\beta}_{IV}$ в регрессии y на x с инструментом z. Исследовательница Аграфена строит обычную МНК оценку в регрессии $\hat{y} = \hat{\beta}_x x + \hat{\beta}_w w$.

a) Выразите w через x, z и y так, чтобы оценка $\hat{\beta}_{IV}$ Мишель и оценка $\hat{\beta}_x$ Аграфены совпали.

b) Сформулируйте ещё одну интерпретацию оценки $\hat{\beta}_{IV}$;

16.4 Величины x_i, z_i и u_i имеют совместное распределение, параметры которого известны:

$$\operatorname{Var}\left(\begin{pmatrix} x_i \\ z_i \\ u_i \end{pmatrix}\right) = \begin{pmatrix} 5 & 1 & -1 \\ 1 & 9 & 0 \\ -1 & 0 & 4 \end{pmatrix}; \quad \mathbb{E}\left(\begin{pmatrix} x_i \\ z_i \\ u_i \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix}$$

Наблюдения с разными номерами независимы и одинаково распределены.

Рассмотрим модель $y_i = \beta_1 + \beta_2 x_i + u_i$.

a) Найдите $\operatorname{plim} \hat{\beta}_2^{LS}$ и $\operatorname{plim} \hat{\beta}_1^{LS}$; Являются ли оценки состоятельными?

b) Храбрый исследователь Афанасий использует двухшаговый МНК. На первом шаге он строит регрессию x_i на константу и z_i, $\hat{x}_i = \hat{\gamma}_1 + \hat{\gamma}_2 z_i$. А на втором регрессию $\hat{y}_i = \hat{\beta}_1^{IV} + \hat{\beta}_2^{IV} \hat{x}_i$. Найдите $\operatorname{plim} \hat{\beta}_2^{IV}$ и $\operatorname{plim} \hat{\beta}_1^{IV}$; Являются ли оценки состоятельными?

c) Как изменятся $\operatorname{plim} \hat{\beta}_2^{IV}$ и $\operatorname{plim} \hat{\beta}_1^{IV}$, если Афанасий забудет включить константу на первом шаге?

16.5 Приведите примеры дискретных случайных величин u и x, таких, что

a) $\mathbb{E}(u) = 0$, $\mathbb{E}(\varepsilon \mid x) = 0$, но величины зависимы. Чему в этом случае равно $\operatorname{Cov}(\varepsilon, x)$?

b) $\mathbb{E}(u) = 0$, $\operatorname{Cov}(\varepsilon, x) = 0$, но $\mathbb{E}(\varepsilon \mid x) \neq 0$. Зависимы ли эти случайные величины?

16.6 Economaîtresse Agnesse would like to estimate the model $y_i = \beta_1 + \beta_2 x_i + u_i$. Unfortunately the variable x_i is unobserved. Instead she observes x_i^*. The variable x_i^* includes a measurement error, $x_i^* = x_i + a_i$. Variances are known $\mathrm{Var}(x_i) = 9$, $\mathrm{Var}(a_i) = 4$, $\mathrm{Var}(u_i) = 1$. Variables x_i, a_i and u_i are normally distributed and independent. Observations are independent.

Agnesse estimates uses ordinary least squares for the regression $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i^*$.

a) Find $\operatorname{plim} \hat{\beta}_2$.
b) Is the estimator consistent?
c) Find $\mathbb{E}(\hat{\beta}_2 \mid x^*)$ and $\mathbb{E}(\hat{\beta}_2)$.
d) Is the estimator unbiased?

16.7 Эконометресса Анжелла хочет оценить модель $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + u_i$, но, к сожалению, величина w_i ненаблюдаема. Известно, что $\mathrm{Var}(x_i) = 9$, $\mathrm{Var}(w_i) = 4$, $\mathrm{Var}(u_i) = 1$ и $\mathrm{Cov}(x_i, w_i) = -2$. Случайная составляющая не коррелированна с регрессорами. Наблюдения представляют собой случайную выборку.

За неимением w_i Анжелла оценивает регрессию $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ с помощью МНК.

a) Найдите $\operatorname{plim} \hat{\beta}_2$.
b) Являются ли оценки, получаемые Анжеллой, состоятельными?

16.8 Эконометресса Венера оценивает регрессию

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

А на самом деле $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^3 + u_i$, причём $x_i \sim \mathcal{N}(0;1)$ и ошибки $u_i \mid x_i \sim \mathcal{N}(0;\sigma^2)$. Все остальные предпосылки теоремы Гаусса-Маркова выполнены.

a) Будут ли оценки $\hat{\beta}_1$ и $\hat{\beta}_2$, получаемые Венерой, состоятельными?
b) В чём состоит трудность проверки несмещённости оценок?
c) Будут ли состоятельны оценки МНК, если в качестве инструмента Венера возьмёт $z_i = \sin x_i$? А $z_i = \cos x_i$?
d) Как изменятся ответы на предыдущие пункты, если истина имеет вид $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + w_i$?

16.9 Вектора (x_1, u_1), (x_2, u_2), …независимы и одинаково распределены. Также известно, что $x_i \sim \mathcal{N}(10; 9)$ и $\mathbb{E}(u_i \mid x_i) = 0$.

Найдите $\operatorname{plim} \left(\frac{1}{n} x'x\right)^{-1}$, $\operatorname{plim} \frac{1}{n} x'u$ и $\operatorname{plim} (x'x)^{-1} x'u$

16.10 Возможно ли, что $\mathbb{E}(x_i \mid u_i) = 0$ и $\mathbb{E}(u_i \mid x_i) = 0$, но при этом x_i и u_i зависимы?

16.11 В некотором интституте на некотором факультете задумали провести эксперимент: раздать студентам учебники разных цветов случайным образом и посмотреть на итоговую успеваемость по эконометрике. Учебники есть двух цветов: зелёненькие и красненькие. Поэтому модель имеет вид:

$$y_i = \beta_1 + \beta_2 green_i + u_i$$

Здесь y_i — результат по эконометрике, $green_i$ — дамми-переменная на зелёный учебник и u_i — прочие характеристики студента. Зелёные и красненькие учебники планировалось раздавать равновероятно. Однако библиотекарь всё прошляпил и разрешил студентам самим выбирать учебник, какой понравится. В результате вместо переменной $green_i$ получилась переменная $green_i^*$. Известно, что $\mathbb{E}(green_i^*) = \alpha$ и $\mathrm{Cov}(green_i^*, u_i) = \gamma$.

Де-факто оценивалась модель

$$\hat{y}_i = \hat{\theta}_1 + \hat{\theta}_2 green_i^*$$

a) Найдите $\operatorname{plim} \hat{\theta}_1$, $\operatorname{plim} \hat{\theta}_2$.

b) Найдите $\mathbb{E}\,\hat{\theta}_1$, $\mathbb{E}\,\hat{\theta}_2$.

16.12 Если возможно, придумайте такие случайные величины x_1, x_2, u_1, что x_1 и x_2 независимы и одинаково распределены, и при этом выполнены условия:

a) $\mathbb{E}(u_1 \mid x_1) = 0$, но $\mathbb{E}(u_1 \mid x_1, x_2) \neq 0$;

b) $\mathbb{E}(u_1 \mid x_1) \neq 0$, но $\mathbb{E}(u_1 \mid x_1, x_2) = 0$.

16.13 Рассмотрим классическую парную регрессию со стохастическим регрессором. Всего три наблюдения:

y_1	x_1
y_2	x_2
y_3	x_3

a) Соедините линиями независимые случайные величины.

b) Соедините линиями одинаково распределённые случайные величины.

16.14 Let z_i be binary variable.

a) Prove that
$$\frac{\operatorname{Cov}(y_i, z_i)}{\operatorname{Cov}(x_i, z_i)} = \frac{\mathbb{E}(y_i \mid z_i = 1) - \mathbb{E}(y_i \mid z_i = 0)}{\mathbb{E}(x_i \mid z_i = 1) - \mathbb{E}(x_i \mid z_i = 0)}.$$

b) Prove that
$$\frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0},$$
where $\bar{x}_1$ is the average value of x when $z = 1$, $\bar{x}_0$ is the average value of x when $z = 0$, and same notation applies for $\bar{y}_1$ and $\bar{y}_0$.

17 Системы уравнений

Прочесть про системы уравнений можно, например, в [Pie][2].

17.1 Наблюдения представляют собой случайную выборку и удовлетворяют системе:

$$\begin{cases} q_t = \alpha p_t + u_{t1} \\ q_t = \beta p_t + u_{t2}. \end{cases},$$

Ковариационная матрица вектора u_t пропорциональна единичной.

a) Наивный Иван оценивает первое уравнение с помощью МНК. Какую оценку $\hat{\alpha}$ он получит при больших n?

b) На секунду предположим, что Ивану наблюдал бы также значения ошибки u_{t2}. Смог бы в этой ситации Иван получить состоятельные оценки α и β?

17.2 Наблюдения представляют собой случайную выборку. Зависимые переменные y_{t1} и y_{t2} находятся из системы:

$$\begin{cases} y_{t1} = \beta_{11} x_t + u_{t1} \\ y_{t2} = \beta_{21} z_t + \beta_{21} y_{t1} + u_{t2} \end{cases},$$

[2] В этих заметках написано что-то странное про неидентифицируемость тождеств. А в целом заметки шикарны.

где вектор ошибок u_t имеет совместное нормальное распределение

$$u_t \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Переменные x_t и z_t — экзогенные.

Эконометресса Анжела оценивает с помощью МНК первое уравнение, а эконометресса Эвридика — второе.

a) Найдите пределы по вероятности получаемых ими оценок.
b) Будут ли оценки состоятельными?

17.3 Экзамен 2018-2019. Рассмотрим систему одновременных уравнений

$$\begin{cases} y_1 = \gamma_1 y_2 + \beta_{11} x_1 + \beta_{12} x_2 + \beta_{13} x_3 + u_1 \\ y_2 = \gamma_2 y_1 + \beta_{21} x_1 + \beta_{22} x_2 + \beta_{23} x_3 + u_2 \end{cases}.$$

Здесь y_j — эндогенные переменные, а x_j — экзогенные. Наблюдения представляют собой случайную выборку. С помощью условий ранга и порядка проверьте идентифицируемость системы

a) в общем случае;
b) при наложении дополнительных ограничений $\gamma_1 - \gamma_2 = \beta_{23} = 0$.

17.4 Экзамен 2018-2019. Рассмотрим систему одновременных уравнений

$$\begin{cases} y_{1t} = \gamma_{10} + \beta_{12} y_{2t} + \beta_{13} y_{3t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} + \varepsilon_{1t} \\ y_{2t} = \gamma_{20} + \beta_{21} y_{1t} + \gamma_{21} x_{1t} + \varepsilon_{2t} \\ y_{3t} = \gamma_{30} + \beta_{31} y_{1t} + \beta_{32} y_{2t} + \gamma_{31} x_{1t} + \gamma_{33} x_{3t} + \varepsilon_{3t} \end{cases}$$

Наблюдения представляют собой случайную выборку. Переменные y_i — эндогенные, переменные x_i — экзогенные.

a) Идентифицируемо ли каждое из уравнений системы?
b) К чему приведёт применение к первому уравнению двухшагового метода наименьших квадратов?

17.5 У Вовочки имеется выборка в 100 наблюдений. Вовочка применяет двухшаговый МНК к первому уравнению системы одновременных уравнений.

a) Что произойдёт, если для первого уравнения не выполнено условие порядка?
b) Что произойдёт, если для первого уравнения выполнено условие порядка, но не выполнено условие ранга?
c) Что произойдёт, если для первого уравнения выполнены условие порядка и условие ранга?

18 GMM

Прочесть про обобщённый метод моментов можно, например, [Zso10].

18.1 Исследователь Максимилиан оценивает параметр θ с помощью двух моментных условий, $\mathbb{E}(y_i) = 2\theta$ и $\mathbb{E}(1/y_i) = \theta$. С трудом Максимилиан нашёл 200 наблюдений и оказалось, что $\sum 1/y_i = 1.5$. Сначала Максимилиан оценил θ с помощью простого метода моментов и первого моментного условия, получил $\hat{\theta} = 1$.

Затем Максимилиан решил применить обобщённый метод моментов, чтобы учесть оба момента. В процессе получения GMM оценки Максимилиан обнаружил, что $\text{Cov}(y_i, 1/y_i) = -\theta^2$, $\text{Var}(1/y_i) = 9\theta^2$, $\mathbb{E}(y_i^2) = 20\theta^2$.

a) Найдите ковариационную матрицу моментных условий, $\text{Var}(g)$;

b) Найдите оптимальную теоретическую матрицу весов W;

c) Оцените матрицу весов $\hat{W}$;

d) Найдите оценку GMM с использованием найденной оценки матрицы весов;

18.2 Величины X_i равномерны на отрезке $[-a; 3a]$ и независимы. Есть несколько наблюдений, $X_1 = 0.5$, $X_2 = 0.7$, $X_3 = -0.1$.

a) Найдите $\mathbb{E}(X_i)$ и $\mathbb{E}(|X_i|)$.

b) Постройте оценку метода моментов, используя $\mathbb{E}(X_i)$.

c) Постройте оценку метода моментов, используя $\mathbb{E}(|X_i|)$.

d) Постройте оценку обобщёного метода моментов используя моменты $\mathbb{E}(X_i)$, $\mathbb{E}(|X_i|)$ и взвешивающую матрицу.

$$W = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

e) Найдите оптимальную теоретическую взвешивающую матрицу для обобщённого метода моментов

f) Постройте двухшаговую оценку обобщённого метода моментов, начав со взвешивающей матрицы W

g) С помощью полученных оценок постройте 95%-ый доверительный интервал для неизвестного параметра a

18.3 Винни-Пух и Пятачок оценивают неизвестный параметр правильности пчёл θ. Когда Винни-Пух проводит очередное измерение параметра правильности, он получает значение X_i нормально распределенное вокруг неизвестного параметра, $X_i \sim \mathcal{N}(\theta, 1)$. Когда Пятачок проводит измерение параметра правильности, он получает значение Y_i, также нормально распределенное вокруг θ, но имеющее большую дисперсию, $Y_i \sim \mathcal{N}(\theta, 4)$. Различные измерения независимы между собой.

a) Найдите $\mathbb{E}(X_i)$ и постройте соответствующую оценку метода моментов.

b) Найдите $\mathbb{E}(Y_i)$ и постройте соответствующую оценку метода моментов.

c) Используя два указанных момента найдите обобщённую оценку метода моментов для взвешивающей матрицы

$$W = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}.$$

d) Найдите оптимальную взвешивающую матрицу W.

18.4 Начинающий футболист делает независимые удары по воротам. С вероятностью θ он попадает левее ворот, с вероятностью 2θ — правее ворот и попадает с вероятностью $1 - 3\theta$. Из n ударов он попал N_L раз левее ворот и N_R раз — правее.

a) Найдите $\mathbb{E}(N_L)$ и постройте соответствующую оценку θ методом моментов.

b) Найдите $\mathbb{E}(N_R)$ и постройте соответствующую оценку θ методом моментов.

c) Используя два указанных момента постройте оценку обобщённого метода моментов со взвешивающей матрицей

$$W = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}.$$

d) Найдите оптимальную теоретическую взвешивающую матрицу.

e) Для каждой из найденных оценок постройте 95%-ый доверительный интервал, если $N_L = 10$, $N_R = 30$, $n = 200$.

18.5 Можно ли получить МНК-оценки в классической задаче регрессии как оценки обобщённого метода моментов? Можно ли получить оценки метода максимального правдоподобия как оценки обобщённого метода моментов?

18.6 Равшан и Джамшут измеряют длину θ оставшегося куска рулона обоев много раз. Измерения Равшана, X_i, распределены нормально, $\mathcal{N}(2\theta, \theta^2+100)$. Измерения Джамшута, Y_i, также распределены нормально $\mathcal{N}(\theta, \theta^2+10)$. Поскольку Равшан и Джамшут спорят друг с другом, их измерения зависимы, $\mathrm{Cov}(X_i, Y_i) = -1$.

Оказалось, что по 200 (?проверить?) измерениям $\sum X_i = 300$, $\sum Y_i = 100$.

Насяльника хочет измерить параметр θ.

a) Запишите два моментных условия на $\mathbb{E}(X_i)$ и $\mathbb{E}(Y_i)$ в виде

$$\begin{cases} \mathbb{E}(g_1(X_i, \theta)) = 0 \\ \mathbb{E}(g_2(Y_i, \theta)) = 0 \end{cases}$$

b) Найдите ковариационную матрицу $\mathrm{Var}(g)$ и теоретическую оптимальную матрицу весов W для обобщённого метода моментов;

c) Найдите оценку параметра θ методом моментов с единичной матрицей весов;

d) Найдите оценку параметра θ методом моментов, предварительно оценив оптимальную матрицу с помощью $\hat{\theta}$ из предыдущего пункта;

19 Панельки

Хорошо изложена теория у Kurt Schmidtheiny, https://www.schmidheiny.name/teaching/shortguides.htm.

19.1 Рассмотрим FE-модель $y_{it} = \alpha + \beta x_{it} + \gamma z_i + c_i + u_{it}$.

Есть всего два момента времени, $t = 1$ и $t = 2$. Визина использует within-оценку $\hat{\beta}_W$, а Федя — оценку в первых разностях, $\hat{\beta}_{FD}$.

a) Как связаны между собой оценки $\hat{\beta}_W$ и $\hat{\beta}_{FD}$?

b) Как связаны между собой $\mathrm{SS}^{\mathrm{res}}_W$ и $\mathrm{SS}^{\mathrm{res}}_{FD}$?

c) Являются ли оценки $\hat{\beta}_W$ и $\hat{\beta}_{FD}$ состоятельными?

d) Как построить 95%-ый доверительный интервал с помощью $\hat{\beta}_{FD}$ и $\hat{\beta}_W$ для β?

19.2 У храброй исследовательницы Аграфены были панельные данные за три периода, $t = 1$, $t = 2$, $t = 3$. По старинке для хранения данных Аграфена распечатала все данные на листочках. Резвый кот Борис безнадёжно испортил листок, относившийся к $t = 2$. И тогда Аграфена решила восстановить данные потерянного периода $t = 2$ с помощью линейной аппроксимации, $x_{i2} = (x_{i1} + x_{i3})/2$, $y_{i2} = (y_{i1} + y_{i3})/2$.

Успешно восстановив испорченные игривым котом Борисом данные, Аграфена приступила к анализу FE-модели

$$y_{it} = \alpha + \beta x_{it} + \gamma z_i + c_i + u_{it}.$$

Аграфена оценила коэффициент β с помощью within-оценки $\hat{\beta}_W$ и оценки в первых разностях, $\hat{\beta}_{FD}$.

a) Как связаны между собой оценки $\hat{\beta}_W$ и $\hat{\beta}_{FD}$?

b) Как связаны между собой $\mathrm{SS}^{\mathrm{res}}_W$ и $\mathrm{SS}^{\mathrm{res}}_{FD}$?

19.3 Рассмотрим модель $y_{it} = \alpha_i + \beta x_{it} + c_i + u_{it}$, оцениваемую по нескольким точкам:

a) Нарисуйте линию сквозной регрессию (pooled ols).

b) Нарисуйте данные в осях первых разностей и нарисуйте линию регрессии в первых разностях.

c) Нарисуйте данные в осях отклонений от среднего и нарисуйте линию within-регрессии.

20 Solutions

1.1.

a) $\hat{\beta} = 13/9$;

b)

c) $\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$;

d) $\hat{\beta}_1 = -1$, $\hat{\beta}_2 = 2$;

e)

f) $\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

1.2.

a) correct;

b) correct;

c) correct;

d) wrong, correct formula should be $\sum_{ij}(x_i - x_j)(y_i - y_j) = 2n\sum_i(x_i - \bar{x})(y_i - \bar{y})$;

e) wrong, correct formula should be $\sum_i(x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - n\bar{x}\bar{y}$.

1.3.

a) 0

b) 0

c) 0

d) $\sum x_i^2$

1.4.

a) $\hat{\theta} = \frac{\sum y_i(1+x_i)}{\sum(1+x_i)^2}$

b) $\hat{\theta} = \frac{\sum(y_i-1)x_i}{\sum x_i^2}$

c) $\hat{\theta} = \frac{\sum(y_i/x_i)}{\sum(1/x_i^2)}$

d) $\hat{\theta} = \sum\left((y_i - z_i)(x_i - z_i)\right) / \sum\left(x_i - z_i\right)^2$

1.5. $\hat{\alpha} = 0, \hat{\beta} = 1$

1.6. Рассмотрим регрессию суммы $(y_i + z_i)$ на саму себя. Естественно, в ней

$$\widehat{y_i + z_i} = 0 + 1 \cdot (y_i + z_i).$$

Отсюда получаем, что $\hat{\alpha} + \hat{\gamma} = 0$ и $\hat{\beta} + \hat{\delta} = 1$.

1.7.
Исходя из условия, нужно оценить методом МНК коэффициенты двух следующих моделей:

$$y_i = \alpha + \beta x_i + u_i$$

$$y_i = \frac{\gamma}{2} + \frac{\delta}{2} x_i + \frac{1}{2} v_i$$

Заметим, что на минимизацию суммы квадратов остатков коэффициент $1/2$ не влияет, следовательно:

$$\hat{\gamma} = 2\hat{\alpha}, \ \hat{\delta} = 2\hat{\beta}$$

1.8. Выпишем задачу:

$$\begin{cases} \mathrm{SS}^{\mathrm{res}} = \sum_{i=1}^n (y_i - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i)^2 \to \min_{\hat{\beta}_1, \hat{\beta}_2} \\ \hat{\beta}_1 + \hat{\beta}_2 = 1 \end{cases}$$

Можем превратить ее в задачу минимизации функции одного аргумента:

$$\mathrm{SS}^{\mathrm{res}} = \sum_{i=1}^n (y_i - x_i - \hat{\beta}_2 (z_i - x_i))^2 \to \min_{\hat{\beta}_2}$$

Выпишем условия первого порядка:

$$\frac{\partial \mathrm{SS}^{\mathrm{res}}}{\partial \hat{\beta}_2} = \sum_{i=1}^n 2(y_i - x_i - \hat{\beta}_2 (z_i - x_i))(x_i - z_i) = 0$$

Отсюда:

$$\sum_{i=1}^n (y_i - x_i)(x_i - z_i) + \hat{\beta}_2 \sum_{i=1}^n (z_i - x_i)^2 = 0 \Rightarrow \hat{\beta}_2 = \frac{\sum_{i=1}^n (y_i - x_i)(z_i - x_i)}{\sum_{i=1}^n (z_i - x_i)^2}$$

А $\hat{\beta}_1$ найдется из соотношения $\hat{\beta}_1 + \hat{\beta}_2 = 1$.

1.9. Обозначив вес первого слитка за β_1, вес второго слитка за β_2, а показания весов за y_i, получим, что

$$y_1 = \beta_1 + u_1, \ y_2 = \beta_2 + u_2, \ y_3 = \beta_1 + \beta_2 + u_3$$

Тогда

$$(300 - \beta_1)^2 + (200 - \beta_2)^2 + (400 - \beta_1 - \beta_2)^2 \to \min_{\beta_1, \ \beta_2}$$

$$\hat{\beta}_1 = \frac{800}{3}, \ \hat{\beta}_2 = \frac{500}{3}$$

1.10. Можем воспользоваться готовой формулой для регрессии на константу:

$$\hat{\beta} = \bar{y} = \frac{10 + 10 + 3}{3} = \frac{23}{3}$$

(можно решить задачу $2(10 - \beta)^2 + (3 - \beta)^2 \to \min$)

1.11.

a) Да.

b) Да.

c) Да.

d) Нет. Из условия первого порядка для первой выборки следует, что $\sum_A y_i = \hat{\beta}_1 + \hat{\beta}_2 \sum x_i$. Значит $\sum_A y_i > 0$. Аналогично, $\sum_B y_i > 0$, но $\sum y_i < 0$.

1.12.

1.13. Поскольку значения y остались теми же, $TSS_1 = TSS_2$.
Добавление ещё одного регрессора не уменьшит точность оценки, то есть $\mathrm{SS}_2^{\mathrm{res}} \leqslant \mathrm{SS}_1^{\mathrm{res}}$, $ESS_2 \geqslant ESS_1$.
Тогда и коэффициент детерминации $R^2 = ESS/TSS$ не уменьшится, то есть $R_2^2 \geqslant R_1^2$.

1.14. Пусть $\bar{y}$ — средний y до добавления нового наблюдения, $\bar{y}'$ — после добавления нового наблюдения. Будем считать, что изначально было n наблюдений. Заметим, что

$$\bar{y}' = \frac{(y_1 + \ldots + y_n) + y_{n+1}}{n+1} = \frac{n\bar{y} + y_{n+1}}{n+1} = \frac{n}{n+1}\bar{y} + \frac{1}{n+1}y_{n+1}$$

Покажем, что TSS может только увеличится при добавлении нового наблюдения (остается неизменным при $y_{n+1} = \bar{y}$):

$$TSS' = \sum_{i=1}^{n+1}(y_i - \bar{y}')^2 = \sum_{i=1}^{n}(y_i - \bar{y} + \bar{y} - \bar{y}')^2 + (y_{n+1} - \bar{y}')^2 =$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \bar{y}')^2 + (y_{n+1} - \bar{y}')^2 = TSS + \frac{n}{n+1}(y_{n+1} - \bar{y})^2$$

Следовательно, $TSS' \geqslant TSS$.

Также сумма $\mathrm{SS}^{\mathrm{res}}$ может только вырасти или остаться постоянной при добавлении нового наблюдения. Действительно, новое $(n+1)$-ое слагаемое в сумме неотрицательно. А сумма n слагаемых минимальна при старых коэффициентах, а не при новых.

ESS и R^2 могут меняться в обе стороны. Например, рассмотрим ситуацию, где точки лежат симметрично относительно некоторой горизонтальной прямой. При этом $ESS = 0$. Добавим наблюдение — ESS вырастет, удалим наблюдение — ESS вырастет.

1.15.

a) R^2 упал до нуля.

b) Да, можно. Если добавить точку далеко слева внизу от исходного набора данных, то наклон линии регрессии будет положительный. Если далеко справа внизу, то отрицательный. Будем двигать точку так, чтобы поймать нулевой наклон прямой. Получим $ESS = 0$.

1.16. На две неизвестных a и b нужно два уравнения. Эти два уравнения — ортогональность вектора остатков плоскости регрессоров. А именно:

$$\begin{cases} \sum_i (y_i - \hat{y}_i) = 0 \\ \sum_i (y_i - \hat{y}_i)\hat{y}_i = 0 \end{cases}$$

В нашем случае

$$\begin{cases} -1 + (6-a) + (6-b) = 0 \\ -1 + (6-a)a + (6-b)b = 0 \end{cases}$$

Решаем квадратное уравнение и получаем два решения: $a = 4$ и $a = 7$. Итого: $a = 4$, $b = 7$.

1.17. Обе ситуации возможны.

1.18.

a) нет, да

b) нет, нет

c) да, да, нет, нет

1.19. Replace factor variable with indicators for every its value. Using this predictors obtain forecasts of binary or quantitative variable. Calculate R^2 and take the square root.

1.20. График примерно симметричен относительно $x = y$. Линии регрессий примерно равны $\hat{y}_i \approx 0.5x_i$ и $\hat{x}_i \approx 0.5y_i$, что при выражении в обычных координатах даёт $y_i \approx 2\hat{x}_i$.

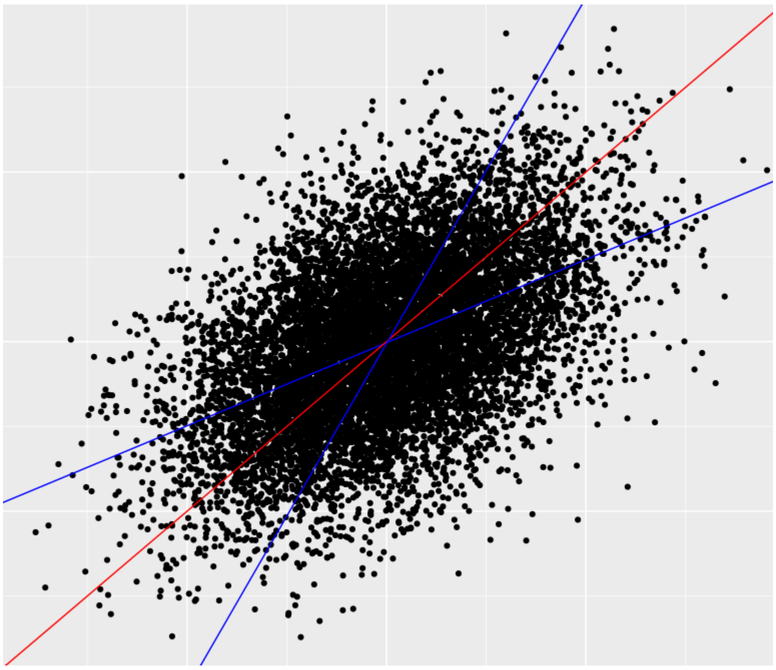
```
library(mvtnorm)
library(tidyverse)

V = matrix(c(1, 0.5, 0.5, 1), nrow = 2)
xy = rmvnorm(n = 10^4, mean = c(0, 0), sigma = V)
colnames(xy) = c('x', 'y')
xy_df = as_tibble(xy)

mod_yx = lm(data = xy_df, y ~ x)

mod_xy = lm(data = xy_df, x ~ y)
xy_slope = 1 / coef(mod_xy)[2]
xy_int = - coef(mod_xy)[1] / coef(mod_xy)[2]

qplot(data = xy_df, x = x, y = y) +
  geom_abline(slope = coef(mod_yx)[2], intercept = coef(mod_yx)[1], color =
  geom_abline(slope = 1, intercept = 0, color = 'red') +
  geom_abline(slope = xy_slope, intercept = xy_int, color = 'blue')
```



1.21.

1.22.

a) $\hat{\beta}_2 = \sum_{ij} w_{ij}\hat{\beta}_{2ij}$, где веса считаются по формуле $w_{ij} = q_{ij}/\sum_{ij} q_{ij}$.

b) $\hat{\beta}_1 = \sum_{ij} w_{ij}\hat{\beta}_{1ij}$.

2.1.

a) $f'(x) = 2x + 3$, $df = 2xdx + 3dx$, $df = 1.3$

b) $df = 2x_1dx_1 + 3dx_1 \cdot x_2^3 + 3x_1 \cdot 3x_2^2dx_2$, $df = 1.7$

2.2.

a) $A(dR)B$

b) $2r'dr$

c) $r'(A' + A)dr$

d) $-R^{-1} \cdot dR \cdot R^{-1}$

e) $-\sin(r'r) \cdot 2r'dr$

f) $\frac{r'(A'+A)dr \cdot r'r - r'Ar2r'dr}{(r'r)^2}$

2.3.

a) $dQ(\hat{\beta}) = 2(y - X\hat{\beta})^T(-X)d\hat{\beta}$, $d^2Q(\hat{\beta}) = 2d\hat{\beta}^T X^T X d\hat{\beta}$

b) $dQ(\hat{\beta}) = 0$

c) $\hat{\beta} = (X^TX)^{-1}X^Ty$

2.4.

a) $dQ(\hat\beta) = -2(y - X\hat\beta)^T X d\hat\beta + 2\lambda\hat\beta^T d\hat\beta$, $d^2Q(\hat\beta) = 2d\hat\beta^T(X^TX + \lambda I)d\hat\beta$

b) $dQ(\hat\beta) = 0$

c) $\hat\beta = (X^TX + \lambda I)^{-1}X^Ty$

2.5.

a) $\hat\mu = \sum y_i/n$

b)

c)

3.1.

$$\begin{cases}\hat\beta_1 \sum z_i^2 + \hat\beta_2 \sum x_i z_i = \sum z_i y_i \\ \hat\beta_1 \sum x_i z_i + \hat\beta_2 \sum x_i^2 = \sum x_i y_i\end{cases}$$

$$\begin{cases}\hat\beta_1 n + \hat\beta_2 \sum x_i = \sum y_i \\ \hat\beta_1 \sum x_i + \hat\beta_2 \sum x_i^2 = \sum x_i y_i\end{cases}$$

$$X'X\hat\beta = X'y$$

3.2.

a) $n = 5$

b) $k = 3$

c) $TSS = 10$

d) $\hat\beta = \begin{pmatrix}\hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3\end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix}2 \\ 2 \\ 1\end{pmatrix}$

e) $\hat y = X\hat\beta$

f) $\mathrm{SS}^{\mathrm{res}} = 2$

g) $R^2 = 1 - \frac{\mathrm{SS}^{\mathrm{res}}}{TSS} = 0.8$. R^2 высокий, построенная эконометрическая модель хорошо описывает данные

3.3. $\sum y_i^2 = \sum \hat y_i^2 + \sum \hat u_i^2$, $TSS = ESS + \mathrm{SS}^{\mathrm{res}}$,

3.4. $\mathrm{sCorr}(\hat y, y) = \frac{\mathrm{sCov}(\hat y, y)}{\sqrt{\mathrm{sVar}(\hat y)\,\mathrm{sVar}\,(y)}}$

$\mathrm{sCorr}(\hat y, y)^2 = \frac{(\mathrm{sCov}(\hat y, y))^2}{\mathrm{sVar}(\hat y)\,\mathrm{sVar}\,(y)}$

$R^2 \cdot TSS/(n-1) \cdot ESS/(n-1) = (\mathrm{sCov}(\hat y, y))^2 = (\mathrm{sCov}(\hat y - \bar y, y - \bar y))^2$ Отсюда можно понять, что ковариация для двухмерного случая равна произведению длин векторов $\hat y - \bar y$ и $y - \bar y - \sqrt{ESS}$ и $\sqrt{TSS}$ на косинус угла между ними ($\sqrt{R^2}$). Геометрически скалярное произведение можно изобразить как произведение длин одного из векторов на проекцию второго вектора на первый. Если будет проецировать $y - \bar y\vec{1}$ на $\hat y - \bar y\vec{1}$, то получим как раз ESS — тот квадрат на рисунке, что уже построен.

$\mathrm{sCov}(\hat y, y) = \sqrt{ESS^2/(n-1)^2} = ESS/(n-1)$

3.5. Спроецируем единичный столбец на «плоскость», обозначим его $1'$. Делаем проекцию y на «плоскость» и на $1'$. Далее аналогично.

3.6. Проекция y на $\hat{y}$ это $\hat{y}$, поэтому оценки коэффициентов будут 0 и 1. Оценка дисперсии $\frac{\text{SS}^{\text{res}}}{(n-2)ESS}$. Нарушены предпосылки теоремы Гаусса-Маркова, например, ошибки новой модели в сумме дают 0, значит коррелированы.

3.7. Либо в регрессию включена константа, либо единичный столбец (тут была опечатка, столбей) можно получить как линейную комбинацию регрессоров, например, включены дамми-переменные для каждого возможного значения качественной переменной.

3.8. Для удобства центрируем мысленно все переменные. Это не меняет ни корреляций, ни выборочных дисперсий, ни угловых коэффициентов в регрессиях. В регрессиях при этом оценка коэффициента при константе превращается в ноль, но какое нам до неё дело? :) Зато при нулевом среднем выборочные корреляции превратились в косинус угла между векторами:

$$\text{sCorr}(x,y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \cos(x,y)$$

И при нулевом среднем выборочная дисперсия — это длина вектора с точностью до умножения на $(n-1)$:

$$\text{sVar}(x) = \frac{\sum x_i^2}{n-1} = \frac{\|x\|^2}{n-1}$$

Начать можно с геометрического смысла оценок МНК:

$$\begin{cases} 1.4 = \frac{\|y\|}{\|x\|}\cos(x,y) \\ 0.6 = \frac{\|x\|}{\|y\|}\cos(x,y) \end{cases}$$

Отсюда находим $\|y\|/\|x\| = \hat{\sigma}_y/\hat{\sigma}_x$ и $\cos(x,y) = \text{sCorr}(x,y)$

Дальше можно решить по теореме косинусов.

a) $\text{sCorr}(x,y) = \sqrt{0.84}$, $\text{sCorr}(y,z) = \frac{\sqrt{70}}{10}$, $\text{sCorr}(x,z) = -\frac{\sqrt{30}}{10}$

b) $\frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} = \frac{3}{7}$, $\frac{\hat{\sigma}_z^2}{\hat{\sigma}_y^2} = \frac{8}{35}$, $\frac{\hat{\sigma}_x^2}{\hat{\sigma}_z^2} = \frac{15}{8}$

3.9. all of them :)

4.1. $\dim V = 1$, $\dim W = 2$, $\dim V \cap W = 0$, $\dim V^{\perp} = 2$, $\dim W^{\perp} = 1$. Эти же числа и будут степенями свободы хи-квадрат распределения.

4.2. $\dim V = 1$, $\dim W = n-1$, $\dim V \cap W = 0$, $\dim V^{\perp} = n-1$, $\dim W^{\perp} = 1$.

4.3.

4.4.

4.5. Сферы с центром в начале координат. Проекция имеет хи-квадрат распределение с тремя степенями свободы. Для нахождения максимальной вероятности максимизируем функцию

$$\exp(-R^2/2) \cdot ((R+t)^3 - R^3) \to \max_R$$

, где R — радиус мякоти, а t — толщина кожуры апельсина. Оставляем только линейную часть по t и затем максимизируем.

Наибольшая вероятность попасть в апельсин радиуса $R = 1$.

5.1.

5.2.

5.3.

a) $\operatorname{Var}(u_1) = \operatorname{Var}(u)_{(1,1)} = 4 \cdot I_{(1,1)} = 4$

b) $\operatorname{Var}(\beta_1) = 0$, так как β_1 — детерминированная величина.

c) $\operatorname{Var}(\hat{\beta}_1) = \sigma^2 (X'X)^{-1}_{(1,1)} = 0.5\sigma^2 = 0.5 \cdot 4 = 2$

d) $\widehat{\operatorname{Var}}(\hat{\beta}_1) = \hat{\sigma}^2 (X'X)^{-1}_{(1,1)} = 0.5\hat{\sigma}^2_{(1,1)} = 0.5\frac{\mathrm{SS}^{\mathrm{res}}}{5-3} = 0.25\mathrm{SS}^{\mathrm{res}} = 0.25y'(I - X(X'X)^{-1}X')y = 0.25 \cdot 1 = 0.25$
$\hat{\sigma}^2 = \frac{\mathrm{SS}^{\mathrm{res}}}{n-k} = \frac{1}{2}$.

e) Так как оценки МНК являются несмещёнными, то $\mathbb{E}(\hat{\beta}) = \beta$, значит:

$$\mathbb{E}(\hat{\beta}_1) - \beta_1^2 = \mathbb{E}(\hat{\beta}_1) - (\mathbb{E}(\hat{\beta}_1))^2 = \widehat{\operatorname{Var}}(\hat{\beta}_1) = 0.25$$

f) $\operatorname{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2 (X'X)^{-1}_{(2,3)} = 4 \cdot \left(-\frac{1}{2}\right) = -2$

g) $\widehat{\operatorname{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = \widehat{\operatorname{Var}}(\hat{\beta})_{(2,3)} = \hat{\sigma}^2 (X'X)^{-1}_{(2,3)} = \frac{1}{2} \cdot \left(-\frac{1}{2}\right) = -\frac{1}{4}$

h) $\operatorname{Var}(\hat{\beta}_2 - \hat{\beta}_3) = \operatorname{Var}(\hat{\beta}_2) + \operatorname{Var}(\hat{\beta}_3) + 2\operatorname{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)} = 4(1 + 1.5 + 2 \cdot (-0.5)) = 6$

i) $\widehat{\operatorname{Var}}(\hat{\beta}_2 - \hat{\beta}_3) = \widehat{\operatorname{Var}}(\hat{\beta}_2) + \widehat{\operatorname{Var}}(\hat{\beta}_3) + 2\widehat{\operatorname{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)} = \frac{1}{2} \cdot 1.5 = 0.75$

j) $\operatorname{Var}(\beta_2 - \beta_3) = 0$

k) $\operatorname{Corr}(\hat{\beta}_2, \hat{\beta}_3) = \frac{\operatorname{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\operatorname{Var}(\hat{\beta}_2)\operatorname{Var}(\hat{\beta}_3)}} = \frac{-2}{\sqrt{4 \cdot 6}} = -\frac{\sqrt{6}}{6}$

l) $\widehat{\operatorname{Corr}}(\beta_2, \beta_3) = \frac{\widehat{\operatorname{Cov}}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\widehat{\operatorname{Var}}(\hat{\beta}_2)\widehat{\operatorname{Var}}(\hat{\beta}_3)}} = \frac{-\frac{1}{4}}{\sqrt{\frac{1}{2} \cdot \frac{3}{4}}} = -\frac{\sqrt{6}}{6}$

m) $(n-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$.

$$\mathbb{E}\left((n-k)\frac{\hat{\sigma}^2}{\sigma^2}\right) = n - k$$

$$\mathbb{E}\left(\frac{\hat{\sigma}^2}{2}\right) = 1$$

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

n) $\hat{\sigma}^2 = \frac{\mathrm{SS}^{\mathrm{res}}}{n-k} = \frac{1}{2}$

5.4. На 40 наблюдений 15 переменных — это явная переподгонка. Асимптотических свойств МНК мы явно не можем использовать. Априори разумно предполагать гетероскедастичность, с которой на 40 наблюдениях никак не поборешься. Для использования робастных стандартных ошибок слишком мало наблюдений. Цель Аполлона слишком размыта. Правильнее было уточнить: построить модель, чтобы прогнозировать, или построить модель, чтобы проверить, влияет ли регрессор z на зависимую переменную.

Вероятно, самое разумное, это применить LASSO выбрав параметр регуляризации так, чтобы осталось буквально два-три регрессора.

5.5.

a) $\hat{\beta}_{\text{ols}} = (y_1 + 2y_2)/5$;

b) $\text{Var}(\hat{\beta}_{\text{ols}} \mid x) = 1/5$;

c) Заметим, что по величине $2y_1 - y_2$ можно однозначно восстановить величины ошибок u_1 и u_2. Например, если $2y_1 - y_2 = 3$, то $u_1 = 1$, $u_2 = -1$.

$$\hat{\beta}_{\text{best}} = \begin{cases} y_1 + 1, & \text{если } 2y_1 - y_2 < 0, \\ y_1 - 1, & \text{если } 2y_1 - y_2 > 0. \end{cases}$$

d) Unexpectedly, $\text{Var}(\hat{\beta}_{\text{best}} \mid x) = 0$.

e) Построенная оценка $\hat{\beta}_{\text{best}}$ является нелинейной по y, а теорема Гаусса — Маркова гарантирует только, что метод наименьших квадратов порождает несмещённую оценку с наименьшей дисперсией среди линейных по y оценок.

5.6.

a) $\alpha = 0.05$;

b) $\alpha = 0.05$;

5.7.

a) $\alpha = 0.05$;

5.8.

a) The estimator $\hat{\beta}_2^A$ is unbiased and consistent;

b) The estimator $\hat{\beta}_2^B$ is unbiased and unconsistent. Intuition behind unconsistency: the first observation has high influence even for large n. One may formally consider a particular case as a counter-example to consistency. Let x_i be uniform on $[0; 1]$, let u_i be equiprobable $+1$ or -1. The jump in $\hat{\beta}_2^B$ caused by two values of u_1 will not go to zero when n tends to infinity.

5.9.

a) Error covariance matrix will contain blocks B on the main diagonal

$$\mathrm{Var}(u) = \begin{pmatrix} B & 0 & 0 & \dots \\ 0 & B & 0 & \dots \\ 0 & 0 & B & \dots \\ \dots & & & \end{pmatrix},$$

where each block is equal to $B = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix}$.

b) Error covariance matrix will consist of four blocks: the top left is proportional to the identity matrix, bottom right block will have all elements equal to σ^2 and two other blocks will be zero:

$$\mathrm{Var}(u) = \begin{pmatrix} \sigma^2 \cdot I & 0 \\ 0 & S \end{pmatrix},$$

where I is an identity matrix, and all $S_{ij} = \sigma^2$.

6.1.

6.2. $F = t^2$;

6.3.

6.4.

7.1.

7.2.

7.3.

7.4.

8.1.

a) $M = \begin{pmatrix} AC + BE & AD + BF \end{pmatrix}$

Предположим, что матрицы A и B имеют m строк. Тогда размерность блока $AC + BE - m \times p$, $AD + BF - m \times q$.

b) $M = \begin{pmatrix} CA + DB \\ EA + FB \end{pmatrix}$

Предположим, что матрицы A и B имеют n столбцов. Тогда размерность блока $CA + DB - p \times n$, $EA + FB - q \times n$.

c) $M = \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix}$

Блок A^T имеет размерность $p \times q$, $B^T - q \times q$, $C^T - p \times p$, $D^T - q \times p$.

8.2.

a) $M^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$

b) $M^{-1} = \begin{pmatrix} 0 & B^{-1} \\ A^{-1} & 0 \end{pmatrix}$

c) $M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}CB^{-1} \\ 0 & B^{-1} \end{pmatrix}$

d) $M^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -B^{-1}CA^{-1} & B^{-1} \end{pmatrix}$

8.3.

a)

$$\left(\begin{array}{cc|cc} A & C & I & 0 \\ D & B & 0 & I \end{array}\right) \sim \left(\begin{array}{cc|cc} I & A^{-1}C & A^{-1} & 0 \\ D & B & 0 & I \end{array}\right) \sim \left(\begin{array}{cc|cc} I & A^{-1}C & A^{-1} & 0 \\ 0 & B - DA^{-1}C & -DA^{-1} & I \end{array}\right) \sim$$
$$\left(\begin{array}{cc|cc} I & A^{-1}C & A^{-1} & 0 \\ 0 & I & -(B - DA^{-1}C)^{-1}DA^{-1} & (B - DA^{-1}C)^{-1} \end{array}\right) \sim$$
$$\left(\begin{array}{cc|cc} I & 0 & A^{-1} + A^{-1}C(B - DA^{-1}C)^{-1}DA^{-1} & -A^{-1}C(B - DA^{-1}C)^{-1} \\ 0 & I & -(B - DA^{-1}C)^{-1}DA^{-1} & (B - DA^{-1}C)^{-1} \end{array}\right)$$

То есть $X = A^{-1} + A^{-1}C(B - DA^{-1}C)^{-1}DA^{-1}$.

b) Из равенства

$$\begin{pmatrix} A & C \\ D & B \end{pmatrix}\begin{pmatrix} X & Z \\ Y & W \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

получаем систему:

$$\begin{cases} AX + CY = I \\ DX + BY = 0 \end{cases} \Rightarrow \begin{cases} X = A^{-1}(I - CY) \\ DX + BY = 0 \end{cases}$$

Подставляя первое уравнение во второе, получим:

$$DA^{-1}(I - CY) = -BY \Rightarrow DA^{-1} = (DA^{-1}C - B)Y \Rightarrow I = (C - AD^{-1}B)Y \Rightarrow Y = (C - AD^{-1}B)^{-1}$$

И окончательно из второго уравнения:

$$DX = -B(C - AD^{-1}B)^{-1} \Rightarrow -(C - AD^{-1}B)B^{-1}DX = I \Rightarrow X = (A - CB^{-1}D)^{-1}$$

c)

$$\begin{aligned} &(A - CB^{-1}D)(A^{-1} + A^{-1}C(B - DA^{-1}C)^{-1}DA^{-1}) = \\ &\qquad I - CB^{-1}DA^{-1} + (C - CB^{-1}DA^{-1}C)(B - DA^{-1}C)^{-1}DA^{-1} = \\ &\qquad I - CB^{-1}DA^{-1} + CB^{-1}(B - DA^{-1}C)(B - DA^{-1}C)^{-1}DA^{-1} = \\ &\qquad\qquad I - CB^{-1}DA^{-1} + CB^{-1}DA^{-1} = I \end{aligned}$$

8.4.

9.1.

a) $\hat{\beta} = (X^TX + \lambda I)^{-1}X^Ty$;

b) $\lim_{\lambda\to\infty}\hat{\beta} = (X^TX)^{-1}X^Ty = 0$;

c) $\lim_{\lambda\to 0}\hat{\beta} = (X^TX)^{-1}X^Ty = \hat{\beta}_{\text{ols}}$;

d) $\mathbb{E}(\hat\beta \mid X) = (X^TX + \lambda I)^{-1}X^TX\beta$;

e) $\mathrm{Var}(\hat\beta \mid X) = \sigma^2(X^TX + \lambda I)^{-1}X^TX(X^TX + \lambda I)^{-1}$;

f) If $\lambda \to \infty$ then $\mathbb{E}(\hat\beta \mid X) \to 0$ and $\mathrm{Var}(\hat\beta \mid X) \to 0$;

g) If $\lambda \to 0$ then $\mathbb{E}(\hat\beta \mid X) \to \beta$ and $\mathrm{Var}(\hat\beta \mid X) \to \sigma^2(X^TX)^{-1}$;

9.2. One can add $y^+ = 0$, $X^+ = \sqrt{\lambda}I$.

9.3.

9.4.

a) $R_{ab} = \pm 3/\sqrt{10}$;

b)

$$(X^TX)^{-1} = \frac{1}{1-R^2_{ab}}\begin{pmatrix} 1/\mathrm{SST}_a & -R_{ab}/\sqrt{\mathrm{SST}_a\mathrm{SST}_b} \\ -R_{ab}/\sqrt{\mathrm{SST}_a\mathrm{SST}_b} & 1/\mathrm{SST}_b \end{pmatrix}$$

9.5.

a) $\hat\beta_1 = \hat\beta_2 = 0.5(x^Tx + \lambda/2)^{-1}x^Ty$;

b) $\hat\beta_j \to 0$;

c) $\hat\beta_j \to \hat\beta_{\text{ols}}/2$, $\hat y = \hat\beta_{\text{ols}}x$.

9.6.

9.7.

9.8.

10.1.

10.2.

10.3. $(Ux)^TUx = x^TU^TUx = x^Tx$, $(Ux)^TUy = x^TU^TUy = x^Ty$.

10.4. $\begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt2 & 1/\sqrt2 \\ 1/\sqrt2 & -1/\sqrt2 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt6 & 2/\sqrt6 & 1/\sqrt6 \\ 2/\sqrt5 & -1/\sqrt5 & 0 \\ 1/\sqrt{30} & 2/\sqrt{30} & -5/\sqrt{30} \end{pmatrix}$

$$X'X = \begin{pmatrix} 1/\sqrt2 & -1/\sqrt2 \\ 1/\sqrt2 & 1/\sqrt2 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1/\sqrt2 & 1/\sqrt2 \\ 1/\sqrt2 & -1/\sqrt2 \end{pmatrix}$$

$$XX' = \begin{pmatrix} 2/\sqrt6 & 0 & 1/\sqrt3 \\ 1/\sqrt6 & 1/\sqrt2 & -1/\sqrt3 \\ -1/\sqrt6 & 1/\sqrt2 & 1/\sqrt3 \end{pmatrix}\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 2/\sqrt6 & 1/\sqrt6 & -1/\sqrt6 \\ 0 & 1/\sqrt2 & 1/\sqrt2 \\ 1/\sqrt3 & -1/\sqrt3 & 1/\sqrt3 \end{pmatrix}$$

$$X = \begin{pmatrix} 2/\sqrt6 & 0 & 1/\sqrt3 \\ 1/\sqrt6 & 1/\sqrt2 & -1/\sqrt3 \\ -1/\sqrt6 & 1/\sqrt2 & 1/\sqrt3 \end{pmatrix}\begin{pmatrix} \sqrt3 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt2 & 1/\sqrt2 \\ -1/\sqrt2 & 1/\sqrt2 \end{pmatrix}$$

10.5.

10.6.

10.7.

10.8.

10.9.

10.10.

10.11.

11.1.

$$\mathbb{E}(y_i(1) - y_i(0)) = \mathbb{E}(y_i \mid w_i = 1) - \mathbb{E}(y_i \mid w_i = 0)$$

11.2.

11.3.

a) $\hat{\mu} = \hat{\alpha}$

b) $\hat{\delta}_c = \hat{\delta}_d$

c) $\mathrm{SS}_1^{\mathrm{res}} = \mathrm{SS}_2^{\mathrm{res}}$

d) $se(\hat{\delta}_c) \neq se(\hat{\delta}_d)$

12.1.

12.2.

a) $\ell = n \ln \lambda - \lambda \sum y_i$

b) $\hat{\lambda} = \frac{\sum y_i}{n} = \frac{1}{2}$

c) $I(\lambda) = \frac{n}{\lambda^2}$

d) $LR = 2(n \ln \frac{n}{\sum y_i} - n - n \ln \lambda_R + \lambda_R \sum y_i)$

$LM = \left(\frac{n}{\lambda} - \sum y_i\right)^2 \frac{\lambda^2}{n}$

$W = \left(\frac{\sum y_i}{n} - \lambda_R\right)^2 \frac{n}{\lambda^2}$

e) $LR \approx 61.37$, $LM = W = 100$

f) $\chi^2_{1,0.95} = 3.84$, основная гипотеза отвергается.

12.3.

a) $\ell = n \ln \sqrt{2\pi} - \frac{1}{2} \sum (y_i - \beta x_i)^2$

b) $\hat{\beta}_{ML} = \frac{\sum y_i x_i}{\sum x_i^2} = 2.5$

c) $I(\beta) = \sum x_i^2$

d) $LR = -\sum(y_i - \hat{\beta}_{ML}x_i)^2 + \sum(y_i - \beta_R x_i)^2$

$LM = (\sum(y_i x_i - \beta_R x_i^2))^2 \cdot \frac{1}{\sum x_i^2}$

$W = (\hat{\beta}_{ML} - \beta_R)^2 \sum x_i^2$

e) $LR = LM = W = 625$

f) $\chi^2_{1,0.95} = 3.84$, основная гипотеза отвергается.

12.4.

a) $\ell = const + y_1 \ln p_1 + y_2 \ln p_2 + y_3 \ln p_3 + (n - y_1 - y_2 - y_3) \ln(1 - y_1 - y_2 - y_3)$

b) $\hat{p} = \begin{pmatrix} y_1/n \\ y_2/n \\ y_3/n \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.3 \\ 0.2 \end{pmatrix}$

c) $I(p) = \begin{pmatrix} \frac{n}{p_1} + \frac{n}{1-p_1-p_2-p_3} & \frac{n}{1-p_1-p_2-p_3} & \frac{n}{1-p_1-p_2-p_3} \\ \frac{n}{1-p_1-p_2-p_3} & \frac{n}{p_2} + \frac{n}{1-p_1-p_2-p_3} & \frac{n}{1-p_1-p_2-p_3} \\ \frac{n}{1-p_1-p_2-p_3} & \frac{n}{1-p_1-p_2-p_3} & \frac{n}{p_3} + \frac{n}{1-p_1-p_2-p_3} \end{pmatrix}$

d) $I^{-1}(p) = \begin{pmatrix} \frac{p_1(1-p_1)}{n} & -\frac{p_1 p_2}{n} & -\frac{p_1 p_3}{n} \\ -\frac{p_1 p_2}{n} & \frac{p_2(1-p_2)}{n} & -\frac{p_2 p_3}{n} \\ -\frac{p_1 p_3}{n} & -\frac{p_2 p_3}{n} & \frac{p_3(1-p_3)}{n} \end{pmatrix}$

12.5.

a) $\hat{\theta}^{UR} = h(y)$

b) $\theta^R = \begin{pmatrix} \theta_1^0 \\ h_2(y) - C^{-1}B^T(\theta_1^0 - h_1(y)) \end{pmatrix}$

3—6. $LR = LM = W = (\theta_1^0 - h_1(y))^T(A - BC^{-1}B^T)(\theta_1^0 - h_1(y))$

12.6.

2. $Q = \frac{1}{\sigma^2}X^T X, h(y) = (X^T X)^{-1}X^T y$

12.7.

12.8.

12.9. Мы знаем, что $(SS_R^{\text{res}} - SS_{UR}^{\text{res}}) \cdot (60 - 10)/5SS_{UR}^{\text{res}}$ имеет в точности F-распределение. Находим критическое значение для него по таблице. Выражаем три статистики через F-распределение. Получаем точные критические значения.

13.1.

a) $\hat{\beta}_{OLS} = 11/9$

b) $se(\hat{\beta}) = \sqrt{5/162}$

c) $se_{HC0}(\hat{\beta}) = \sqrt{168}/81, se_{HC3}(\hat{\beta}) = \sqrt{2649}/180$

d) $\hat{\beta} = 7/6$

13.2. $\text{Var}(u_i) = cx_i^4$

13.3. Протестируем гетероскедастичность ошибок при помощи теста Голдфельда- Квандта. H_0 : $\text{Var}(u_i) = \sigma^2$, H_a : $\text{Var}(u_i) = f(x_i)$

a) Тестовая статистика $GQ = \frac{\text{SS}_3^{\text{res}}/(n_3-k)}{\text{SS}_1^{\text{res}}/(n_1-k)}$, где $n_1 = 11$ — число наблюдений в первой подгруппе, $n_3 = 11$ — число наблюдений в последней подгруппе, $k = 3$ — число факторов в модели, считая единичный столбец.

b) Распределение тестовой статистики при верной H_0: $GQ \sim F_{n_3-k,n_1-k}$

c) Наблюдаемое значение $GQ_{obs} = 1.41$

d) Область, в которой H_0 не отвергается: $GQ \in [0; 3.44]$

e) Статистический вывод: поскольку $GQ_{obs} \in [0; 3.44]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 не может быть отвергнута. Таким образом, тест Голдфельда-Квандта не выявил гетероскедастичность.

13.4. Протестируем гетероскедастичность ошибок при помощи теста Уайта. H_0 : $\text{Var}(u_i) = \sigma^2$, H_a : $\text{Var}(u_i) = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i$.

a) Тестовая статистика $W = n \cdot R^2_{aux}$, где n — число наблюдений, R^2_{aux} — коэффициент детерминации для вспомогательной регрессии.

b) Распределение тестовой статистики при верной H_0: $W \sim \chi^2_{k_{aux}-1}$, где $k_{aux} = 6$ — число регрессоров во вспомогательной регрессии, считая константу.

c) Наблюдаемое значение тестовой статистики: $W_{obs} = 18$

d) Область, в которой H_0 не отвергается: $W \in [0; W_{crit}] = [0; 11.07]$

e) Статистический вывод: поскольку $W_{obs} \notin [0; 11.07]$, то на основании имеющихся наблюдений на уровне значимости 5% основная гипотеза H_0 отвергается. Таким образом, тест Уайта выявил гетероскедастичность.

13.5. $k(k+1)/2$

13.6. 0.0752, 5, 10

13.7.

13.8.

13.9. Одинаковые.

13.10.

$$\widehat{\text{Var}}(\hat{\beta}_1) = TSS\frac{1}{n_0}\frac{1}{n-2}$$

$$\widehat{\text{Var}}(\hat{\beta}_1) = TSS_0\frac{n}{n_0}\frac{1}{n_0}\frac{1}{n-2}$$

13.11. В предположении о гомоскедастичности, $\gamma_2 = 0$, оценка правдоподобия совпадает с МНК-оценкой, значит $\hat{\beta} = \sum y_i x_i / \sum x_i^2$. И $\hat{\sigma}_i^2 = \text{SS}^{\text{res}}/n$, значит $\hat{\gamma}_1 = \ln(\text{SS}^{\text{res}}/n)$.

13.12.

$$\begin{aligned}\operatorname{Cov}(\hat\beta_{GLS}, u) = \operatorname{Cov}\left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y, u\right) = \\ = \operatorname{Cov}\left((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon, u\right) = \\ = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\operatorname{Cov}(\varepsilon, u) = \\ = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\Sigma = (X'\Sigma^{-1}X)^{-1}X'\end{aligned}$$

13.13.

14.1. $\hat{\mathbb{P}}(y_i = 1 \mid x_i) = \frac{1}{1+\exp(-\beta_1-\beta_2 x_i)}$

a) $loss(\beta_1, \beta_2) = -\sum_{i=1}^{l}\left([y_i = 1]\ln\frac{1}{1+\exp(-\beta_1-\beta_2 x_i)} + [y_i = -1]\ln\left(1 - \frac{1}{1+\exp(-\beta_1-\beta_2 x_i)}\right)\right)$

b) $\frac{\partial loss}{\partial \beta_1} = -\sum_{i=1}^{l}\left([y_i = 1]\cdot\frac{1}{1+\exp(\beta_1+\beta_2 x_i)} + [y_i = -1]\cdot(-1)\cdot\frac{1}{1+\exp(-\beta_1-\beta_2 x_i)}\right)$

c) $y_4 = 1$, $x_4 = 0.8$

14.2.

a) $\Lambda(w) + \Lambda(-w) = 1$

b) $\Lambda'(w) = -\Lambda'(-w)$

c)

d) $\Lambda(0) = 0.5$, $\Lambda'(0) = 0.25$, $\ln\Lambda(0) = -\ln 2$

e) $\Lambda^{-1}(p) = \ln\frac{p}{1-p}$

f) $\frac{d\ln\Lambda(w)}{dw} = \Lambda(-w)$

g) $\frac{d\ln\Lambda(-w)}{dw} = -\Lambda(w)$

h)

14.3.

a) Выпишем аппроксимацию функции потерь:

$$\operatorname{loss}(\beta_1, \beta_2) \approx 100\ln 2 + 6\beta_1 + 12\beta_2 + \frac{1}{2}(25\beta_1^2 + 2\cdot 12\beta_1\beta_2 + 12\beta_2^2) \to \min_{\beta_1,\beta_2}$$

Взяв производные по β_1 и β_2, получим $\hat\beta_1 = \frac{6}{13}$, $\hat\beta_2 = -\frac{19}{13}$.

b) $\hat{\mathbb{P}}(\text{honey}_i = 1 \mid \text{bee}_i = 0) = \frac{1}{1+\exp(-6/13)} \approx 0.615$.

Это же число можно было получить из таблицы: $\frac{32}{32+20} \approx 0.61$.

14.4. Предельный эффект максимален при максимальной производной $\Lambda'(\hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z)$, то есть при $\hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z = 0$.

14.5. Ввести штраф в жанре LASSO или гребневой регрессии.

14.6. $z = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{3}{0.3} = 10$, H_0 отвергается. Предельный эффект равен $\hat{\beta}_2 \Lambda'(-0.8) \approx 0.642$. Для нахождения $se(\hat{\mathbb{P}})$ найдём линейную аппроксимацию для $\Lambda(\hat{\beta}_1 + \hat{\beta}_2 x)$ в окрестности точки $\hat{\beta}_1 = 0.7$, $\hat{\beta}_2 = 3$. Получаем

$$\Lambda(\hat{\beta}_1 + \hat{\beta}_2 x) \approx \Lambda(\beta_1 + \beta_2 x) + \Lambda'(\beta_1 + \beta_2 x)(\hat{\beta}_1 - \beta_1) + \Lambda'(\beta_1 + \beta_2 x) x (\hat{\beta}_2 - \beta_2).$$

14.7. Если в пробит-уравнении ненаблюдаемой переменной домножить все коэффициенты и стандартную ошибку на произвольную константу, то в результате получится ровно та же модель. Следовательно, модель с $u_i \sim \mathcal{N}(0; \sigma^2)$ не идентифицируема. Поэтому надо взять какое-то нормировочное условие. Можно взять, например, $\beta_2 = 42$, но традиционно берут $u_i \sim \mathcal{N}(0; 1)$.

14.8.

Если посчитать ожидание матрицы Гессе, то оно не зависит от игреков. А значит информация Фишера не зависит от игреков. Поэтому для подсчёта изменения стандартных ошибок достаточно подставить новые оценки вместо истинных значений параметра. Можно также воспользоваться тем, что новые оценки являются линейным преобразованием старых.

14.9.

14.10. $F(t) = 1/(1 + \exp(-t))$.

14.11. Они абсолютно равны. Фактически они просто разбили исходную выборку на части и в качестве прогноза для каждой части наблюдений берут долю игреков, равных единице на этой части выборки.

15.1.

15.2.

15.3.

15.4.

15.5.

15.6.

15.7.

15.8.

16.1.

16.2.

16.3.

16.4.

a) $\operatorname{plim} \hat{\beta}_2^{LS} = \beta_2 - 1/5$

b) $\operatorname{plim} \hat{\beta}_2^{IV} = \beta_2$

c)

16.5.

16.6.

$$\operatorname{plim} \hat{\beta}_2 = \frac{\operatorname{Cov}(y_1, x_1^*)}{\operatorname{Var}(x_1^*)} = \frac{9}{14}\beta_2$$

The estimator $\hat{\beta}_2$ is biased and unconsistent.

16.7.

$$\operatorname{plim} \hat{\beta}_2 = \frac{\operatorname{Cov}(x_1, y_1)}{\operatorname{Var}(x_1)} = \frac{\beta_2 \cdot 9 + \beta_3(-2) + 0}{9} = \beta_2 - \frac{2}{9}\beta_3 \neq \beta_2$$

Оценка несостоятельна.

16.8. Для $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^3 + w_i$. Оценка $\hat{\beta}_2$ несостоятельна, $\operatorname{plim} \hat{\beta}_2 = \beta_2 + 3\beta_3$, а $\hat{\beta}_1$ состоятельная.

Трудность проверки несмещённости состоит в зависимости числителя и знаменателя в формуле условного ожидания.

Для $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + w_i$. Обе оценки будут состоятельными. Вызвано это тем, что $u_i = \beta_3 x_i^2 + w_i$ некоррелировано с x_i.

16.9. $\operatorname{plim} \left(\frac{1}{n}x'x\right)^{-1} = 109^{-1}$, $\operatorname{plim} \frac{1}{n}x'u = 0$ и $\operatorname{plim}(x'x)^{-1}x'u = 0$

16.10. Да, например, равномерное распределение (u_i, x_i) на круге или на окружности. Или равновероятное на восьми точках, $(\pm 1, \pm 1)$, $(\pm 2, \pm 2)$.

16.11.

$\operatorname{plim} \hat{\theta}_2 = \frac{\gamma}{\alpha(1-\alpha)}$

Заметим, что $\operatorname{Cov}(green_i, green_i^*) = 0$, так как $green_i$ — это результат подбрасывания монетки, $green_i^*$ определяется характеристиками студента.

Ожидание такое же, хотя считается по-другому, $\mathbb{E}(\hat{\theta}_2) = \frac{\gamma}{\alpha(1-\alpha)}$.

16.12.

a) Например, можно взять $u_1 = x_2$, и все величины $\mathcal{N}(0; 1)$.

b) Невозможно, в силу $\mathbb{E}(u_1 \mid x_1) = \mathbb{E}(\mathbb{E}(u_1 \mid x_1, x_2) \mid x_1)$.

16.13. Одинаково распределены: $y_1 \sim y_2 \sim y_3$, $x_1 \sim x_2 \sim x_3$. Независимы переменные с разными номерами.

16.14.

17.1. $\operatorname{plim} \hat{\alpha} = (\alpha + \beta)/2$

Переменную u_{t2} можно использовать как инструмент для p_t. При известной u_{t2} величина β восстанавливается идеально точно.

17.2. В первом уравнении нет проблемы эндогенности, оценка $\hat{\beta}_{11}$ будет состоятельной.

17.3. Без дополнительных ограничений уже по критерию порядка система не идентифицируема. Поэтому без ограничений критерий ранга можно даже не проверять. При наложении ограничений первое уравнение остаётся неидентифицируемым, второе — становится идентифицируемым.

17.4.

a) Условие порядка о том, что количество не включённых в правую часть уравнения переменных должно быть не меньше числа включённых эндогенных, выполняется только для второго уравнения системы. Остаётся проверить для второго уравнения условие ранга, оно также окажется выполнено в точке общего положения.

   Нарушается условие ранга если

$$\begin{cases} \gamma_{12} = 0 \\ \beta_{13}\gamma_{33} = 0 \end{cases}$$

b) К жёсткой мультиколлинеарности на втором шаге.

   На первом шаге будут построены регрессии:

$$\hat{y}_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{\alpha}_3 x_{3t}$$
$$\hat{y}_{3t} = \hat{\delta}_0 + \hat{\delta}_1 x_{1t} + \hat{\delta}_2 x_{2t} + \hat{\delta}_3 x_{3t}$$

   На втором шаге строим регрессию

$$\hat{y}_{1t} = \hat{\gamma}_{01} + \hat{\beta}_{12}\hat{y}_{2t} + \hat{\beta}_{13}\hat{y}_{3t} + \hat{\gamma}_{11} x_{1t} + \hat{\gamma}_{12} x_{12t}$$

   Однако $\hat{y}_{2t}$ и $\hat{y}_{3t}$ являются линейными комбинациями x_{1t}, x_{2t}, x_{3t}, а значит, в регрессии второго шага есть линейно зависимые регрессоры. Поэтому МНК-оценки получить нельзя.

17.5.

a) На втором шаге возникнет проблема жёсткой мультиколлинеарности.

b) На втором шаге жёсткой мультиколлинеарности не будет, наблюдений слишком мало. Вероятно, Вовочка ничего не заметит. Оценки IV при малых n смещены. О состоятельности по 100 наблюдениям говорить бессмысленно. Если бы наблюдений было не 100, а гораздо больше, то из-за проблем с условием ранга, на втором шаге вылезала бы практически жёсткая мультиколлинеарность.

c) Оценки IV при малых n смещены. О состоятельности по 100 наблюдениям говорить бессмысленно.

18.1.

$$\mathrm{Var}(g) = \mathrm{Var}\left(\begin{pmatrix} y_i - 2\theta \\ 1/y_i - \theta \end{pmatrix}\right)$$

18.2.

18.3.

18.4.

18.5. да, да

18.6. Оценка при единичной весовой матрице равна $\hat{\theta}_{W=I} = 1.5$. С точностью до деления на определитель ковариационной матрицы оценка матрицы весов имеет вид:

$$\hat{W} = \begin{pmatrix} 12 & 1 \\ 1 & 102 \end{pmatrix}$$

19.1. $\hat{\beta}_W = \hat{\beta}_{FD}$, $\mathrm{SS}^{\text{res}}_W = 2\mathrm{SS}^{\text{res}}_{FD}$, обе оценки состоятельны.

19.2. $\hat{\beta}_W = \hat{\beta}_{FD}$

19.3.

21 Sources of Wisdom

[Sch] Kurt Schmidheiny. *Lecture Notes in Microeconometrics: Instumental Variables*. URL: https://www.schmidheiny.name/teaching/iv.pdf. У Курта Шмидхайни очень аккуратные и понятные маленькие заметки по разным темам.

[Pie] Richard G. Pierse. *Some lecture notes on Econometrics: Simultaneous Equations Models: Identification, Estimation and Testing*. URL: http://rpierse.esy.es/rpierse/files/ec7.pdf. У Ричарда Пирса много заметок по разным темам, в основном по временным рядам.

[Zso10] Peter Zsohar. *Short introduction to the generalized method of moments*. 2010. URL: http://www.ksh.hu/statszemle_archive/2012/2012_K16/2012_K16_150.pdf.