

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

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

Для удобства расчетов приведены матрицы

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

- Найдите вектор МНК-оценок коэффициентов  $\hat{\beta}$ .
  - Найдите коэффициент детерминации  $R^2$ .
  - Предполагая нормальное распределение вектора  $\varepsilon$ , проверьте гипотезу  $H_0: \beta_2 = 0$  против альтернативной  $H_a: \beta_2 \neq 0$ .
2. Для линейной регрессии  $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$  была выполнена сортировка наблюдений по возрастанию переменной  $x$ . Исходная модель оценивалась по разным частям выборки:

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

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием.

- Предполагая гомоскедастичность остатков на уровне значимости 5% проверьте гипотезу, что исследуемая зависимость одинакова на всех трёх частях всей выборки.
- Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.
- Какой тест можно на гетероскедастичность можно было бы использовать, если бы не было уверенности в нормальности остатков? Опишите пошагово процедуру этого теста.

3. По 2040 наблюдениям оценена модель зависимости стоимости квартиры в Москве (в 1000\$) от общего метража и метража жилой площади.

	Estimate	Std. Error	t value	Pr(> t )
Константа	-88.81	4.37	-20.34	0.00
Общая площадь	1.70	0.10	17.78	0.00
Жилая площадь	1.99	0.18	10.89	0.00

Оценка ковариационной матрицы  $\widehat{Var}(\hat{\beta})$  имеет вид

	(Intercept)	totsp	livesp
(Intercept)	19.07	0.03	-0.45
totsp	0.03	0.01	-0.02
livesp	-0.45	-0.02	0.03

- (a) Можно ли интерпретировать коэффициент при переменной *totsp* как стоимость одного метра нежилой площади?
- (b) Проверьте гипотезу о том, что коэффициенты при регрессорах *totsp* и *livesp* равны.
- (c) Постройте 95%-ый доверительный интервал для ожидаемой стоимости квартиры с жилой площадью 30 м<sup>2</sup> и общей площадью 60 м<sup>2</sup>.
- (d) Постройте 95%-ый прогнозный интервал для фактической стоимости квартиры с жилой площадью 30 м<sup>2</sup> и общей площадью 60 м<sup>2</sup>.
4. Предположим, что в классической линейной модели ошибки имеют нормальное распределение, т.е.

$$y_i = \beta_1 + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i} + \varepsilon_i$$

где  $\varepsilon_i$  нормальны  $N(0, \sigma^2)$  и независимы

- (a) Найдите оценки для  $\beta$  и  $\sigma^2$  методом максимального правдоподобия.
- (b) Являются ли полученные оценки  $\hat{\beta}_{ML}$  и  $\hat{s}_{ML}^2$  несмещенными?
- (c) Выведите формулу  $LR$ -статистики у теста отношения правдоподобия для тестирования гипотезы об адекватности регрессии  $H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$ .
5. В модели есть три регрессора,  $x_1$ ,  $x_2$  и  $x_3$ . Для удобства будем считать, что они центрированы и нормированы, т.е. выборочное среднее каждого регрессора равно нулю, а выборочная дисперсия — единице. Эти три регрессора являются столбцами матрицы  $X$ . Известно, что

$$X'X = \begin{pmatrix} 100 & 0 & 0 \\ 0 & 100 & 90 \\ 0 & 90 & 100 \end{pmatrix}$$

- (a) Найдите число обусловленности матрицы  $X'X$ .
- (b) Выразите первые две главные компоненты через  $x_1$ ,  $x_2$  и  $x_3$

6. По данным о пассажирах Титаника оценивается логит-модель. Зависимая переменная *survived* равна 1, если пассажир выжил.

	Model 1
(Intercept)	1.92*** (0.28)
age	-0.01 (0.01)
sexmale	-2.84*** (0.21)
AIC	633.45
BIC	646.80
Log Likelihood	-313.72
Deviance	627.45
Num. obs.	633

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Таблица 1: Statistical models

- (a) Оцените вероятность выжить для мужчины 30 лет
- (b) Оцените предельный эффект увеличения возраста для мужчины 30 лет
- (c) С помощью какого метода оценивается логит-модель? Каким образом получаются оценки стандартных ошибок коэффициентов?

Дополнительная задача

7. Рассмотрите модель  $y_i = \beta x_i + \varepsilon_i$ .

- (a) Как выглядит МНК оценка  $\hat{\beta}_{ols}$ ?
- (b) Как выглядит оценка методом Ridge regression с коэффициентом штрафа  $\lambda$ ?
- (c) Как выглядит оценка методом LASSO с коэффициентом штрафа  $\lambda$ ?