

# Задачник по эконометрике-1

(с шахматами и поэтэссами)

Дмитрий Борзых, Борис Демешев

16 октября 2012 г.

## 1 Неклассифицировано

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

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

- (a) Укажите число наблюдений.
  - (b) Укажите число регрессоров с учетом свободного члена.
  - (c) Рассчитайте  $TSS = \sum (y_i - \bar{y})^2$ ,  $RSS = \sum (y_i - \hat{y}_i)^2$  и  $ESS = \sum (\hat{y}_i - \bar{y})^2$ .
  - (d) Рассчитайте при помощи метода наименьших квадратов  $\hat{\beta}$ , оценку для вектора неизвестных коэффициентов.
  - (e) Чему равен  $\hat{\varepsilon}_5$ , МНК-остаток регрессии, соответствующий 5-ому наблюдению?
  - (f) Чему равен  $R^2$  в модели? Прокомментируйте полученное значение с точки зрения качества оцененного уравнения регрессии.
  - (g) Используя приведенные выше данные, рассчитайте несмещенную оценку для неизвестного параметра  $\sigma^2$  регрессионной модели.
  - (h) Рассчитайте  $\widehat{\text{Cov}}(\hat{\beta})$ , оценку для ковариационной матрицы вектора МНК-коэффициентов  $\hat{\beta}$ .
  - (i) Найдите  $\widehat{\text{Var}}(\hat{\beta}_1)$ , несмещенную оценку дисперсии МНК-коэффициента  $\hat{\beta}_1$ .
  - (j) Найдите  $\widehat{\text{Var}}(\hat{\beta}_2)$ , несмещенную оценку дисперсии МНК-коэффициента  $\hat{\beta}_2$ .
  - (k) Найдите  $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$ , несмещенную оценку ковариации МНК-коэффициентов  $\hat{\beta}_1$  и  $\hat{\beta}_2$ .
  - (l) Найдите  $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2)$ ,  $\widehat{\text{Var}}(\hat{\beta}_1 - \hat{\beta}_2)$ ,  $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)$ ,  $\widehat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2 - 2\hat{\beta}_3)$
  - (m) Найдите  $\widehat{\text{Corr}}(\hat{\beta}_1, \hat{\beta}_2)$ , оценку коэффициента корреляции МНК-коэффициентов  $\hat{\beta}_1$  и  $\hat{\beta}_2$ .
  - (n) Найдите  $s_{\hat{\beta}_1}$ , стандартную ошибку МНК-коэффициента  $\hat{\beta}_1$ .
2. Априори известно, что парная регрессия должна проходить через точку  $(x_0, y_0)$ .

- (a) Выведите формулы МНК оценок;
  - (b) В предположениях теоремы Гаусса-Маркова найдите дисперсии и средние оценок
3. Слитки-вариант. Перед нами два золотых слитка и весы, производящие взвешивания с ошибками. Взвесив первый слиток, мы получили результат 300 грамм, взвесив второй слиток — 200 грамм, взвесив оба слитка — 400 грамм. Предположим, что ошибки взвешивания — независимые одинаково распределенные случайные величины с нулевым средним.
- (a) Найдите несмещенную оценку веса первого шара, обладающую наименьшей дисперсией.
  - (b) Как можно проинтерпретировать нулевое математическое ожидание ошибки взвешивания?
4. Вася считает, что  $sCov(y, \hat{y}) = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}}$  это неплохая оценка для  $Cov(y_i, \hat{y}_i)$ . Прав ли он?
5. Сгенерировать набор данных, обладающий следующим свойством. Если попытаться сразу выкинуть регрессоры  $x$  и  $z$ , то гипотеза о их совместной незначимости отвергается. Если вместо этого попытаться выкинуть отдельно  $x$ , или отдельно  $z$ , то гипотеза о незначимости не отвергается.
6. Сгенерировать набор данных, обладающий следующим свойством. Если попытаться сразу выкинуть регрессоры  $x$  и  $z$ , то гипотеза о их совместной незначимости отвергается. Если вместо сначала выкинуть отдельно  $x$ , то гипотеза о незначимости не отвергается. Если затем выкинуть  $z$ , то гипотезы о незначимости тоже не отвергается.
7. К эконометристу Вовочке в распоряжение попали данные с результатами контрольной работы студентов по эконометрике. В данных есть результаты по каждой задаче, переменные  $p_1, p_2, p_3, p_4$  и  $p_5$ , и суммарный результат за контрольную, переменная  $kr$ . Чему будут равны оценки коэффициентов, их стандартные ошибки,  $t$ -статистики,  $P$ -значения,  $R^2$ ,  $RSS$ , если
- (a) Вовочка построит регрессию  $kr$  на константу,  $p_1, p_2, p_3, p_4$  и  $p_5$
  - (b) Вовочка построит регрессию  $kr$  на  $p_1, p_2, p_3, p_4$  и  $p_5$  без константы
8. Как построить доверительный интервал для вершины параболы? ...

## 2 МНК без матриц и вероятностей

1. Даны  $n$  пар чисел:  $(x_1, y_1), \dots, (x_n, y_n)$ . Мы прогнозируем  $y_i$  по формуле  $\hat{y}_i = \hat{\beta}x_i$ . Найдите  $\hat{\beta}$  методом наименьших квадратов.
2. Даны  $n$  чисел:  $y_1, \dots, y_n$ . Мы прогнозируем  $y_i$  по формуле  $\hat{y}_i = \hat{\beta}$ . Найдите  $\hat{\beta}$  методом наименьших квадратов.
3. Даны  $n$  пар чисел:  $(x_1, y_1), \dots, (x_n, y_n)$ . Мы прогнозируем  $y_i$  по формуле  $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$ . Найдите  $\hat{\beta}_1$  и  $\hat{\beta}_2$  методом наименьших квадратов.
4. Даны  $n$  пар чисел:  $(x_1, y_1), \dots, (x_n, y_n)$ . Мы прогнозируем  $y_i$  по формуле  $\hat{y}_i = 1 + \hat{\beta}x_i$ . Найдите  $\hat{\beta}$  методом наименьших квадратов.
5. 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.
6. 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.
7. Регрессия на дамми-переменную...

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

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

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

- (а) Возможно ли, что  $\hat{\beta}_2 > 0$ ,  $\hat{\gamma}_2 > 0$ , но  $\hat{\eta}_2 < 0$ ?
- (б) Возможно ли, что  $\hat{\beta}_1 > 0$ ,  $\hat{\gamma}_1 > 0$ , но  $\hat{\eta}_1 < 0$ ?
- (с) Возможно ли одновременное выполнение всех упомянутых условий?

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

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

- (а)  $y$  на константу и  $1_f$
- (б)  $y$  на константу и  $1_m$
- (с)  $y$  на  $1_f$  и  $1_m$  без константы
- (д)  $y$  на константу,  $1_f$  и  $1_m$

12. Какие из указанные моделей можно представить в линейном виде?

- (а)  $y_i = \beta_1 + \frac{\beta_2}{x_i} + \varepsilon_i$
- (б)  $y_i = \exp(\beta_1 + \beta_2 x_i + \varepsilon_i)$
- (с)  $y_i = 1 + \frac{1}{\exp(\beta_1 + \beta_2 x_i + \varepsilon_i)}$
- (д)  $y_i = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i + \varepsilon_i)}$
- (е)  $y_i = x_i^{\beta_2} e^{\beta_1 + \varepsilon_i}$

### 3 Инструментальные переменные

1. Табличка 2 на 2. Найдите  $\mathbb{E}(\varepsilon)$ ,  $\mathbb{E}(\varepsilon|x)$ ,  $\text{Cov}(\varepsilon, x)$ .
2. Все предпосылки классической линейной модели выполнены,  $y = \beta_1 + \beta_2 x + \varepsilon$ . Рассмотрим альтернативную оценку коэффициента  $\beta_2$ ,

$$\hat{\beta}_{2,IV} = \frac{\sum z_i (y_i - \bar{y})}{\sum z_i (x_i - \bar{x})} \quad (2)$$

- (а) Является ли оценка несмещенной?
- (б) Любые ли  $z_i$  можно брать?
- (с) Найдите  $\text{Var}(\hat{\beta}_{2,IV})$

3.

## 4 Проекция, Картинка

1. Найдите на Картинке четыре прямоугольных треугольника. Сформулируйте четыре теоремы Пифагора.
2. Покажите на Картинке  $TSS$ ,  $ESS$ ,  $RSS$ ,  $R^2$ ,  $sCov(\hat{y}, y)$
3. Предложите аналог  $R^2$  для случая, когда константа среди регрессоров отсутствует. Аналог должен быть всегда в диапазоне  $[0; 1]$ , совпадать с обычным  $R^2$ , когда среди регрессоров есть константа, равняться единице в случае нулевого  $\hat{\varepsilon}$ .
4. 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 коэффициента при  $\hat{y}$ ? Почему оценка коэффициента неслучайна, а оценка её дисперсии положительна?
5. При каких условиях  $TSS = ESS + RSS$ ?

## 5 МЕГАМАТРИЦА

1. В рамках классической линейной модели найдите ковариационные матрицы всех пар случайных векторов:  $\varepsilon$ ,  $y$ ,  $\hat{y}$ ,  $\hat{\varepsilon}$ ,  $\hat{\beta}$
2. Найдите  $\mathbb{E}(\sum(\varepsilon_i - \bar{\varepsilon})^2)$ ,  $\mathbb{E}(RSS)$
3.  $\mathbb{E}(TSS)$ ,  $\mathbb{E}(ESS)$  — громоздкие
4. Вася строит регрессию  $y$  на некий набор объясняющих переменных и константу. А на самом деле  $y_i = \beta_1 + \varepsilon_i$ . Чему равно  $\mathbb{E}(TSS)$ ,  $\mathbb{E}(RSS)$ ,  $\mathbb{E}(ESS)$  в этом случае?

## 6 Голая линейная алгебра

Здесь будет собран минимум задач по линейной алгебре.

1. Приведите пример таких  $A$  и  $B$ , что  $\det(AB) \neq \det(BA)$ .
2. Для матриц-проекторов  $\pi = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$  и  $P = X(X'X)^{-1}X'$  найдите  $\text{tr}(\pi)$ ,  $\text{tr}(P)$ ,  $\text{tr}(I - \pi)$ ,  $\text{tr}(I - P)$ .

## 7 Компьютерные упражнения

1. Скачайте результаты двух контрольных работ по теории вероятностей, с описанием данных, . Скачайте табличку соответствия имени и пола, . Наша задача попытаться предсказать результат второй контрольной работы зная позадачный результат первой контрольной.
  - (a) Какая задача из первой контрольной работы наиболее существенно влияет на результат второй контрольной?
  - (b) Влияет ли пол на результат второй контрольной?
  - (c) Влияет ли редкость имени на результат второй контрольной?
  - (d) Что можно сказать про влияние группы, в которой учится студент?
2. Задача Макар-Лиманова. У торговца 55 пустых стаканчиков, разложенных в несколько стопок. Пока нет покупателей он развлекается: берет верхний стаканчик из каждой стопки и формирует из них новую стопку. Потом снова берет верхний стаканчик из каждой стопки и формирует из них новую стопку и т.д.
  - (a) Напишите функцию 'makar\_step'. На вход функции подаётся вектор количества стаканчиков в каждой стопке до переукладывания. На выходе функция возвращает количества стаканчиков в каждой стопке после одного переукладывания.

- (b) Изначально стаканчики были разложены в две стопки, из 25 и 30 стаканчиков. Как разложатся стаканчики если покупателей не будет достаточно долго?
3. Напишите функцию, которая бы оценивала регрессию методом наименьших квадратов. На вход функции должны подаваться вектор зависимых переменных  $y$  и матрица регрессоров  $X$ . На выходе функция должна выдавать список из  $\hat{\beta}$ ,  $\widehat{\text{Var}}(\hat{\beta})$ ,  $\hat{y}$ ,  $\hat{\varepsilon}$ ,  $ESS$ ,  $RSS$  и  $TSS$ . По возможности функция должна проверять корректность аргументов, например, что в  $y$  и  $X$  одинаковое число наблюдений и т.д.
  4. Сгенерируйте вектор  $y$  из 300 независимых нормальных  $N(10, 1)$  случайных величин. Сгенерируйте 40 «объясняющих» переменных, по 300 наблюдений в каждой, каждое наблюдение — независимая нормальная  $N(5, 1)$  случайная величина. Постройте регрессию  $y$  на все 40 регрессоров и константу.
    - (a) Сколько регрессоров оказалось значимо на 5% уровне?
    - (b) Сколько регрессоров в среднем значимо на 5% уровне?
    - (c) Эконометрист Вовочка всегда использует следующий подход: строит регрессию зависимой переменной на все имеющиеся регрессоры, а затем выкидывает из модели те регрессоры, которые оказались незначимы. Прокомментируйте Вовочкин эконометрический подход.

## 8 Вопросы теоретического характера

1. Что означают слова автокорреляция, гетероскедастичность, гомоскедастичность?
2. Напишите формулу для оценок коэффициентов в парной регрессии без матриц
3. Напишите формулу для оценок коэффициентов в множественной регрессии
4. Аналогично для дисперсий