

# Задачник по эконометрике-1

(с шахматами и поэтэссами)

Дмитрий Борzych, Борис Демешев

12 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  $y =$ ,  $X =$ . Для удобства расчетов ниже приведены матрицы  $X'X =$  и  $(X'X)^{-1} =$ .
  - (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  $\text{Corr}(\hat{\beta}_1, \hat{\beta}_2)$ , 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

Вроде бы равносильно переносу начала координат и применению результата для регрессии без свободного члена. Должна остаться несмещенность.

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?

Как отсутствие систематической ошибки.

4. Вася считает, что  $sCov(y, \hat{y}) = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2}}$  это неплохая оценка для  $Cov(y_i, \hat{y}_i)$ . Прав ли он? Не прав. Ковариация  $Cov(y_i, \hat{y}_i)$  зависит от  $i$ , это не одно неизвестное число, для которого можно предложить одну оценку.
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. Сгенерировать сильно коррелированные  $x$  и  $z$
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

## 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.  $\hat{\beta} = \sum x_i y_i / \sum x_i^2$
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.  $\hat{\beta} = \bar{y}$
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.  $\hat{\beta}_2 = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sum (x_i - \bar{x})^2$ ,  $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$
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.  $\hat{\beta} = \sum x_i (y_i - 1) / \sum x_i^2$
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.  $(300 - \hat{\beta}_1)^2 + (200 - \hat{\beta}_2)^2 + (400 - \hat{\beta}_1 - \hat{\beta}_2)^2 \rightarrow \min$
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.  $2 \cdot (10 - \hat{\beta})^2 + (3 - \hat{\beta})^2 \rightarrow \min$
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  $\hat{y} = \hat{m}_1 + \hat{m}_2 x$  по всем наблюдениям.

- (a) Возможно ли, что  $\hat{\beta}_2 > 0$ ,  $\hat{\gamma}_2 > 0$ , но  $\hat{m}_2 < 0$ ?
- (b) Возможно ли, что  $\hat{\beta}_1 > 0$ ,  $\hat{\gamma}_1 > 0$ , но  $\hat{m}_1 < 0$ ?
- (c) Возможно ли одновременное выполнение всех упомянутых условий?

да, возможно. Два вытянутых облачка точек. Первое облачко даёт первую регрессию, второе — вторую. Прямая, соединяющая центры облачков, — общую.

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? Нет. Коэффициенты можно интерпретировать только «при прочих равных», т.е. при равных  $x$ . Из-за разных  $x$  может оказаться, что у мужчин  $\bar{y}$  меньше, чем  $\bar{y}$  для женщин.

### 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  $\beta_2$ ,

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

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

Да, является. Любые, кроме констант.  $\text{Var}(\hat{\beta}_{2,IV}) = \sigma^2 \sum (z_i - \bar{z})^2 / (\sum (z_i - \bar{z})x_i)^2$ .

3.

### 4 Проекция, Картинка

1. Найдите на Картинке четыре прямоугольных треугольника. Сформулируйте четыре теоремы Пифагора.  $\sum y_i^2 = \sum \hat{y}_i^2 + \sum \varepsilon_i^2$ ,  $TSS = ESS + RSS$ ,
2. Покажите на Картинке  $TSS$ ,  $ESS$ ,  $RSS$ ,  $R^2$ ,  $s\text{Cov}(\hat{y}, y)$
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  $\hat{\varepsilon}$ . Спроецируем единичный столбец на «плоскость», обозначим его  $1'$ . Делаем проекцию  $y$  на «плоскость» и на  $1'$ . Далее аналогично.
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? Проекция  $y$  на  $\hat{y}$  это  $\hat{y}$ , поэтому оценки коэффициентов будут 0 и 1. Оценка дисперсии

$\frac{RSS}{(n-2)ESS}$ . Нарушены предпосылки теоремы Гаусса-Маркова, например, ошибки новой модели в сумме дают 0, значит коррелированы.

### 5 МЕГАМАТРИЦА

1. В рамках классической линейной модели найдите ковариационные матрицы всех пар случайных векторов:  $\varepsilon$ ,  $y$ ,  $\hat{y}$ ,  $\hat{\varepsilon}$ ,  $\hat{\beta}$   $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
2. Найдите  $\mathbb{E}(\sum(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2)$ ,  $\mathbb{E}(RSS)$   $(n-1)\sigma^2$ ,  $(n-k)\sigma^2$
3.  $\mathbb{E}(TSS)$ ,  $\mathbb{E}(ESS)$  — громоздкие  $\mathbb{E}(TSS) = (n-1)\sigma^2 + \beta'X'(I - \pi)X\beta$

## 6 Голая линейная алгебра

Здесь будет собран минимум задач по линейной алгебре.

1. Приведите пример таких  $A$  и  $B$ , что  $\det(AB) \neq \det(BA)$ . Например,  $A = (1, 2, 3)$ ,  $B = (1, 0, 1)'$
2. Для матриц-проекторов  $\pi = \vec{1}(\vec{1}'\vec{1})^{-1}\vec{1}'$  и  $P = X(X'X)^{-1}X'$  найдите  $\text{tr}(\pi)$ ,  $\text{tr}(P)$ ,  $\text{tr}(I - \pi)$ ,  $\text{tr}(I - P)$ .  $\text{tr}(I) = n$ ,  $\text{tr}(\pi) = 1$ ,  $\text{tr}(P) = k$

## 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 Макара-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.