# Конспект семинаров по метрике-2016

Студенты ИП и Борис Демешев

2017-02-27

# Оглавление

# О конспекте

Метрика с R :)

Напутствия:

1. Идеальный конспект — интересный и без отклонений от здравомыслия. Примочки и пеночки нужны! Не будь занудой!

2. Конспект одного семинара должен иметь одно заглавие уровня # и несколько, скажем от двух до пяти, подзаголовков уровня ##. После «решёточек» должен идти пробел. После заголовка должен стоять краткий английский уникальный идентификатор, например, {#04_matrix_algebra}.

3. Помни об оформлении знаков препинания: после запятой есть пробел, а до запятой — нет. Существует длинное тире, —, которое отличается от просто дефиса -.

4. Рисунки оформляй в открытом софте (tex + tikz, inkscape, graphviz, geogebra, draw.io и прочее) и прикладывай к работе. Есть ещё Ipe и tikzit.

5. Рисунки клади в подпапку images

6. Соблюдай конвенцию о названиях файлов: файлы относящиеся к третьему семинару должны начинаться с 03_, и сам конспект и рисунки. Имена файлов не должны содержать русских букв и пробелов.

7. Никаких здесь и тут в ссылках. Текст, замещающий ссылку, должен быть осмысленным! Например, [Весёлый поисковик](http://www.yandex.ru)

8. Про язык разметки маркдаун. Подробно рассказано как вставлять картинки, списки, цитаты, оформлять разделы и подразделы.

9. Про язык латех от Воронцова. Подробно рассказано, как набирать дроби, суммы и другие формулы. Частые косяки:

   - Функции оформляются со слэшем спереди, например, \ln, \cos

   - Забудь про $$! Формулы на всю строчку пишутся в окружении \[...\], а именно в три строки:

     ```
     \[
     a^2 + b^2 = c^2
     \]
     ```

   - Русский текст внутри формул пишется с помощью \text{Привет!}.

10. Посмотри, как сделали конспект другие, и сделай лучше! :) Обрати внимание на название .Rmd файлов, на структуру внутри .Rmd файлов.

11. Каждый кусок кода должен иметь уникальное название, например, {r, "plotting_histogram"}

12. Уважай букву ё – ставь над ней точки! :)

```
library("tidyverse") # ggplot2 for plots, dplyr for data manipulation, broom and more
library("sandwich") # оценка Var для гетероскедастичности
library("lmtest") # тест Бройша-Пагана
library("data.table") # манипуляции с данными
library("reshape2") # преобразование длинных таблиц в широкие
```

Данная версия конспекта скомпилирована для latex.

# Глава 1

# Метод наименьших квадратов

Конспект: Бердникович Алеся, Головина Мария

дата: 05.09.2016

## 1.1 Основная задача

Маша каждый день ловит покемонов и решает задачи по теории вероятностей. Пусть x и y - случайные величины, $x_i$ - количество решённых в i-тый день задач, а $y_i$ - количество пойманных в i-тый день покемонов. Результаты наблюдения за действиями Маши представлены в таблице:

| День | $x_i$ | $y_i$ |
|---|---|---|
| 1 | 1 | 10 |
| 2 | 2 | 0 |
| 3 | 0 | 4 |

Необходимо определить, как количество пойманных за день покемонов зависит от количества решённых за день задач. Предположим, что регрессионная модель имеет линейный вид $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, где коэффициенты $\beta_1, \beta_2$ неизвестны и должны быть оценены, а $\epsilon_i$ - случайная величина. Тогда прогнозируемая зависимость имеет вид $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$.

### 1.1.1 Метод наименьших квадратов (OLS):

$y_i - \hat{y}_i$ - ошибка прогноза, которую нужно минимизировать. Штрафная функция:

$$Q(\hat{\beta}_1, \hat{\beta}_2) = (y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+(y_3-\hat{y}_3)^2 = (y_1-(\hat{\beta}_1+\hat{\beta}_2 x_1))^2+(y_2-(\hat{\beta}_1+\hat{\beta}_2 x_2))^2+(y_3-(\hat{\beta}_1+\hat{\beta}_2 x_3))^2 \to min_{\hat{\beta}_1,\hat{\beta}_2}$$

### 1.1.2 Метод наименьших модулей (LAD):

Альтернативный метод минимизации ошибок прогноза. Отличие заключается в виде штрафной функции:

$$Q(\hat{\beta}_1, \hat{\beta}_2) = |y_1 - \hat{y_1}|^2 + |y_2 - \hat{y}_2|^2 + |y_3 - \hat{y}_3|^2 \to min_{\hat{\beta}_1, \hat{\beta}_2}$$

Найдём $\hat{\beta}_1, \hat{\beta}_2$ в нашей задаче методом наименьших квадратов:

$$Q(\hat{\beta}_1, \hat{\beta}_2) = (y_1 - (\hat{\beta}_1 + \hat{\beta}_2 x_1)^2 + (y_2 - (\hat{\beta}_1 + \hat{\beta}_2 x_2))^2 + (y_3 - (\hat{\beta}_1 + \hat{\beta}_2 x_3))^2$$

$$Q(\hat{\beta}_1, \hat{\beta}_2) = (10 - (\hat{\beta}_1 + \hat{\beta}_2)^2 + (0 - (\hat{\beta}_1 + 2\hat{\beta}_2))^2 + (4 - (\hat{\beta}_1))^2 \to min$$

$$\frac{\partial Q}{\partial \hat{\beta}_1} = -28 + 6\hat{\beta}_1 + 6\hat{\beta}_2 = 0$$

$$\frac{\partial Q}{\partial \hat{\beta}_2} = -20 + 6\hat{\beta}_1 + 10\hat{\beta}_2 = 0$$

$$\hat{\beta}_1 = \frac{20}{3}, \hat{\beta}_2 = -2$$

Искомая оценка зависимости числа пойманных покемонов от числа решённых задач:

$$\hat{y}_i = \frac{20}{3} - 2x_i$$

## 1.2 Реализация в R:

```
x <- c(1, 2, 0)
y <- c(10, 0, 4)
md <- data.frame(problem = x, pokemon = y)
md
```

```
##   problem pokemon
## 1       1      10
## 2       2       0
## 3       0       4
```

Восстановление линейной зависимости методом наименьших квадратов:

```
model_1_ols <- lm(data = md, pokemon~problem)
summary(model_1_ols)
```

```
##
## Call:
## lm(formula = pokemon ~ problem, data = md)
##
## Residuals:
##      1      2      3
##  5.333 -2.667 -2.667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.667      5.963   1.118    0.465
## problem       -2.000      4.619  -0.433    0.740
##
## Residual standard error: 6.532 on 1 degrees of freedom
## Multiple R-squared:  0.1579, Adjusted R-squared:  -0.6842
## F-statistic: 0.1875 on 1 and 1 DF,  p-value: 0.7399
```

Подключаем нужный пакет:

```
library("quantreg")
```

Если пакета не установлен, то это исправляется командой

```
install.packages("quantreg")
```

Восстановление линейной зависимости методом наименьших модулей:

```
model_1_lad <- rq(data = md, pokemon~problem)
summary(model_1_lad)
```

```
##
## Call: rq(formula = pokemon ~ problem, data = md)
##
## tau: [1] 0.5
```

```
##
## Coefficients:
##             coefficients  lower bd       upper bd
## (Intercept)  4.000000e+00 -1.797693e+308  1.797693e+308
## problem     -2.000000e+00 -1.797693e+308  1.797693e+308
```

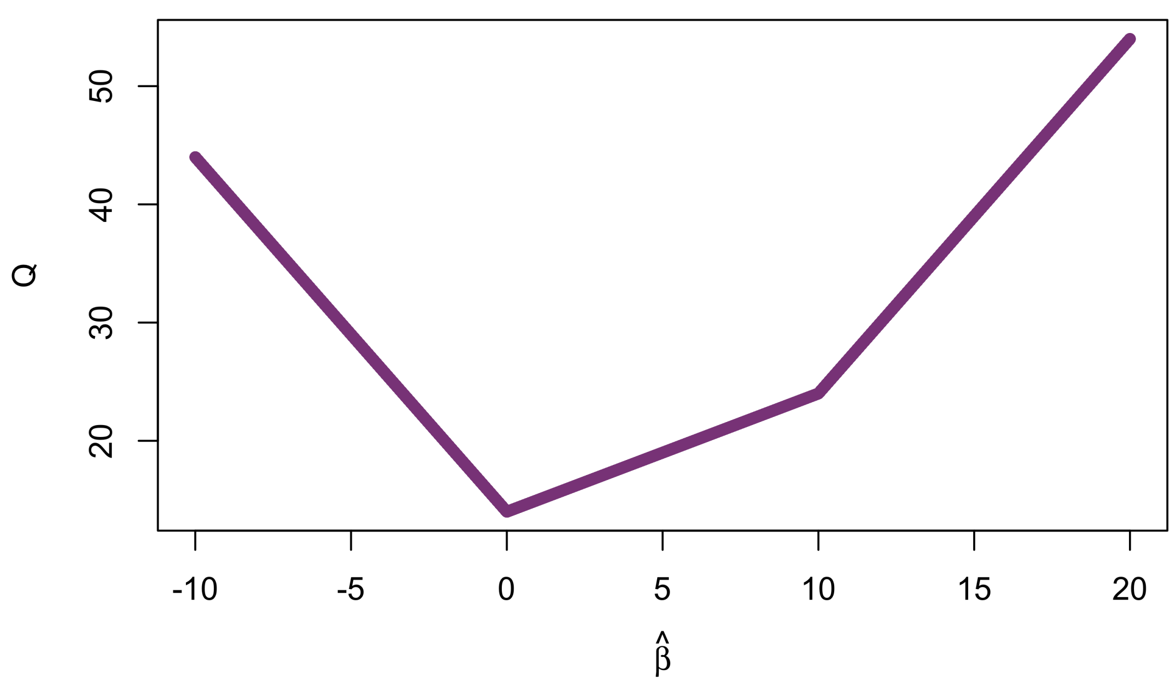
Предположим теперь иную модель зависимости $y_i = \hat{\beta} x_i$, ищем оценку единственного неизвестного коэффициента $\hat{\beta}$ с помощью метода наименьших модулей. Штрафная функция примет вид

$$Q(\hat{\beta}) = |10 - \hat{\beta}| + |0 - \hat{\beta}| + |4 - 0| \to min$$

Точки изломов функции находятся в нулях подмодульных выражений: $\hat{\beta} = 0$ и $\hat{\beta} = 10$. Функция принимает наименьшее значение при $\hat{\beta} = 0$ (см. график), что говорит об отсутствии зависимости числа пойманных покемонов от числа решённых задач.

### 1.2.1 График штрафной функции:

```r
x <- seq(-10, 20, 0.001)
fx <- (x <= 0) * (14 - 3 * x) +
      (x > 0 & x < 10) * (14 + x) +
      (x >= 10) * (3*x - 6)
plot(x = x, y = fx, xlab = expression(hat(beta)), ylab = 'Q', pch = 20, col = 'orchid4')
```



## 1.3 Домашнее задание

Вывести общие формулы для коэффициентов $\hat{\beta}, \hat{\beta}_1, \hat{\beta}_2$, используя МНК-оценку, при условии, что:

1. $y_i = \beta x_i + \epsilon_i,\ \hat{y}_i = \hat{\beta} x_i;$
2. $y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \epsilon_i,\ \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \epsilon_i.$

# Глава 2

# Вспомнить всё

1. Найдите длины векторов $a = (1, 1, 1)$ и $b = (1, 2, 3)$ и косинус угла между ними. Найдите один любой вектор, перпенидкулярный вектору $b$.

2. Сформулируйте теорему о трёх перпендикулярах и обратную к ней

3. На плоскости $\alpha$ лежит прямая $\ell$. Вне плоскости $\alpha$ лежит точка $C$. Ромео проецирует точку $C$ на прямую $\ell$ и получает точку $R$. Джульетта проецирует точку $C$ сначала на плоскость $\alpha$, а затем проецирует полученную точку $A$ на прямую $\ell$. После двух действий Джульетта получает точку $D$. Обязательно ли $R$ и $D$ совпадают?

4. Для матрицы
$$A = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$$
   - Найдите собственные числа и собственные векторы матрицы
   - Найдите $\det(A)$, $\operatorname{tr}(A)$
   - Найдите собственные числа матрицы $A^{2016}$, $\det(A^{2016})$ и $\operatorname{tr}(A^{2016})$

5. Известно, что $X$ — матрица размера $n \times k$ и $n > k$, известно, что $X'X$ обратима. Рассмотрим матрицу $H = X(X'X)^{-1}X'$. Укажите размер матрицы $H$, найдите $H^{2016}$, $\operatorname{tr}(H)$, $\det(H)$, собственные числа матрицы $H$. Штрих означает транспонирование.

6. Занудная халява: известно, что $\mathbb{C}ov(X, Y) = 5$, $\mathbb{V}ar(X) = 10$, $\mathbb{V}ar(Y) = 20$, $\mathbb{E}(X) = 10$, $\mathbb{E}(Y) = -10$. Найдите $\mathbb{C}ov(X + 2Y, Y - X)$, $\mathbb{V}ar(X + 2Y)$, $\mathbb{E}(X + 2Y)$.

7. За 100 дней Ромео посчитал все глубокие вздохи Джульетты. Настроение Джульетты столь спонтанно, что глубокие вздохи за разные дни можно считать независимыми. В сумме оказалось 890 вздохов. Сумма квадратов оказалась равна 8000. Постройте 95%-ый доверительный интервал для математического ожидания ежедневного количества глубоких вздохов Джульетты. На~уровне значимости 5%-ов проверьте гипотезу, что математическое ожидание равно~9.

8. Ромео подкидывает монетку два раза. Если монетка выпадает орлом, то Ромео кладет в мешок черный шар, если решкой — белый. Джульетта не знает, как выпадала монетка, и достает шары из мешка наугад по очереди. Первый шар оказался черного цвета. Какова вероятность того, что второй шар Джульетты будет белым?

---

Что-то с памятью моей стало…

Линейная алгебра:

1. Великолепный учебник! Strang, Introduction to linear algebra. Ознакомиться можно на gen.lib.rus.ec :)

Глава 3

# Геометрия МНК

конспект: Света Колесниченко

дата: 19 сентября 2016

## 3.1 Обозначения

Варианты представления регрессии:

- Скалярный вариант: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i$
- Векторный вариант: $\hat{y} = \hat{\beta}_1 e + \hat{\beta}_2 x + \hat{\beta}_3 z$

$$e = \vec{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \text{ - единичный вектор размерности } n \times 1,$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \text{ - векторы переменных}$$

Количество наблюдений $= n$, количество коэффициентов $\beta =$ количество регрессоров $= k$.

- Матричный вариант: $\hat{y} = X \hat{\beta}$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \text{ - вектор размера } k \times 1,$$

$$X = \begin{pmatrix} 1 & x_1 & z_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{pmatrix} \text{ - матрица размера } n \times k$$

Конвенция об обозначениях:

1. $y, \beta, \hat{\beta}, x, z$ - векторы
2. $y_i, \beta_j, \hat{\beta}_7, x_{45}, z_{37}$ - числа (скаляры)
3. $\Omega, X, H$ - матрицы

## 3.2 Ныряем в $n$-мерное пространство

$$\min_{i \in I} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min_{i \in I} \sum_{i=1}^{n} |y_i - \hat{y}_i|^2 \text{ - минимизируем квадрат длины вектора}$$

$$\begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \bar{y} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \bar{y} \cdot \vec{1}$$

$\bar{y} = \hat{y}$ - т.к. $\bar{y} \cdot \vec{1} = \hat{y} \cdot \vec{1}$

среднее значение = среднее значение прогнозов $(\hat{y}_i)$

“Лапа” = $Lin(e, x, z)$ $\leftarrow$ выбираем через e, x, z положение $\hat{y}$

$\hat{y}$ - проекция y на “лапу”

y - линейная комбинация e, x, z $\rightarrow$ лежит в линейной оболочке этих векторов

$\hat{\varepsilon} = y - \hat{y}$ - вектор “остатков”/ошибок прогнозов/resideals

$\hat{\varepsilon} \perp e, \hat{\varepsilon} \perp x, \hat{\varepsilon} \perp y$

$\hat{\varepsilon} \cdot \vec{1} = 0$, $\hat{\varepsilon} \cdot x = 0$, $\hat{\varepsilon} \cdot z = 0$ $\leftarrow$ скалярное произведение векторов (ссыль подробнее) перпендикулярных векторов равно 0.

$\hat{\varepsilon} \perp$ Лапа $\rightarrow$ $\hat{\varepsilon} \perp$ любому вектору, лежащему в Лапе

Великая Теорема о 3 перпендикулярах и аж в 2 формулировках и с чертёжиком.

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0, \sum_{i=1}^{n} \hat{\varepsilon}_i x_i = 0$$

$$X' = \begin{pmatrix} 1 & \dots & 1 \\ x & \dots & x_n \\ z & \dots & z_n \end{pmatrix} \text{ - единичный вектор размерности } n \times 1 \quad \hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} \text{ - единичный вектор размерности } n$$

Условие ортогональности: $X' \cdot \hat{\varepsilon} = 0$ - размерность этого нуля - $k \times 1$

$\hat{y} = X \cdot \hat{\beta} \rightarrow \hat{\beta} = \frac{\hat{y}}{X} = \frac{\hat{\varepsilon} - Y}{X}$

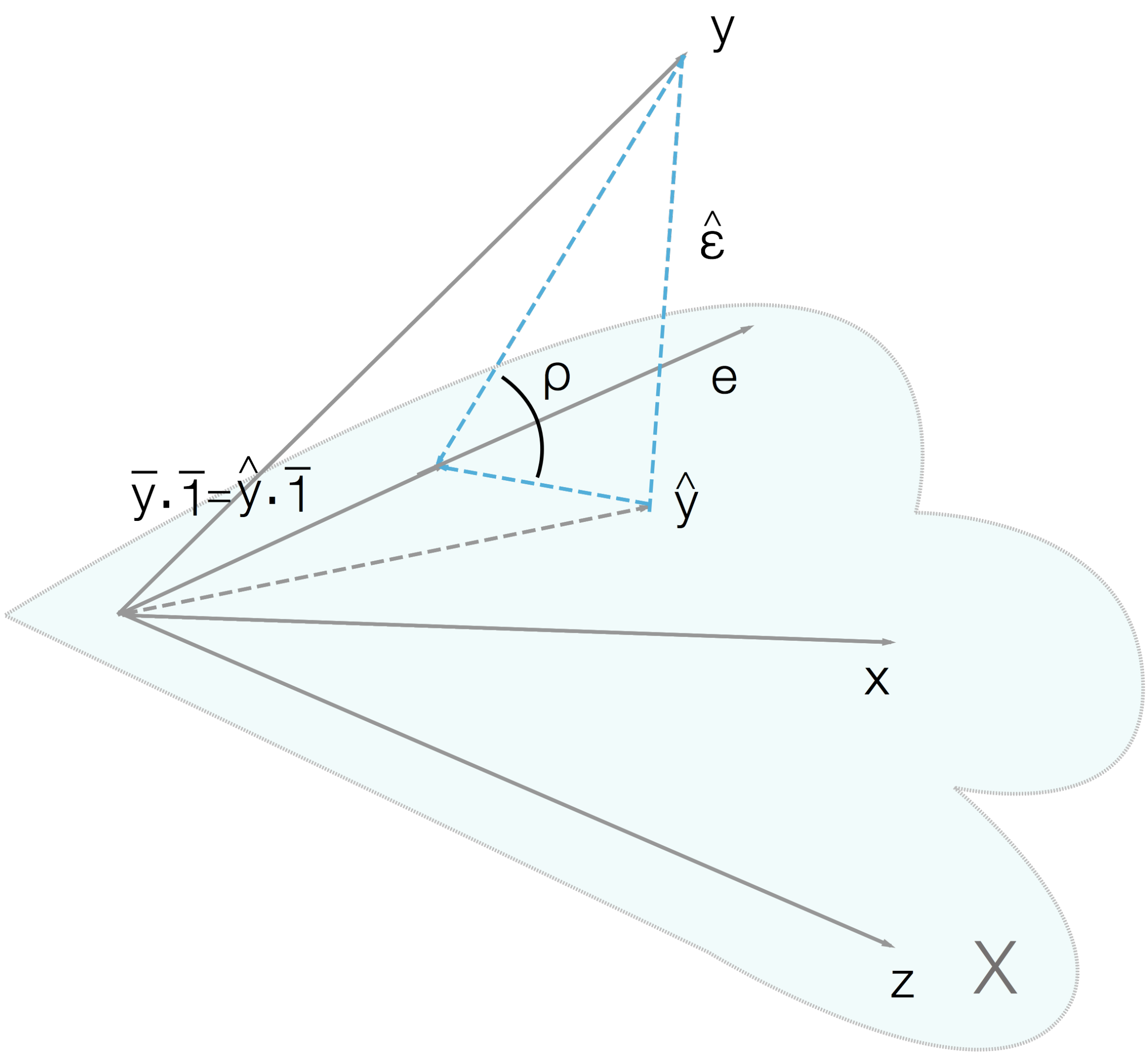



Рис. 3.1: Картиночка Лапы

### 3.2.1 Упражнение 1

Выведите $\hat{\beta}$ из $X' \cdot (y - X \cdot \hat{\beta}) = 0$

$\hat{\varepsilon} = y - \hat{y} = y - X \cdot \hat{\beta}$

$X$ - задает "лапу".

$\hat{\beta}$ - отвечает за то, с каким весом в $\hat{y}$ входят базисные векторы «лапы».

$\hat{\beta} = (\sum_{i=1}^{n} x_i^2)^{-1} \sum_{i=1}^{n} x_i y_i = 0$ - для $\hat{y}_i = \hat{\beta}\, x_i$

$X' \cdot y = X' \cdot X \cdot \hat{\beta}$ $(X' \cdot X)^{-1} \cdot X' \cdot y = (X' \cdot X)^{-1} \cdot X' \cdot X \cdot \hat{\beta}$ $\hat{\beta} = (X' \cdot X)^{-1} \cdot X' \cdot y$

## 3.3 Больше проекций

### 3.3.1 Упражнение 2

Спроецируйте вектор $\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ на прямую, порождённую вектором $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

$$\hat{\beta} = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = 4^{-1} \cdot 10 \cdot \tfrac{1}{4} \cdot 10 = 2.5$$

$$\hat{y} = X \cdot \hat{\beta} = \begin{pmatrix} 2.5 \\ 2.5 \\ 2.5 \\ 2.5 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \end{pmatrix}$$

$$\bar{y} = \frac{1+2+3+4}{4} = 2.5$$

Проекция вектора на прямую из единиц даёт вектор из средних.

By the way, крутые читщиты по матрицам и основам линейной алгебры.

### 3.3.2 Упражнение 3

Сформулируйте все теоремы Пифагора

$\{\hat{\varepsilon}, \hat{y} - \bar{y} \cdot \vec{1}, y - \bar{y} \cdot \vec{1}\}$

По Теореме Пифагора:

$$|y - \bar{y} \cdot \vec{1}|^2 = |\hat{\varepsilon}|^2 + |\hat{y} - \bar{y} \cdot \vec{1}|^2$$

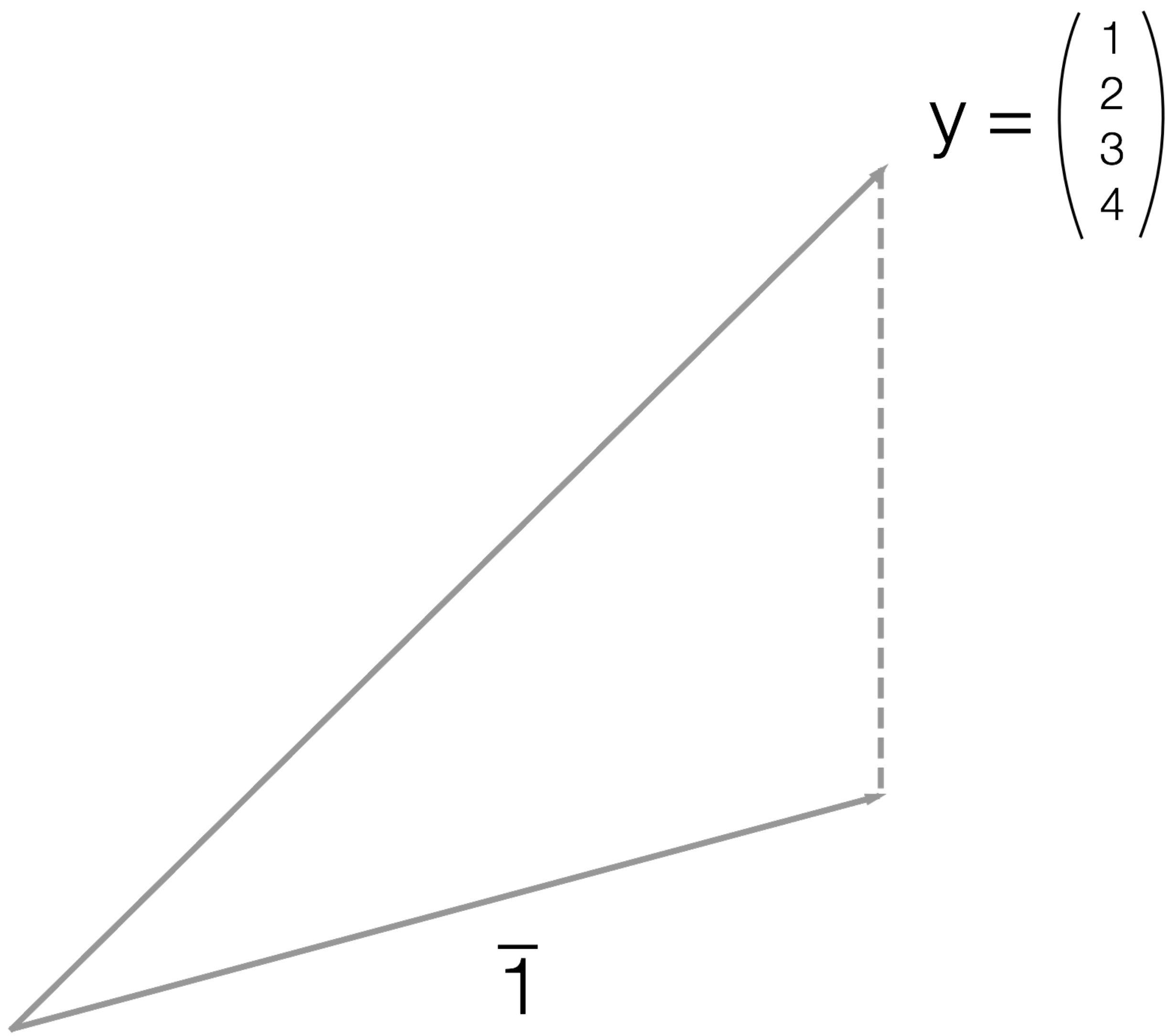



Рис. 3.2: Визуализация задачи

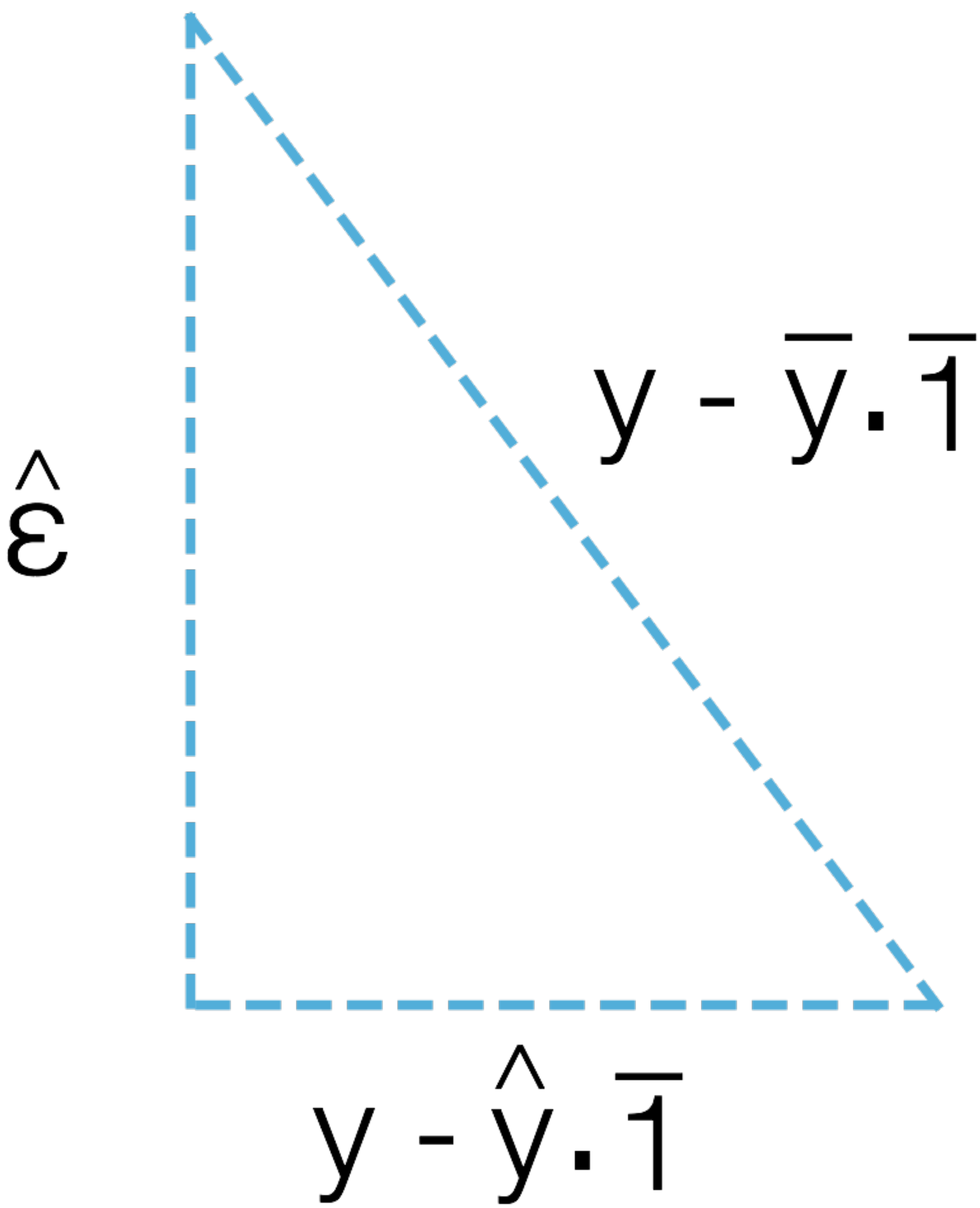



Рис. 3.3: Чертёжик одного из треугольников

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}\hat{\varepsilon}_i^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$y_i-\bar{y}=\begin{pmatrix}y_1-\bar{y}\\ \vdots\\ y_n-\bar{y}\end{pmatrix}$$

Полное задание см. в Задачнике по координатам: 4.23, 4.24, 4.25

Коэффициент детерминации ($R^2$) - примитивный показатель качества прогнозов.

$$R^2=\frac{ESS}{TSS}=\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{\sum_{i=1}^{n}(y_i-\hat{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{\text{residial sum of squares (сумма квадратов остатков)}}{\text{total sum of squares (полная сумма квадратов)}}$$

$$ESS=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\text{ explained sum of squares ("объясненная" сумма квадратов)}$$

В МНК работает (и точно только в нём!) соотношение: $RSS+ESS=TSS$.

В МНК решается задача минимизации RSS.

Если прогнозы $\hat{y}_i$ идеально совпадают с $y_i$, то $R^2=1\Rightarrow ESS=TSS$.

$R^2\in[0;1]\,,R^2=\cos^2\rho$

$\hat{y}$ ближе к $y$ с ростом «лапы» $\Rightarrow\angle\rho\downarrow\to R^2$ т.к. $\cos^2\rho\uparrow$

#### 3.3.2.1 ДЗ:

1.1, 1.2, 1.7, 1.12, 1.13, 4.13 (1-6), 4.23, 4.24, 4.25 из Задачника

##### 3.3.2.1.1 Полезные ссылки:

Репозиторий курса метрики и теории вероятностей

# Глава 4

# Борьба с матрицами

дата: 26 сентября 2016

конспект: Вика Шрамова, Эдуард Аюнц

Семинар посвящен работе с матрицами - матричному дифференцированию и представлению многомерных случайных величин при помощи матриц.

Перед тем как приступить к работе с матрицами, полезно повторить основные свойства операций над матрицами:

1. $A(B+C) = AB + AC$
2. $(A+B)^T = A^T + B^T$
3. $(AB)^T = B^T A^T$
4. $(AB)^{-1} = B^{-1}A^{-1}$
5. $(A^{-1})^T = (A^T)^{-1}$

Производные следа и определителя:

1. $tr(AB)'_A = B^T$
2. $det(A)'_A = det(A)(A^{-1})^T$
3. $(logdetA(x))'_x = tr(A^{-1}A'_x)$

След и определитель:

1. $det(AB) = det(A)det(B)$
2. $det(A^{-1}) = 1/det(A)$
3. $det(A) = \prod_j \lambda_j$
4. $tr(A) = \sum_j A_{jj} = \sum_j \lambda_j$
5. $tr(ABC) = tr(BCA) = tr(CAB)$

Для начала напомним о разнице между одномерными и многомерными случайными величинами.

Обозначим $y$ как случайный вектор $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Одномерную случайную величину будем обозначать маленькими латинскими буквами с индексами: $y_1$.

| Многомерные с.в. | Одномерные с.в. |
|---|---|
| $\mathbb{E}(Ay+b) = A\mathbb{E}(y) + b$<br>$Var(Ay+b) = AVar(y)A^T$<br>$Cov(Ay+b, Cz+d) = ACov(y,z)D^T$ | $\mathbb{E}(Ay_1+b) = A\mathbb{E}(y_1) + b$<br>$Var(ay_1+b) = a^2Var(y_1)$<br>$Cov(a_1y_1+b_1, a_2y_2+b_2) = a_1a_2Cov(y_1,y_2)$ |

$$Var(y) = \begin{pmatrix} Var(y_1) & Cov(y_{1,2}) & \dots Cov(y_{1,n}) \\ \vdots & \vdots & \vdots \\ Cov(y_k, y_1) & Var(y_k) & \dots Cov(y_k, y_n) \\ \vdots & \vdots & \vdots \\ Cov(y_n, y_1) & Cov(y_{n,2}) & \dots Var(y_n) \end{pmatrix}$$

Из такой записи ковариции векторов очевидно, что если в формуле ковариации поменять местами векторы, то их матрица ковариации будет являться транспонированной матрицой ковариации векторов в исходной последовательности. $Cov(y,z) = Cov(z,y)\hat{}T$

Упражнение

Дана матрица $A = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$ и случайный вектор $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ с матожиданием $E(y) = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$ и дисперсией $Var(y) = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$

Требуется найти $E(z), Var(z), Cov(y,z)$.

Решение

1. $E(z) = A \cdot E(y) = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 7 \end{pmatrix} = \begin{pmatrix} 4 \\ 23 \end{pmatrix}$
2. $Var(z) = A \cdot Var(y) \cdot A\hat{}T = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 1 & 7 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 12 & 12 \\ 12 & 72 \end{pmatrix}$
3. $t = Ay = \begin{pmatrix} 2y_1 \\ y_1 + 3y_2 \end{pmatrix}$
4. $Cov(y,z) = \begin{pmatrix} Cov(y_1, z_1) & Cov(y_1, z_2) \\ Cov(y_2, z_1) & Cov(y_1, z_2) \end{pmatrix} = Cov(y, Ay) = Cov(y,y)A\hat{}T = \begin{pmatrix} 3 & 1 \\ 1 & 7 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$

Упражнение

Предположим, существует истинная зависимость $y = X\beta + \varepsilon$ между оцениваемыми величинами. При оценивании параметров модели МНК будут фигурировать следующие величины: $y, \hat{y}, \varepsilon, \hat{\varepsilon}, \beta, \hat{\beta}$. Оценим все матожидания, дисперсии и ковариации указанных величин.

Перед тем, как начать, необходимо сделать оговорку, что существуют две парадигмы исследования: 1) Предполагается, что матрица X является детерминированной. 2) Матрица X состоит из случайных величин.

В ходе нашего курса мы будем работать со случайными X, однако пока будем считать, что матрица X детерминирована.

Решение

Оценка $\hat{\beta} = (X'X)^T Xy$, $\hat{\varepsilon} = y - \hat{y}$, $\hat{y} = X\hat{\beta}$

- Найдем матожидания:

1. $E(\beta) = beta$, так как $\beta$ – вектор неизвестных констант

2. $E(\epsilon) = 0$ (по предпосылкам МНК)

3. $E(y) = E(X\beta + \epsilon) = XE(\beta) + E(\epsilon) = X\beta$

4. $E(\hat{y}) = E(X\hat{\beta}) = XE(\hat{\beta}) = XE((X'X)^{-1}X'y) = X(X'X)^{-1}X'E(y) = X(X'X)^{-1}X'X\beta = X\beta$

5. $E(\hat{\epsilon}) = E(y - \hat{y}) = E(y) - E(\hat{y}) = X\beta - X\beta = 0$

6. $E(\hat{\beta}) = (X'X)^{-1}X'E(y) = (X'X)^{-1}X'X\beta = \beta$

- Найдем, к примеру, $Cov(\epsilon, \hat{\beta})$:

$Cov(\epsilon, \hat{\beta}) = Cov(\epsilon, (X'X)^{-1}X'y) = Cov(\epsilon, (X'X)^{-1}X'(X\beta + \epsilon)) = Cov(\epsilon, (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon) = Cov(\epsilon, \epsilon)((X'X)^{-1}X')' = \sigma^2 I((X'X)^{-1}X')' = \sigma^2 X''((X'X)^{-1})' = \sigma^2 X((X'X)')^{-1} = \sigma^2 X(X'X)^{-1}$

### 4.0.1 Матрицы и Производные

Существует 2 традиции матричного дифференцирования, суть различия которых заключается в представлении вектора (матрицы) производной — в виде столбца или в виде строки. Основные различия представлены в следующей таблице:

| | Традиция 1 | Традиция 2 |
|---|---|---|
| $x$ | вектор размера $n \times 1$ | вектор размера $n \times 1$ |
| $X$ | матрица размера $n \times k$ | матрица размера $n \times k$ |
| $f(x)$ | скаляр | скаляр |
| $g(x)$ | вектор размера $k \times 1$ | вектор размера $k \times 1$ |
| $\frac{\partial f}{\partial x}$ | вектор размера $n \times 1$ | вектор размера $1 \times n$ |
| $\frac{\partial f}{\partial X}$ | матрица размера $n \times k$ | матрица размера $k \times n$ |
| $\frac{\partial g}{\partial x}$ | матрица размера $n \times k$ | матрица размера $k \times n$ |

Таким образом, мы, придерживаясь 1 традиции, будем записывать:

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \qquad \frac{\partial f}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1k}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{n1}} & \cdots & \frac{\partial f}{\partial x_{nk}} \end{pmatrix} \qquad \frac{\partial g}{\partial x} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_k}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \frac{\partial g_k}{\partial x_n} \end{pmatrix} \tag{4.1}$$

### 4.0.2 Свойства матричного дифференцирования

1. Если $x$ – вектор, $f(x) = Ax$, то $\frac{\partial f}{\partial x} = A'$
2. Если $f(x) = x'Ax$, то $\frac{\partial f}{\partial x} = (A + A')x$
3. Если $f(X) = det(X)$, то $\frac{\partial f}{\partial x} = det(X)(X^{-1})'$

Доказательство второго свойства и не только можно почитать здесь

Глава 5

# Доказательство свойств и оценка моделей. {#05_models_evaluation}

конспект: Артём Калинин, Кирилл Улыбин

дата: 9 сентября 2016

### 5.0.1 Пример из домашки

Сначала разобрали пример из домашки прошлого семинара:

Искали

$$cov(\hat{y}, \hat{\epsilon}) = ?$$
$$cov(\hat{y}, \hat{\epsilon}) = cov(X\hat{\beta}, y - \hat{y}) = cov(y - \hat{y}, X\hat{\beta})' = cov((I - \underbrace{X(X'X)^{-1}X'}_{H})y, X(X'X)^{-1}X'y))' =$$
$$= ((I - H)cov(y, y)(H)')' = (\sigma^2(IH - H))' = (\sigma^2(H - H))' = 0$$

Интуитивное объяснение результата - предсказания не должны зависеть от ошибок. Например, если бы предсказания положительно зависели от ошибок, можно было бы сделать поправку в предсказании на известную величину ошибки.

Вспомогательно:

H - матрица-шляпница! Почему?

H*(любой вектор)=(его проекция)

$$Hy = \hat{y}$$

$$H' = H$$

$$H^2 = X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X' = H$$

### 5.0.2 Упражнение 1

Даны модель A: $y_i = \beta x_i + \epsilon_i$, модель B: $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$

$\sum_{i=1}^n x_i = 50;\sum_{i=1}^n x_i y_i = -50;\ \sum_{i=1}^n x_i^2 = 2000;\ \sum_{i=1}^n y_i = 20;\ \sum_{i=1}^n y_i^2 = 500;\ n = 100$

Найти для модели B (для модели A - дома):

$1)X-?\ 2)X'X-?\ 3)\hat{\beta}-?\ 4)Var(\hat{\beta})-?\ 5)X'y-?$

Решение

$$1)X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$2)X' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

$$(X'X) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix}$$

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & x_1 + x_2 + ...x_n \\ x_1 + x_2 + ... + x_n & x_1^2 + x_2^2 + ... + x_n^2 \end{pmatrix} = \begin{pmatrix} 100 & 50 \\ 50 & 2000 \end{pmatrix}$$

$$3)X'y = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} 20 \\ -50 \end{pmatrix}$$

$$4)\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 100 & 50 \\ 50 & 2000 \end{pmatrix}^{-1} = \begin{pmatrix} 20 \\ -50 \end{pmatrix} = \frac{1}{197500} \begin{pmatrix} 2000 & -50 \\ -50 & 100 \end{pmatrix} \begin{pmatrix} 20 \\ -50 \end{pmatrix} =$$
$$\frac{1}{197500} \begin{pmatrix} 42500 \\ -51000 \end{pmatrix}$$

$$5)Var(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 * \frac{1}{197500} \begin{pmatrix} 2000 & -50 \\ -50 & 100 \end{pmatrix}$$

### 5.0.3 Упражнение 2

Доказать, что $Var(Ay) = AVar(y)A'$

$$\text{Напомним, что } Var(z) = \begin{pmatrix} var(z_1) & cov(z1, z2) & \dots & cov(z_1, z_n) \\ \vdots & \vdots & \vdots & \vdots \\ cov(z_n, z_1) & \dots & \dots & var(z_n) \end{pmatrix}$$

Сначала докажем вспомогательные утверждения:

а)$E(Ay) = AE(y)$
б)$E(zB) = E(z)B$

$$\text{в)} Var(z) = E(zz') - E(z)E(z')$$

$$\text{а)} E(Ay) = AE(y)$$
Левая часть: $E(Ay) = Left_{ij} = E(\sum_{k=1}^{s} a_{ik}y_{kj}) = \sum_{k=1}^{s} a_{ik}E(y_{kj})$
Правая часть: $AE(y) = Right_{ij} = \sum_{k=1}^{s} a_{ik}E(y_{kj})$

б)$E(zB) = E(z)B$; Док-во такое же как в (а)

$$\text{в)} Var(z) = E(zz') - E(z)E(z')$$
Левая часть: $Var(z) = Left_{ij} = cov(z_i, z_j)$
Правая часть: $Right_{ij} = E(zz')_{ij} - (E(z)E(z'))_{ij} = E(z_i z_j) - E(z_i)E(z_j) = cov(z_i, z_j)$

$$E(zz') = \begin{pmatrix} E(z_1^2) & E(z_1 z_2) & \dots & E(z_1 z_n) \\ \vdots & \vdots & \vdots & \vdots \\ E(z_n z_1) & \dots & \dots & E(z_n^2) \end{pmatrix}$$

$$E(z)E(z') = \begin{pmatrix} E(z_1)E(z_1) & \dots & E(z_1)E(z_n) \\ \vdots & \vdots & \vdots \\ E(z_n)E(z_1) & \dots & E(z_n)E(z_n) \end{pmatrix}$$

Теперь вернемся к доказательству $Var(Ay) = AVar(y)A'$.

$$Var(Ay) = E(Ay(Ay)') - E(Ay)E((Ay)') = E(Ayy'A') - E(Ay)E(y'A') =$$
$$AE(yy')A' - AE(y)E(y')A' = A\underbrace{(E(yy') - E(y)E(y'))}_{Var(y)}A' = AVar(y)A'$$

### 5.0.4 Упражнение 3

Дано $X_{n\times 1}$, $A_{n\times n}$

$$f(A)_{1\times 1} = X'_{1\times n} A_{n\times n} X_{n\times 1}$$

Найти $\frac{\partial f}{\partial A}$, т.е. скаляр дифференцируем по каждому элементу матрицы А.

Решение:

$$\frac{\partial f}{\partial A}_{ij} = \frac{\partial f}{\partial a_{ij}}$$

$$X'A = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix} =$$
$$\begin{pmatrix} \sum_{i=1}^{n} x_i a_{i1} & \sum_{i=1}^{n} x_i a_{i2} & \dots & \sum_{i=1}^{n} x_i a_{in} \end{pmatrix}_{1\times n}$$

$$X'AX = \begin{pmatrix} \sum_{i=1}^{n} x_i a_{i1} & \sum_{i=1}^{n} x_i a_{i2} & \dots & \sum_{i=1}^{n} x_i a_{in} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{j=1}^{n} (x_j \sum_{i=1}^{n} x_i a_{ij})$$

Тогда можно переписать в виде:

$$f(A) = \sum_{j=1}^{n} (x_j \sum_{i=1}^{n} x_i a_{ij}) = \sum_{i=1,j=1}^{n} x_i x_j a_{ij}$$

То есть $\frac{\partial f}{\partial A_{ij}} = x_i x_j$

$$\frac{\partial f}{\partial A} = \begin{pmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2^2 & \dots & x_2 x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_n x_1 & \dots & \dots & x_n^2 \end{pmatrix} = XX'$$

### 5.0.5 Упражнение 4. Оценим модель с помощью ML

Пусть истинная зависимость $y = X\beta + \varepsilon$, причем $\varepsilon \sim N(0; \sigma^2 I)$

Всего n наблюдений и k регрессоров

Найти:

а) $\hat{\beta_{ML}}$, $\hat{\sigma^2_{ML}}$

б) $E(\hat{\beta_{ML}})$, $E(\hat{\sigma}^2_{ML})$

c) $Var(\hat{\beta_{ML}})$, $Var(\hat{\sigma}^2_{ML})$

Решение

а) Найдем ML оценки

X и y известны, оцениваем $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ и $\sigma^2$

Из $\varepsilon \sim N(0; \sigma^2 I))$ и $y = X\beta + \varepsilon$ следует, что $y \sim N(X\beta; \sigma^2 I)$

Запишем формулу плотности многомерного нормального распределения:

$$p(y) = \frac{1}{\sqrt{(2\pi)^n}\sqrt{det(\sigma^2 I)}} e^{-\frac{1}{2}(y-X\beta)'(\sigma^2 I)^{-1}(y-X\beta)}$$

Для удобства логарифмируем и получим задачу максимизации фукции правдоподобия:

$$Q = ln(p(y)) = -\frac{n}{2}ln(2\pi) - \frac{1}{2}ln(det(\sigma^2 I)) - \frac{1}{2}(y - X\beta)'(\sigma^2 I)^{-1}(y - X\beta) \to \max_{\beta, \sigma^2}$$

Заметим, что первое слагаемое не влияет на решение задачи максимизации, а $det(\sigma^2 I) = \sigma^{2n}$

$\frac{\partial Q}{\partial \beta}$: $Q = -\frac{1}{2\sigma^2}(y' - X'\beta')(y - X\beta) = -\frac{1}{2\sigma^2}(y'y - \beta' X' y - y' X\beta + \beta' X' X\beta)$

Заметим: $y' X\beta$ - скаляр, причем $(y' X\beta)' = \beta' X' y$, тогда можно записать в виде:

$$Q = -\frac{1}{2\sigma^2}(y'y - 2y'X\beta + \beta' X' X\beta)$$

$$\frac{\partial Q}{\partial \beta} = -\frac{1}{2\sigma^2}((-2y'X)' + (X'X + (X'X)^{-1})\hat{\beta}) = 0$$

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

$$\hat{\beta} = (X'X)^{-1}X'y$$

$\frac{\partial Q}{\partial \sigma^2}$: $Q = -\frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$

$$\frac{\partial Q}{\partial \sigma^2} = \frac{n}{\hat{\sigma}^2} - \frac{(y-X\hat{\beta})^{'}(y-X\hat{\beta})}{\hat{\sigma}^2}$$

$$\hat{\sigma}^2 = \frac{(y-X\hat{\beta})^{'}(y-X\hat{\beta})}{n}$$

б) По свойству ML оценок: $E(\hat{\beta}) = \beta$, $E(\hat{\sigma}^2) = \sigma^2$

c) Чтобы найти $Var(\hat{\beta_{ML}})$, $Var(\hat{\sigma}^2_{ML})$ посчитаем вторые производные:

$$\frac{\partial^2 Q}{\partial \beta^2} = -\frac{X^{'}X}{\sigma^2}$$

$$\frac{\partial^2 Q}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{(y-X\hat{\beta})^{'}(y-X\hat{\beta})}{(\sigma^2)^3}$$

$$\frac{\partial^2 Q}{\partial \beta \partial \sigma^2} = -\frac{X^{'}(y-X\hat{\beta})}{(\sigma^2)^2} = -\frac{X^{'}(y-(X^{'}X)^{-1}X^{'}Xy)}{(\sigma^2)^2} = 0$$

$$\text{тогда Var}\begin{pmatrix}\hat{\beta} \\ \hat{\sigma}^2\end{pmatrix} = \begin{pmatrix} -\frac{1}{n}\frac{\partial^2 Q}{\partial \beta^2} & 0 \\ 0 & -\frac{1}{n}\frac{\partial^2 Q}{\partial (\sigma^2)^2}\end{pmatrix}^{-1} = \begin{pmatrix}\frac{n\sigma^2}{X'X} & 0 \\ 0 & 2(\sigma^2)^2\end{pmatrix}$$

### 5.0.6 ДЗ

1) В упражнении 1 сделать те же пункты для модели А
2) Даны $A_{r\times s}$, $B_{s\times r}$
   Записать через эти матрицы ( и их преобразования) сумму: $\sum_{i=1,j=1}^{n} a_{ij}b_{ij} =?$

### 5.0.7 Ссылки

В лекции 5 Бостонского университета подробнее изложено про матрицу-шляпницу и её свойста

Да и в целом курс Бостонского университета хорош :)

# Глава 6

# Статистические свойства RSS

- Дата 10/10/16
- Конспект: Мария Такташева, Алексей Панков

Главный спонсор этого семинара – Алеся Бердникович, которая все решила дома.

объявление публикуется на правах рекламы

## 6.1 Упражнение 1

Найти матожидание и дисперсию оценки параметра $\hat{\beta}$ методом максимального правдоподобия. Для этого построим функцию правдоподобия:

$$Q = \ln p(y) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\det\left(\sigma^2 I\right) - \frac{1}{2}(y - X\beta)'(\sigma^2 I)^{-1}(y - X\beta) \to \max_{\hat{\beta},\hat{\sigma^2}}$$

### 6.1.1 Матожидание оценки

Продифференцируем функцию правдоподобия по $\hat{\beta}$, чтобы найти оценку ML параметра $\beta$

$$\frac{\partial Q}{\partial \hat{\beta}} = \frac{\partial}{\partial \hat{\beta}}\left[-\frac{1}{2}\left(y - X\hat{\beta}\right)'\left(\sigma^2 I\right)^{-1}\left(y - X\hat{\beta}\right)\right] = -\frac{1}{2\sigma^2 I}\frac{\partial}{\partial \hat{\beta}}\left[\left(y' - \hat{\beta}'X'\right)\left(y - X\hat{\beta}\right)\right] = \qquad (6.1)$$

$$= -\frac{1}{2\sigma^2 I}\frac{\partial}{\partial \hat{\beta}}\left[y'y - \hat{\beta}'X'y - y'X\hat{\beta} - \hat{\beta}X'X\hat{\beta}\right] = 0 \qquad (6.2)$$

А теперь заметим, что это выражение состоит из скаляров. И действительно:

- $y'y = y'_{1\times n} \times y_{n\times 1} = y'y_{1\times 1}$
- $\hat{\beta}'X'y = \hat{\beta}'_{1\times k} \times X'_{k\times n} \times y_{n\times 1} = \hat{\beta}'X'y_{1\times 1}$
- $y'X\hat{\beta} = y'_{1\times n} \times X_{n\times k} \times \hat{\beta}_{k\times 1} = y'X\hat{\beta}_{1\times 1}$
- ну и без лишних подробностей $\hat{\beta}'X'X\hat{\beta}_{1\times 1}$

При этом известно, что если матрица $A = A_{1\times 1}$ — скаляр, то $A' = A$. Значит

$$\hat{\beta}' X' y = \left(y' X \hat{\beta}\right)' = y' X \hat{\beta}$$

и выражение выше принимает вид

$$-\frac{1}{2\sigma^2 I} \frac{\partial}{\partial \hat{\beta}} \left[y'y - \hat{\beta}' X' y - 2\hat{\beta} X' X \hat{\beta}\right] = 0$$

Вспомнив некоторые правила матричного дифференцирования, можно прийти к виду:

$$\begin{aligned} -\frac{1}{2\sigma^2 I} \left[-2X'y + X'X\hat{\beta} + (X'X)' \hat{\beta}\right] = \\ = -\frac{1}{2\sigma^2 I} \left[-2X'y + 2X'X\hat{\beta}\right] = 0 \end{aligned} \tag{6.3}$$

Для тех, кто не помнит, как работать с матрицами

$$\frac{\partial}{\partial x} x' A x = (A' + A)$$

$$\frac{\partial}{\partial x} x' A = A$$

$$\frac{\partial}{\partial x} A x = A'$$

---

В итоге мы получаем оценку $\hat{\beta}_{ML} = (X'X)^{-1} X' y$, которая совпадает с оценкой $\hat{\beta}_{OLS}$, построенной методом наименьших квадратов.

### 6.1.2 Дисперсия оценки

До того как мы будем искать производную функции правдоподобия, неплохо бы заметить, что из-за свойств определителя диагональной матрицы

---

$$\det\left(\sigma^2 I\right) = \left(\sigma^2\right)^n$$

Зная этот хинт, можно дифференцировать функцию правдоподобия по $\hat{\sigma^2}$:

$$\frac{\partial Q}{\partial \hat{\sigma^2}} = \frac{1}{2} \left(y - X\hat{\beta}\right)' \left(y - X\hat{\beta}\right) \cdot \frac{1}{(\hat{\sigma^2})^2} - \frac{1}{2} \cdot \frac{n}{\hat{\sigma^2}} = 0 \tag{6.4}$$

$$\left(y - X\hat{\beta}\right)' \left(y - X\hat{\beta}\right) = n\hat{\sigma^2}$$

$$\hat{\sigma^2} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n}$$

Теперь вспоминаем, что $\hat{y} = X\hat{\beta}$, $\hat{\varepsilon} = y - \hat{y}$, а $\hat{\varepsilon}'\hat{\varepsilon} = RSS$, откуда:

$$\hat{\sigma^2} = \frac{RSS}{n}$$

### 6.1.3 Тривиальщина, которая здорово упрощает жизнь

Пусть у нас есть матрица $a_{1\times 1}$, тогда магически

$$a' = a \quad \det(a) = a \quad tr(a) = a$$

Теперь в более общем виде с $Z_{n\times m}$.

1. А правда ли, что
$$\mathbb{E}\left(Z'\right) = \left(\mathbb{E}\left(Z\right)\right)'$$
? Да! Математическое ожидание матрицы — это матрица, в которой от каждого элемента взято математическое ожидание. Транспонирование просто переставляет элементы матрицы, не изменяя их.

2. А это тоже верно?
$$\det\left(\mathbb{E}\left(Z\right)\right) = \mathbb{E}\left(\det\left(Z\right)\right)$$
? Нет. Приведем простой контрпример: пусть $A$ — матрица $2\times 2$, тогда её определитель легко посчитать по формуле
$$\det\left(\mathbb{E}(A)\right) = \det\begin{pmatrix} \mathbb{E}(a_{1,1}) & \mathbb{E}(a_{1,2}) \\ \mathbb{E}(a_{2,1}) & \mathbb{E}(a_{2,2}) \end{pmatrix} = \mathbb{E}(a_{1,1})\mathbb{E}(a_{2,2}) + ... \neq \mathbb{E}\left(a_{1,1}a_{2,2} + ...\right) = \mathbb{E}(\det(A))$$
Математическое ожидание произведения не равно произведению математических ожиданий в общем случае. Однако, это может быть верно, когда элементы матрицы не зависят друг от друга.

3. Может быть
$$tr\left(\mathbb{E}\left(Z\right)\right) = \mathbb{E}\left(tr\left(Z\right)\right)$$
? Точно! Ведь след — это сумма диагональных элементов

Вывод: математическое ожидание любит след и транспонирование

Продолжаем.

$$RSS = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y - X(X'X)^{-1}X'y)'(...),$$

где $H = X(X'X)^{-1}X'$ — “матрица-шляпница” или “hat matrix”

### 6.1.4 Магические свойства “матрицы-шляпницы”

А матрица $X$ у Себера называется «матрица плана»

1. Это матрица проекции.

$$Hy = \hat{y}, \quad H \times \text{любой вектор} = \text{проекция этого вектора на "лапу"}$$

   Напоминаем, что “лапа” — линейная оболочка вектора $X$.

2. Два раза проецировать можно, но результат не изменится :)

$$H^2 y = Hy = \hat{y}$$

3. $$H' = H$$

Возвращаемся к дисперсии оценки $\hat{\beta_{ML}}$

$$RSS = ((I - H)y)'((I - H)y) = y'(I - H)'(I - H)y =$$

---

- $(A - B)' = A' - B' \Rightarrow$

$$= y'(II - HI - IH + H^2)y = y'(I - H)y$$

Поскольку $RSS$ имеет размерность $1 \times 1$, можно перейти к следу, для того, чтобы переставить местами множители

$$\mathbb{E}(RSS) = \mathbb{E}(tr(RSS)) = \mathbb{E}(tr(y'(I - H)y) =$$

---

- $tr(A \cdot B) = tr(B \cdot A), \quad tr(A + B) = tr(A) + tr(B) \Rightarrow$

$$= \mathbb{E}(tr((I - H)y'y) = \mathbb{E}(tr(yy') - tr(Hyy')) = tr(\mathbb{E}(yy')) - tr(\mathbb{E}(Hyy')) = tr(\mathbb{E}(yy')) - tr(H\mathbb{E}(yy')) =$$

---

- $\mathbb{E}(yy')_{n \times n} = ?$, можно найти из $Var(y) = \mathbb{E}(yy') - \mathbb{E}(y)\mathbb{E}(y')$
- $Var(y) = \sigma^2 I$
- $\mathbb{E}(y)\mathbb{E}(y') = X\beta\beta'X'$, т.к. $y = X\beta + \varepsilon, \mathbb{E}(y) = \mathbb{E}(X\beta) + \mathbb{E}(\varepsilon) = X\beta$
- $\Rightarrow \mathbb{E}(yy') = \sigma^2 I + (X\beta\beta'X')_{n \times n}$

$$= tr((I - H)(\sigma^2 I + X\beta\beta'X') = tr(\sigma^2(I - H)) =$$

- $(I-H)(X\beta\beta'X') = X\beta\beta'X' - X\beta\beta'X' = 0$

  Хорошее свойство следа: $tr(X) = \sum_i \lambda_i$ — равен сумме собственных чисел матрицы

$$= \sigma^2(\lambda_1 + ... + \lambda_n) =$$

### 6.1.5 Собственные числа

Для матрицы $I - \underbrace{1...1}_{\text{n штук}}$.

Для матрицы $H - \underbrace{1...1}_{\text{k штук}}\ \underbrace{0...0}_{\text{n-k штук}}$. Множество собственных чисел устроено так, поскольку

$$(I-H)v = \begin{cases} 1 \cdot v \text{ для перпендикуляров лапе} \\ 0 \cdot v \text{ для лежащих в лапе} \end{cases}$$

$$= (n-k)\sigma^2 = E(RSS)$$

Наконец-то мы посчитали $E(RSS)$. Теперь ясно, что оценка дисперсии случайной ошибки $\varepsilon_i$ методом максимального правдоподобия — смещенная!

$$\mathbb{E}(\hat{\sigma}^2_{ML}) = \frac{n-k}{n}\sigma^2 \neq \sigma^2$$

Зато можно построить несмещенную

$$\hat{\sigma}^2_{\text{скорр}} = \frac{RSS}{n-k}$$

## 6.2 Теорема

Если $M$ — проектор — выполняет проецирование ($M' = M$, $M^2 = M$) и вектор $u \sim \mathbb{N}(0, I)$, то $u'Mu \sim \chi^2_{rk(M)}$. Для проектора $rk(M) = tr(M)$ — количество линейно независимых собственных векторов (или просто столбцов в матрице $M$)

## 6.3 Домашка

1. Найти как распределено $\dfrac{RSS}{\sigma^2}$

2. Посчитать руками $H$, $I-H$, $rk(H)$, $rk(I-H)$, $tr(H)$, $tr(I-H)$, найти закон распределения $\frac{\varepsilon' H\varepsilon}{\sigma^2}$, $\frac{\varepsilon'(I-H)\varepsilon}{\sigma^2}$, $\frac{y'(I-H)y}{\sigma^2}$ и $\frac{RSS}{\sigma^2}$ если известно, что

$$X = \begin{pmatrix} 1 & 0.1 \\ 1 & 0.2 \\ 1 & 0.3 \end{pmatrix}, \quad y = X\beta + \varepsilon, \quad \varepsilon \sim \mathbb{N}(0, \sigma^2 I)$$

Глава 7

# Подготовка к празднику №2

- Дата: 31.11.2016
- Авторы: Герман Никита, Ишмаева Бэлла

## 7.1 Упражнение 1

Найти как распределена случайная величина:

$$\frac{RSS}{\sigma^2} \sim ?$$

Для этого надо вспомнить теорему:

Если одновременно выполнено:

- $H$ — проектор: $H^2 = H$, $H^T = H$
- $u \sim N(0; I)$

То:

$u^T \cdot H \cdot u \sim \chi^2_k$, где $k$ — это размерность пространства, куда $H$ проецирует.

Вспомним, как распределена случайная ошибка в модели регрессии:

$$\varepsilon \sim N(0; \sigma^2 I)$$

При этом:

$$\frac{RSS}{\sigma^2} = \frac{\hat{\varepsilon}^T \cdot \hat{\varepsilon}}{\sigma^2}$$

Теперь вспомним, что такое наблюдаемая ошибка:

$$\hat{\varepsilon} = y - \hat{y} = y - Hy = (I - H)y = (I - H)(X\beta + \varepsilon) = X\beta - HX\beta + (I - H)\varepsilon$$

Вспомним, что $H = X(X^T X)^{-1} X^T$

Также вспомним, что матрица H проецирует любой вектор на линейную оболочку, порожденную X. Тогда проекция X будет тоже X. То есть $HX = X$. Тогда $\hat{\varepsilon} = (I - H)\varepsilon$. Отсюда:

$$\frac{RSS}{\sigma^2} = \frac{\varepsilon^T (I-H)^T \cdot (I-H)\varepsilon}{\sigma^2}$$

Вспомним, что $(I-H)^T = (I-H)$ и $(I-H)^2 = (I-H)$. Тогда:

$$\frac{RSS}{\sigma^2} = \frac{\varepsilon^T (I-H)\varepsilon}{\sigma^2} = u^T (I-H) u,$$

где $u = \frac{\varepsilon}{\sigma^2}$, $u \sim N(0; I)$.

Вспоминая теорему и, используя свойство, что $\text{rank}(I-H) = tr(I-H) = n-k$, получаем:

$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-k}$$

## 7.2 Упражнение 2

Необходимо посчитать:

$$E\left(y^T \cdot \frac{1}{n} \cdot S \cdot y\right)$$

Если известно, что y задается регрессионной моделью:

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N(0; \sigma^2 I)$$

S — матрица строевого леса:

$$S = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

Заметим, что $(y^T \cdot \frac{1}{n} \cdot S \cdot y)$ имеет размерностнь 1 x 1. Возьмем след, потому что след матрицы 1 x 1 равняется самой матрице:

$$tr\left(E(y^T \cdot \frac{1}{n} \cdot S \cdot y)\right) = E\left(tr(y^T \cdot \frac{1}{n} \cdot S \cdot y)\right) = E\left(\frac{1}{n} S \cdot tr(y^T y)\right) = tr\left(\frac{1}{n} S \cdot E(y^T y)\right)$$

Вспомним, что:

$$E(yy^T) = Var(y) + E(y)E(y^T)$$

Тогда:

$$tr\left(E(y^T \cdot \frac{1}{n} \cdot S \cdot y)\right) = \frac{1}{n} tr\left[S(\sigma^2 I + X\beta\beta^T X^T)\right] = \frac{1}{n} tr(S\sigma^2 I) + \frac{1}{n} tr(SX\beta\beta^T X^T)$$

Заметим, что след матрицы $S\sigma^2 I$ равен $n\sigma^2$. Также заметим, что $SX\beta\beta^T X^T$ - скаляр. Отсюда:

$$tr\left(E(y^T \cdot \frac{1}{n} \cdot S \cdot y)\right) = \sigma^2 + \frac{1}{n} SX\beta\beta^T X^T$$

## 7.3 Подготовка к контрольной!

#### 7.3.0.1 Сначала вспомним модель парной регрессии

Общая формула нахождения оценок $\beta$ для множественной регрессии:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

А для парной:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \text{ где}$$

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_{i^2} \end{pmatrix}$$

$$X^T y = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

В этом случае можно воспользоваться шаманским способом обращения матриц:

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{detA} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

#### 7.3.0.2 А теперь перейдем к задачам про доверительный интервал

Когда мы решали подобного рода задачи в прошлом году, мы обычно находили статистику и сравнивали ее с критическим значением нормального распределения, распределения Стьюдента или Фишера, в зависимости от того, как распределена была сама статистика. В нашем случае нам нужно проверять гипотезы и строить доверительные интервалы для $\beta$, а значит нужно понять, какое распределение имеет статистика:

$$\frac{\hat{\beta} - \beta}{se\left(\hat{\beta}\right)}$$

Если с числителем еще более или менее понятно: $\hat{\beta}$ имеет нормальное распределение, то вот со знаменателем стоит разобраться, что мы и будем делать дальше.

В домашнем задании мы уже выводили, что:

$$\begin{cases} Cov\left(\hat{\varepsilon}, \hat{\beta}\right) = 0 \\ Cov\left(\hat{\beta}\right) = \sigma^2 \left(X^T X\right)^{-1} \end{cases}$$

А скорректированная $ML$ оценка для сигмы-квадрат :

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

Стоит отметить, что в начале семинара мы уже показали, что, если отмасштабировать, $RSS$ будет иметь распределение $\chi^2_{n-k}$, которое здесь делится на количество степеней свободы $n-k$ (должно напоминать нам $t-$ или $F-$ распределение)

Теперь вернемся к условию $Cov(\hat{\varepsilon}, \hat{\beta}) = 0$: вообще говоря, равенство нулю ковариации не говорит нам о независимости случайных величин, но в нашем случае $\hat{\epsilon}$ и $\hat{\beta}$ имеют совместное нормальное распределение, а значит мы можем сказать, что они независимы.

Имеем:

$$Var\left(\hat{\beta}\right) = \sigma^2\left(X^TX\right)^{-1} \hat{Var}\left(\hat{\beta}\right) = \hat{\sigma^2}\left(X^TX\right)^{-1} \hat{Var}\left(\hat{\beta}\right) = \left(\frac{RSS}{n-k}\right)\left(X^TX\right)^{-1}$$

Матрица $\hat{Var}\left(\hat{\beta}\right)$ имеет вид:

$$\begin{pmatrix} \hat{Var}\left(\hat{\beta}_1\right) & \cdots & \hat{Cov}\left(\hat{\beta}_n, \hat{\beta}_1\right) \\ \vdots & \ddots & \vdots \\ \hat{Cov}\left(\hat{\beta}_1, \hat{\beta}_n\right) & \cdots & \hat{Var}\left(\hat{\beta}_n\right) \end{pmatrix}$$

А каждый элемент матрицы мы можем посчитать исходя из данных.

Вспомним, как выглядит распределение Стьюдента:

$$\frac{N(0,1)}{\sqrt{\frac{\chi^2_r}{r}}} \sim t_r$$

Тогда мы можем получить следующую теорему:

Если:

- $Y = X\beta + \varepsilon$
- $\varepsilon \sim N\left(0, \sigma^2 I\right)$

То:

$$\frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \sim t_{n-k},$$

где $se = \sqrt{\hat{Var}\left(\hat{\beta}_j\right)}$ — это стандартная ошибка.

## 7.4 Упражнение 3

Дано:

$\sum_i x_i = 5$, $\sum_i y_i = 2$, $\sum_i x_i y_i = 20$, $\sum_i x_i^2 = 10$, $\sum_i y_i^2 = 80$, $n = 5$

$$y_i = \beta x_i + \varepsilon_i$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

Найти:

a) $X^TX$, $X^Ty$

b) $\hat{\beta}$, $RSS$, $\hat{\epsilon}^2$, $se\left(\hat{\beta}\right)$

c) Построить 95% доверительный интервал для $\hat{\beta}$ при $\alpha = 0.05$ и проверить гипотезу:

$$H_0 : \beta = 0$$
$$H_a : \beta \neq 0$$

Решение:

a) $X^TX = \sum_i x_i^2 = 10$, $X^Ty = \sum_i x_iy_i = 20$

b) $\hat{\beta} = (X^TX)^{-1}X^Ty = (10^{-1}20) = 2$. Для нахождения $RSS$, нужно вспомнить, что:

$$H = X(X^TX)^{-1}X^T$$

$$RSS = y^T\left(I - H\right)y = y^TIy - y^THy = y^Ty - y^TX\left(X^TX\right)^{-1}X^Ty = \sum_i y_i^2 - \sum_i x_iy_i\left(\sum_i x_i^2\right)^{-1}\sum_i x_iy_i$$

$$RSS = 80 - 20\frac{1}{10}20 = 40$$

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{40}{4} = 10$$

$$\hat{Var}\left(\hat{\beta}\right) = \hat{\sigma}^2\left(X^TX\right)^{-1} = 10\frac{1}{10} = 1$$

$$se\left(\hat{\beta}\right) = \sqrt{\hat{Var}\left(\hat{\beta}\right)} = 1$$

c) В общем случае доверительный интервал выглядит так:

$$\beta \in \left[\hat{\beta} - t_{crit}\cdot se\left(\hat{\beta}\right); \hat{\beta} + t_{crit}\cdot se\left(\hat{\beta}\right)\right]$$

Находим в таблице критическое значение для t-распределения с $(n-1) = 4$ степенями свободы и 5% уровнем значимости: $t_{crit} = 2,77$.

Тогда наш интервал выглядит следующим образом:

$$\beta \in [2 - 2,77; 2 + 2,77]$$

Чтобы проверить гипотезу, найдем $t$ наблюдаемое:

$$\frac{\hat{\beta} - \beta}{se\left(\hat{\beta}\right)} = \frac{2-0}{1} = 2$$

Получаем, что $H_0$ не отвергается.

## 7.5 ДЗ

Демоверсия и разделы 3-4

# Глава 8

# Гипотезы о нескольких ограничениях в регрессии

- Дата ??/??/16
- Конспект: Арсений Лысенко

## 8.1 Примеры:

$$H_0 : \beta_2 = 0$$

$$H_0 : \beta_2 = \beta_7$$

$$H_0 : \begin{cases} \beta_2 = \beta_7 \\ \beta_2 = \beta_3 \end{cases}$$

$$H_0 : \begin{cases} \beta_1 = 1 \\ \beta_2 = 1 \\ \beta_3 = 2 \end{cases}$$

## 8.2 Проверка гипотез по шагам:

1. Строим регрессию, забыв про ограничения; Получим $RSS_{UR} - RSS_{\text{Unrestricted}}$
2. Строим регрессию с учетом ограничений.
   Получим $RSS_R - RSS_{\text{Restricted}}$

## 8.3 Замечание:

МНК минимизирует RSS, поэтому безусловный RSS($RSS_{\text{UR}}$) будет меньше или равен условного RSS($RSS_{\text{R}}$).

## 8.4 Теорема:

Если выполнены предпоссылки теоремы Гаусса-Маркова, $H_0$ верна ии $\epsilon \sim N(0; \sigma * I)$, тогда:

$$\text{F} = \frac{(RSS_R - RSSUR)/(\text{кол-во ограничений})}{RSS_{UR}/(n - k_{UR})} \; F \sim F_{\text{кол-во ограничений};n-k_{UR}}$$

где $k_{UR}$ - колличество коэффициентов в неограниченной модели.

## 8.5 Упражнение:

Харис пытается понять, что лучше помогает решать задачи по эконометрике:
-поедание пирожков(штуки)
-посещение лекций(академические часы)

$$problems_t = \beta_1 + \beta_2 lecture_t + \beta_3 pie_t + u_t$$

$$H_0 : \beta_2 = \beta_3$$

Какую регрессию нужно оценить, чтобы найти $RSS_R$?

### 8.5.1 Решение

Согласно $H_0$ должно выполняться: $\beta_2 = \beta_3$. Тогда:

$$problem_i = \beta_1 + \beta_2(lecture_i + pie_i) + u_i = \beta_1 + \beta_2 lp_i + u_i$$

где $lp_i = lecture_i + pie_i$

А что, если Харис захочет проверить гипотезу о постоянной отдаче от масштабов? Как тогда будут выглядеть $H_0$ и ограниченная регрессия?

UR: $lnproblem_i = \gamma_1 + \gamma_2 \ln lecture_i + \gamma_3 \ln pie_i + u_i$

$$problem_i = e^{\gamma_1} * lecture_i^{\gamma_2} * pie_i^{\gamma_3} * e^{u_i}$$

$$H_0 = \gamma_1 + \gamma_3 = 1$$

R: $lnproblem_i = \gamma_1 + \gamma_2 \ln lecture_i + (1 - \gamma_2) \ln pie_i + u_i$

После преобразований получим:

$$(\ln problem_i - \ln pie_i) = \gamma_1 + \gamma_2(\ln lecture_i - \ln pie_i) + u_i$$

Введём новые переменные:

$$\tilde{y_i} = \ln problem_i - \ln pie_i$$

$$\tilde{x_i} = \ln lecture_i - \ln pie_i$$

Получим:

$$\tilde{y_i} = \gamma_1 + \gamma_2 \tilde{x_i} + u_i$$

## 8.6 Упражнение:

10 наблюдений

$$UR : RSS = 50R^2 = 0,3$$

$$problem_i = \beta_1 + \beta_2 lecture_i + \beta_3 pie_i + u_i$$

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

$H_A:$ хотя бы одна из $\beta_2$ и $\beta_3 \neq 0$
а) Как выглядит ограниченная регрессия?
Чему равен $RSS_R$?
б) Как выглядит $F$?
Проверить $H_0$ на 5% уровне значимости.

### 8.6.1 Решение:

а) $problem_i = \beta_1 + \mu_i$
$ESS = 0$
$TSS = RSS$
$R^2 = \frac{ESS}{TSS}$

$$R^2_{UR} = 0,3 = \frac{TSS_{UR} - 50}{TSS_{UR}} TSS_{UR} = 500/7 = 71$$

$problem_i = \beta_1 + \mu_i$
Следовательно, $TSS_R = TSS_{UR} = RsS_R = 71$
б) $F = \frac{(71-50)/2}{50/(10-3)} = \frac{10,5}{7} = 1,4$
$F_{cr} = 4,7$. Получается, что основная гипотеза не отвергается.

## 8.7 Упражнение:

Пусть Харис решил заново оценить модель после второго модуля чтобы понять, изменилась ли зависимость.
В первом модуле было 10 наблюдений. Во втором модуле было 8 наблюдейний.
Кроме того, известно:
по двум модулям: $RSS = 150$
по первому модулю: $RSS = 50$
по второму модулю: $RSS = 70$
$H_0$: зависимость не и зменилась
$H_A$: зависимость изменилась, но осталась линейной
а) Как выглядят ограниченная и неограниченная регрессии?
б) $RSS_{UR}-?$
$RSS_R-?$
в) Проверить гипотезу $H_0$.

### 8.7.1 Решение:

I модуль: $\beta_1, \beta_2, \beta_3$
II модуль: $\gamma_1, \gamma_2, \gamma_3$

$$H_0 = \begin{cases} \beta_1 = \gamma_1 \\ \beta_2 = \gamma_2 \\ \beta_3 = \gamma_3 \end{cases}$$

Ограниченная модель строится по всем наблюдениям(по 18), то есть $RSS_R = 150$.
Теперь рассмотрим неограниченную модель.

$$X = \begin{pmatrix} 1 & lecture_1 & pie_1 \\ \vdots & \vdots & \vdots \\ 1 & lecture_n & pie_n \end{pmatrix}$$

$$y_I = X_I \cdot \beta + u_I$$

$$y_{II} = X_{II} \cdot \beta + u_I$$

$$\begin{pmatrix} y_I \\ y_{II} \end{pmatrix} = \begin{pmatrix} X_I & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & X_{II} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} u_I \\ u_{II} \end{pmatrix}$$

$$y_i = \beta_1 \cdot m_i^1 + \beta_2 \cdot lect_i \cdot m_i^1 + \beta_3 \cdot pie_i \cdot m_i^1 + \gamma_1 \cdot m_i^2 + \gamma_2 \cdot lect_i \cdot m_i^2 + \gamma_3 \cdot pie_i \cdot m_i^2$$

$$m^2 = 1 - m^1$$

$$=> RSS_{UR} = RSS_1 + RSS = 50 + 70 = 120$$

- Дата 21/11/16
- Конспект: Козловский Евгений, Ермакова Мария

## Глава 9

# Фриш и Ко

Вспомним теорему Фриша-Вау-Ловелла. Допустим, имеется следующая модель:

$$y_i = \beta_1 x_i + \beta_2 z_i + u_i$$

Заметим сначала, что вышеупомянутую теорему можно применять вне зависимости от того, есть ли в модели константа или нет. Мы хотим оценить коэффициент $\beta_1$. Стандартный способ — построить регрессию вида:

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

исходя из которой мы легко с помощью МНК можем вычислить $\hat{\beta}_1$. Теорема имени Фриша-Вау-Ловелла гласит, что есть и другой путь, состоящий из нескольких шагов:

1) строим регрессию $\hat{y}_i = \hat{\lambda} z_i$, вычисляем остатки $\tilde{y}_i = y_i - \hat{y}_i$
2) строим регрессию $\hat{x}_i = \hat{\delta} z_i$, вычисляем остатки $\tilde{x}_i = x_i - \hat{x}_i$
3) теперь нам остается оценить модель $\tilde{y}_i = \alpha \tilde{x}_i + \varepsilon_i$, т.е. нужно построить регрессию остатков $\tilde{y}_i$ на остатки $\tilde{x}_i$: $\hat{\tilde{y}}_i = \hat{\alpha} \tilde{x}_i$. Полученная оценка $\hat{\tilde{\alpha}}$ будет в точности совпадать с $\hat{\beta}_1$ из исходной регрессии!

Проведем аналогию

$$y_i = \beta x_i + u_i$$

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$y_i = \beta_1 x_i + \beta_2 z_i + u_i$$

$$\hat{\beta} = \frac{\sum \tilde{x}_i \tilde{y}_i}{\sum \tilde{x_i^2}}$$

,где $\tilde{y}_i$ и $\tilde{x}_i$ остатки от регрессий $y_i$ на $z_i$ и $x_i$ на $z_i$

## 9.1 Упражнение 1

В качестве примера рассмотрим весьма популярный массив mtcars, встроенный в R, из которого возьмем следующие характеристики автомобилей:

- $mpg$ — количество миль на галлон бензина
- $wt$ — масса автомобиля
- $hp$ — мощность (в л.с.) Пусть имеются следующие модели с соответствующими остатками:

$$mpg_i = \beta_1 + \beta_2 hp_i + u_i; \widetilde{mpg}_i = mpg_i - \widehat{mpg_i}$$

$$wt_i = \gamma_1 + \gamma_2 hp_i + u_i; \widetilde{wt}_i = wt_i - \widehat{wt_i}$$

Пусть также дана ковариационная матрица остатков:

$$\widehat{Var}\begin{pmatrix}\widetilde{mpg}_i \\ \widetilde{wt}_i\end{pmatrix} = \begin{pmatrix}14.4 & -2.10 \\ -2.10 & 0.542\end{pmatrix}$$

Найдем всевозможные коэффициенты в следующих регрессиях:

$$mpg_i = \alpha_1 + \alpha_2 wt_i + \alpha_3 hp_i + \upsilon_i$$

$$wt_i = \delta_1 + \delta_2 mpg_i + \delta_3 hp_i + \varepsilon_i$$

Чтобы отыскать оценку $\alpha_2$, нужно построить регрессию $\widetilde{mpg}_i = \alpha_2 \widetilde{wt}_i$

Можем воспользоваться ковариационной матрицей для нахождения оценки коэффициента $\alpha_2$:

$$\widehat{\alpha_2} = \frac{\sum \widetilde{mpg}_i \widetilde{wt}_i/(n-1)}{\sum \widetilde{wt_i^2}/(n-1)} = \frac{-2.10}{0.542} = -3.87454$$

По такой же схеме можем найти $\hat{\delta}_2$:

$$\widehat{\delta}_2 = \frac{\sum \widetilde{mpg}_i \widetilde{wt}_i/(n-1)}{\sum \widetilde{mpg_i^2}/(n-1)} = \frac{-2.10}{14.4} = -0,14583$$

## 9.2 Упражнение 2

В глубокий тыл противника заброшен майор Пейн. Оказалось, что ему под силу в уме оценивать $\widehat{\beta}$ для регрессий типа $y_i = \beta x_i + u_i$, а также он что-то слышал про теорему Фриша-Вау-Ловелла.

Что мы хотим? Узнать, сколько нужно построить регрессий, чтобы найти все коэффициенты для:

а) $y_i = \gamma_1 z_i + \gamma_2 x_i + u_i$

б) $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + u_i$

Для начала разберемся с пунктом а) $y_i = \gamma_1 z_i + \gamma_2 x_i + u_i$

Оценим $\gamma_2$ (найдем $y_i$ на $x_i$) :

1) Очистим y от z
2) Очистим x от z
3) Строим регрессию очищенного y на очищенный x

Оценим $\gamma_1$

1) Очистим y от x
2) Очистим z от x
3) Строим регрессию очищенного y на очищенный z

Итого необходимо $3 + 3 = 6$ регрессий.

Ну а теперь пункт б:

Для оценки $\beta_1$: (строим тройные регрессии)

1) $y_i$ на $x_i, z_i, w_i$ (k штук)
2) 1 на $x_i, z_i, w_i$ (k штук)
3) $\tilde{y_i}$ на $\tilde{1}$

Итого : $2k + 1$ регрессий для оценивания одного коэффициента, следовательно для оценивания всех коэффициентов понадобится $4(2k + 1)$ однофакторных регрессий, где $k$ — это количество регрессий на одну переменную, необходимых для того, чтобы узнать оценки всех коэффициентов трехфакторной модели. Заметим, что если за $m$ обозначить количество однофакторных регрессий, которые требуется построить, чтобы рассчитать оценки всех коэффиентов трехфакторной модели, то будет выполняться равенство: $k = 3(2m + 1) = 3(2 \times 6 + 1) = 39$. Следовательно $4(2k + 1) = 4(2 \times 39 + 1) = 316$.

На самом деле Фриш и Ко окружают нас повсюду! Попробуем узреть их теорему в стандартной формуле для $\hat{\beta}_2$ в модели $y_i = \beta_1 + \beta_2 x_i + u_i$:

$$\hat{\beta}_2 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

Но, как известно, $\overline{x}$ и $\overline{y}$ — это $\hat{\alpha_1}$ и $\hat{\alpha_2}$ в регрессиях $x_i = \alpha_1$ и $y_i = \alpha_2$ соответственно, а $x_i - \overline{x}$ и $y_i - \overline{y}$ — это остатки. Таким образом, оценку $\hat{\beta}_2$ можно также получить в результате регрессии остатков $\tilde{y_i}$ на остатки $\tilde{x_i}$

## 9.3 Домашка

4.51, 4.49, 4.47, 4.40, 4.45, 4.46, 4.43, 3.61, 3.60, 3.48, 3.49, 3.47

# Глава 10

# Стохастические бобры и прочие условности {#12_everything_about_beavers}

Дата: 05.12.2016

Авторы: Уманец Екатерина, Купцова Анастасия

## 10.1 Упражнение № 0

Дано

| u/x | 2 | 3 |
|---|---|---|
| -1 | 0.1 | 0.2 |
| 1 | 0.1 | 0.6 |

$u$, $x$ - скалярные случайные величины

Хотим найти

- $\mathbb{E}(u|x)$
- $\mathbb{V}ar(u|x)$

Решение

$$E(u|x=2) = 0,5*(-1) + 0,5*1 = 0$$

$$E(u|x=3) = -1*0,25 + 1*0,75$$

$$E(u|x) = \begin{cases} 0, & \text{если } x = 2 \\ 0,5, & \text{если } x = 3 \end{cases} = 0.5(x-2)$$

$$Var(u|x) = E(u^2|x) - E^2(u|x)$$

$$E(u^2|x) = 1 \Rightarrow$$

$$Var(u|x) = \begin{cases} 1/3, x = 2 \\ 3/4, x = 3 \end{cases} = 1 - 0,25(x-2)^2$$

## 10.2 Условные свойства

1. $E(f(x)|x) = f(x)$
2. $Var(f(x)|x) = 0$
3. $E(f(x)y|x) = f(x)E(y|x)$
4. $E(E(y|x)) = E(y)$
5. $Var(y) = Var(E(y|x)) + E(Var(y|x))$

## 10.3 Упражнение № 1

Дано:

$x_1, x_2...$

$x_i \sim N(10, 9)$ - независимы

$\underbrace{\overbrace{(x_1, u_1)}^{can\ be\ dependent} \quad (\mathrm{x}_2, \mathrm{u}_2)}_{independent}$... - независимы и одинаково распеделены

$$X = \begin{pmatrix} x_1 \\ : \\ x_n \end{pmatrix}, u = \begin{pmatrix} u_1 \\ : \\ u_n \end{pmatrix}$$

Хотим найти

a) $plim(\frac{1}{n}X'X)^{-1}$

b) $plim(\frac{1}{n}X'u)$

c) $plim(X'X)^{-1}X'u$

Решение

ЗБЧ:

$Y_1, .., Y_n$ - независимы и одинаково распределены

$\bar{Y_n} \longrightarrow E(Y_1)$

a) $plim(\frac{1}{n}\sum_1^n x_i^2)^{-1} = (plim(\frac{1}{n}\sum_1^n x_i^2))^{-1} = E(x_1^2)^{-1} = (9+100)^{-1} = \frac{1}{109}$

b) $plim(\frac{1}{n}X'u) = [\frac{X'u}{n} = \frac{\sum(x_i u_i)}{n}] = E(x_i, u_i) = E(E(x_i, u_i|x)) = E(x_i \underbrace{E(u_i|x_i)}_{=0}) = 0$

c) $plim(X'X)^{-1}X'u = plim(\frac{1}{n}X'X)^{-1}\frac{1}{n}X'u = plim(\frac{1}{n}X'X)^{-1} * plim(\frac{1}{n}X'u) = 109^{-1} * 0 = 0$

## 10.4 Случайность? Не думаю! Или история о том, как отличить стохастические '{и}ксы

Будем исследовать такой воспрос: как количество выпитого кофе влияет на производительность Бориса.

- Эксперимент №1 (неслучайные иксы): пригласим 100 рандомных Борисов, попросим Бориса номер один в первый день выпить одну кружку кофе, Бориса номер два во второй - две, и так далее, скажем, сто дней; соответственно, на сотый день сотый Борис будет пить сто кружек кофе; и посмотрим, сколько брутальных задачек по эконометрике каждый Борис сможет решить в каждый из этих дней. Внимание: В данном экмперименте ни один Борис не пострадал!!!
- Эксперимент №2 (стохастические иксы): поймаем 100 рандомных Борисов на улице и спросим, сколько кружек кофе каждый из них пьет и сколько брутальных задач решает за день.

Главное отличие между этими двумя экспериментами заключается в том, что в первом случае мы сами выбирали количество кружек кофе, а во втором эта величина получалась случайно.

## 10.5 Теорема (как на все это смотрели Гаусс и Марков)

Если:

- $y = X\beta + u$ - наша регрессионная модель со стохастическими иксами
- $(\chi_{i|...}, y_i)$ - независимы и одинаково распределены (то есть между парами зависимостей нет, а вот внутри пары - угадайте что:) ), где $\chi_{i|...}$ - $i$-ая строка матрицы $X$
- $E(u_i|\chi_{i|...}) = 0$
- $Var(u_i|\chi_{i|...}) = \sigma^2$ - условие гомоскедастичности
- $P$(столбцы матрицы $X$ линейно независимы) $= 1$

Тогда:

- $E(\hat{\beta}|X) = \beta$ и $E(\hat{\beta}) = \beta$
- $Var(\hat{\beta}|X) = \sigma^2(X^TX)^{-1}$
- $\hat{\beta}$ линейна по $y$
- $\hat{\beta}$ эффективна среди линейных по $y$ и несмещенных оценок
- $plim(\hat{\beta}) = \beta$

## 10.6 А что если гетероскедастичность?

Пусть выполняются все предположения предыдущей теоремы, кроме одного - гомоскедастичности. То есть теперь $Var(u_i|\chi_{i|\ldots}) = f(\chi_{i|\ldots})$ - условие гетероскедастичности.

Тогда хотим найти:

1. $\Omega = Var(u|X)$
2. $E(\hat{\beta}|X)$
3. $Var(\hat{\beta}|X)$
4. $plim(\hat{\beta})$

Решение:

1. $\Omega = Var(u|X) = \begin{pmatrix} f(\chi_{1|\ldots}) & \cdots & 0 \\ \vdots & \ldots & \vdots \\ 0 & \cdots & f(\chi_{n|\ldots}) \end{pmatrix}$

Теперь поймем, откуда это взялось:

- во-первых, $Var(u_i|X) = Var(u_i|\chi_{i|\ldots}) = f(\chi_{i|\ldots})$ - элементы на диагонали матрицы $\Omega$
- во-вторых, $Cov(u_i, u_j|X) = 0$ - элементы вне диагонали матрицы $\Omega$

2. $E(\hat{\beta}|X) = \beta$

Действительно:

$$E(\hat{\beta}|X) = E((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^TE(y|X)$$

Так как $E(y|X) = E(X\beta + u|X) = X\beta + E(u|X) = X\beta + 0$ (Note: $E(u|X) = 0$ следует из предпосылок теоремы, а именно $E(u_i|\chi_{i|\ldots}) = 0$),

то $E(\hat{\beta}|X) = (X^TX)^{-1}X^TE(y|X) = (X^TX)^{-1}X^TX\beta = 1 * \beta = \beta$

3. $Var(\hat{\beta}|X) = (X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$

Докажем, что на самом деле получается именно такая сэндвич-формула:

$$Var(\hat{\beta}|X) = Var((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^TVar(y|X)((X^TX)^{-1}X^T)^T = (X^TX)^{-1}X^TVar(y|X)X(X^TX)^{-1}$$

Так как $Var(y|X) = Var(X\beta + u|X) = Var(u|X) = \Omega$,

то $Var(\hat{\beta}|X) = (X^TX)^{-1}X^TVar(y|X)X(X^TX)^{-1} = (X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$

4. $plim(\hat{\beta}) = \beta$

Докажем это:

$$plim(\hat{\beta}) = plim((X^TX)^{-1}X^Ty) = plim((X^TX)^{-1}X^T(X\beta+u)) = plim((X^TX)^{-1}X^TX\beta)+plim((X^TX)^{-1}X^Tu$$
$$= \beta + plim((X^TX/n)^{-1}X^Tu/n) = \beta + plim(X^TX/n)^{-1} \cdot plim(X^Tu/n) = \beta + const \cdot 0 = \beta \tag{10.1}$$

Эти пределы мы уже искали в упражнении 1 :)

Как Уайт предлагает оценивать дисперсию оценки $\beta$ :

$$\hat{Var}(\hat{\beta}|X) = (X^T X)^{-1} X^T \hat{\Omega} X (X^T X)^{-1}$$

$$\text{где } \hat{\Omega} = \begin{pmatrix} \hat{u}_1^2 & \cdots & 0 \\ \vdots & \dots & \vdots \\ 0 & \cdots & \hat{u}_n^2 \end{pmatrix}$$

## 10.7 В чем прелесть гетероскедастичности?

А вот в чем - зная о гетероскедастичности, мы знаем распределение:

$\frac{\hat{\beta}_j - \beta_j}{SE_{HC}(\hat{\beta}_j)} \sim N(0,1)$ - асимптотически, конечно!

# Глава 11

# Упражнения на гетероскедастичость

## 11.1 Упражнение 0

Даны две случайные величины: $u, x$ - случайные величины

(матрица)

$E(u|x) =?$

Посчитаем условные математические ожидания для каждой случайной величины:

$$E(u|x = 2) = -1 * 0.5 + 1 * 0.5 = 0$$

$$E(u|x = 3) = -1 * \frac{0.2}{0.8} + 1 * \frac{0.6}{0.8} = 0.5$$

Совместим их:

$$E(u|x) = \begin{cases} 0 & \text{если x=2} \\ 0.5 & \text{если x=3} \end{cases}$$

Также можем представить это условное математическое ожидание в виде функции

$$E(u|x) = 0.5(x - 2)$$

Значение условной дисперсии $Var(u|x)$:

Способ №1

Считаем отдельные дисперсии: $Var(u|x = 2)$ и $Var(u|x = 3)$. Совмещаем их, по аналогии с математическим ожиданием.

Способ №2

$$Var(u|x) = E(u^2|x) - [E(u|x)]^2$$

$$E(u^2|x) = 1$$

$$Var(u|x) = 1 - 0.25(x - 2)^2$$

Так как x принимает всего два значения ($x = 2, x = 3$) условную дисперсию можно представить в виде системы:

$$Var(u|x) = \begin{cases} 1 & \text{если x=2} \\ 0.75 & \text{если x=3} \end{cases}$$

или в качестве функции:

$$Var(u|x) = 0.25x + 1.5$$

Свойства условного математического ожидания:

1. $E(f(x)|x) = f(x)$
2. $Var(f(x)|x) = 0$
3. $E(f(x)y|x) = f(x)E(y|x)$
4. $E[E(y|x)] = E(y)$
5. $Var(y) = Var(E(y|x)) + E(Var(y|x))$ по теореме Пифагора

## 11.2 Упражнение 1

- Даны случайные величины $X = (x_1, x_2, ..., x_n)$ и $u = (u_1, u_2, ..., u_n)$, где $x_i\tilde{}N(10; 9)$, при этом векторы $(x_1, u_1); (x_2, u_2)..., (x_n, u_n)$ независимы, то есть может существовать зависимость между $x_i, u_i$, однако нет зависимости между парами. $*E(u_i|x_i) = 0$

a) $plim(\frac{1}{n}X'X)^{-1} =?$

b) $plim(\frac{1}{n}X'u) =?$

c) $plim(X'X)^{-1}X'u =?$

Решение:

a) $X'X = x_1^2 + x_2^2 + ... + x_n^2 = \sum_{i=1}^n x_i^2$

$$plim(\frac{1}{n}\sum_{i=1}^n x_i^2)^{-1} = [E(x_i^2)]^{-1} = (109)^{-1} \text{ (по закону больших чисел)}$$

Напоминание:

- $lim(f(a_n)) = f(lim(a_n))$
- ЗБЧ: $Y_1, Y_2, ..., Y_n$ $i.i.d.$, следовательно, $\bar{(}Y_n) -> E(Y_1)$
- $E(x_i^2) = Var(x_i) + (E(x_i))^2$

b) $\frac{1}{n}X'u = \frac{1}{n}\sum_{i=1}^n x_i u_i^2$

По ЗБЧ $plim(frac1n\sum_{i=1}^n x_i u_i^2) = E(x_i u_i)$.

Исходя из свойств условного математического ожидания $E[E(y|x)] = E(y)$ представим $E(x_i u_i)$ как $E(E(x_i u_i|x_i))$.

Так как $x_i$ известно, то по свойству $E(f(x)y|x) = f(x)E(y|x)$ можем вынести известную случайную величину за знак математического ожидания: $E(x_i E(u_i|x_i)) = 0$

Следовательно, $plim(\frac{1}{n}X'u) = 0$

c) $X'X$ и $X'u$ случайные величины домножим $X'X$ и $X'u$ на $\frac{1}{n}$

$$plim(\frac{1}{n}X'X)^{-1}\frac{1}{n}X'u = plim(\frac{1}{n}X'X)^{-1} * plim(\frac{1}{n}X'u) = (109)^{-1} * 0 = 0$$

Следовательно, $plim(X'X)^{-1}X'u = 0$

Напоминание:

- $lim(a_n b_n) = lim(a_n) lim(b_n)$

## 11.3 Стохастические регрессоры

Если:

1. $y = X\beta + u$
2. $\beta$ - константы
3. В матрице X: $x_{i.}$ - i-ая строка; векторы $(x_{i.}, y_i)$ независимы и одинаково распределены. Наблюдения - это случайная выборка
4. $E(u_i|x_i) = 0$
5. Гомоскедастичность $E(u_i^2|x_i) = Var(u_i|x_i) = \sigma^2$
6. $P($ñòîëáöû$X$ëèíåéíîíåçàâèñèìû$) = 1$
7. $\hat{\beta} = (X'X)^{-1}X'y$

## 11.4 Теорема Гаусса Маркова для стохастических регрессоров

1. Несмещенность: $E(\hat{\beta}|X) = \beta$
2. $Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$
3. $\hat{\beta}$ линейна по y
4. $\hat{\beta}$ эффективная и несмещенная оценка среди линейных по y
5. $plim(\hat{\beta}) = \beta$

## 11.5 Упражнение 2

Нарушена предпосылка теоремы Гаусса Маркова об отсутствие гетероскедастичности: $Var(u_i|x_i) = f(x_{i.})$

Найти:

1. Какой вид имеет матрица $\Omega = Var(u|X)$
2. $E(\hat{\beta}|X) =?$
3. $Var(\hat{\beta}|X) =?$
4. $plim(\hat{\beta}) =?$

Решение:

1. $\Omega = \Big( f(x_{1.}) \quad cov(u_1, u_2|X) \quad \dots \quad cov(u_1, u_n|X) \vdots \;\vdots\; \vdots\; \vdots cov(u_1, u_n|X) \quad \dots \quad \dots \quad f(x_{n.}) \Big)$

   Так как выборка случайна, то $cov(u_i, u_j|X) = 0$, при $i \neq j$

   $$\Omega = \Big( f(x_{1.}) \quad 0 \quad \dots \quad 0 \vdots \;\vdots\; \vdots\; \vdots 0 \quad \dots \quad \dots \quad f(x_{n.}) \Big)$$

2. $E(\hat{\beta}|X) = \beta$
3. $Var(\hat{\beta}|X) = Var((X'X)^{-1}X'y|X) = (X'X)^{-1}X' \times Var(y|X) \times ((X'X)^{-1}X')' = (X'X)^{-1}X' \times Var(y|X) \times X(X'X)^{-1} = (X'X)^{-1}X' \times Var(X\beta + u|X) \times X(X'X)^{-1} = (X'X)^{-1}X' \times Var(u|x) \times X(X'X)^{-1} = (X'X)^{-1}X' \times \Omega \times X(X'X)^{-1}$

В действительности возникает проблема при оценке матрицы $\Omega$, так как она имеет размер $n \times n$, то есть нам нужно оценить n чисел на диагонали по n наблюдениям, что является достаточно нетривиальной задачей. Оказывается, что нельзя получить состоятельную оценку матрицы $\Omega$, но можно получить такую оценку этой матрицы, чтобы левая часть, то есть $Var(\hat{\beta}|X)$, была состоятельна.

Например, состоятельной оценкой бцдет следующая:
$\hat{Var}_{HC}(\hat{\beta}|X) = (X'X)^{-1}X' \times \hat{\Omega} \times X(X'X)^{-1}$ где $\hat{\Omega}$ можно представить в виде:(White)

$$\hat{\Omega}_{HC0} = \Big( \hat{u_1^2} \quad 0 \quad \dots \quad 0 \vdots \;\vdots\; \vdots\; \vdots 0 \quad \dots \quad \dots \quad \hat{u_n^2} \Big)$$

- абревиатура HC - heteroscedasticity-consistent

При гетероскедастичности оценки несмещенные и состоятельные, но проблема будет заключаться в t-статистике, так как мы не знаем ее распределение, следовательно, нам необходимо использовать стандартные отклонения из устойчивой к гетерокедастичности матрицы, чтобы получить надежные оценки:

$t_{HC} = \frac{\hat{\beta}_j - \beta_j}{se_{HC}(\hat{\beta}_j)} \sim N(0;1)$ где $se_{HC}(\hat{\beta}_j)$ находятся из матрицы $\hat{Var}_{HC}(\hat{\beta}|X)$

и теперь $t_{HC} \sim N(0;1)$ (асимтотически нормальное)

4. $plim(\hat{\beta}) = plim(X'X)^{-1}X'y = \beta + plim(X'X)^{-1}X'u = \beta$

Глава 12

# Конспект семинара 14. Гетеро или гомо? {hetero_or_homo}

Для начала стоит определить, что такое Гетероскедастичность!!!

Определение, украденное с Википедии: Гетероскедастичность (англ. heteroscedasticity) — понятие, используемое в прикладной статистике (чаще всего — в эконометрике), означающее неоднородность наблюдений, выражающуюся в неодинаковой (непостоянной) дисперсии случайной ошибки регрессионной (эконометрической) модели. Гетероскедастичность противоположна гомоскедастичности, означающей однородность наблюдений, то есть постоянство дисперсии случайных ошибок модели.

Другими словами:

$$Var(u_i|X) \neq const = f(x_i)$$

### 12.0.1 Какие последствия отсюда вытекают?

Хорошие последствия:

- $\widehat{\beta}$ - несмещенно
- $\widehat{\beta}$ - состоятельно

Плохие последствия:

- $\widehat{\beta}$ - неэффективно

$$\left(\frac{\widehat{\beta}_i - \beta_i}{se(\widehat{\beta}_i)}\right) \nrightarrow N(0,1)$$

(!) Для борьбы с гетероскедастичностью нужно знать структуру самой гетероскедастичности!

Например, на прошлом семинаре была доказано, что:

$$\left(\frac{\widehat{\beta}_i - \beta_i}{se_{hco}(\widehat{\beta}_i)}\right) \to N(0,1)$$

В данном случае индекс hco означает heteroscedasticity consistent, т.е. гетероскедастично устойчиво. Говоря языком программистов и статистов, берутся робастные стандартные ошибки.

$se_{hco}(\widehat{\beta}_i)$ добывается с диоганали матрицы $\widehat{var}(\widehat{\beta}|X) = (X'X)^{-1}\widehat{\Omega}X(X'X)$

$$\widehat{\Omega}_{hco} = \begin{pmatrix} \widehat{u}_1^2 & \dots & \dots & 0 \\ 0 & \widehat{u}_2^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \widehat{u}_n^2 \end{pmatrix}$$

### 12.0.2 Обнаружение гетероскедастичности

1. Графическое

Для наглядного примера возьмём датасет про зависимость стоимости квартир от площади.

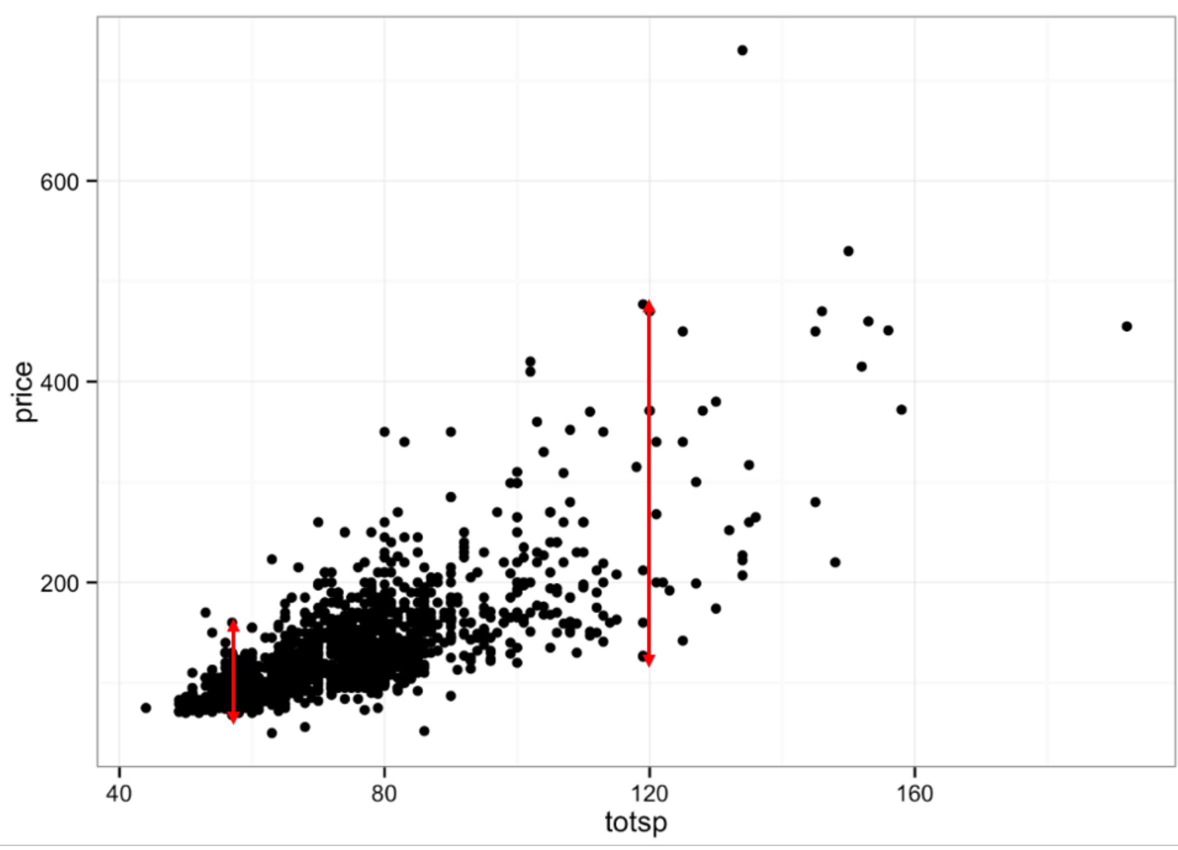



Рис. 12.1: График с явной гетероскедастичностью

На данном графике красными линиями обозначены примерные отклонения. Отчетливо видно, что в зависимости от значения площади наблюдается разный разброс в ценах квартир.

Как правило, причиной возникновения гетероскедастичности является размер объекта.

Например:

- размер семьи влияет на диспресию расходов и доходов
- размер компании влияет на разброс в доходах сотрудников
- можно придумать ещё много примеров

2. Более утончённое графическое

Если $Var(u_i|X) = \sigma^2$, то $var(\widehat{u}|X) = (I - H)\sigma^2$. Можно заметить, что если u - гомоскедастично, тогда дисперсии всё равно будут разными ($H = X(X'X)^{-1}X'$). То есть $Var(\widehat{u}_i|X) = g_i(X)$

Что же делать?

Можно отнормировать остатки!

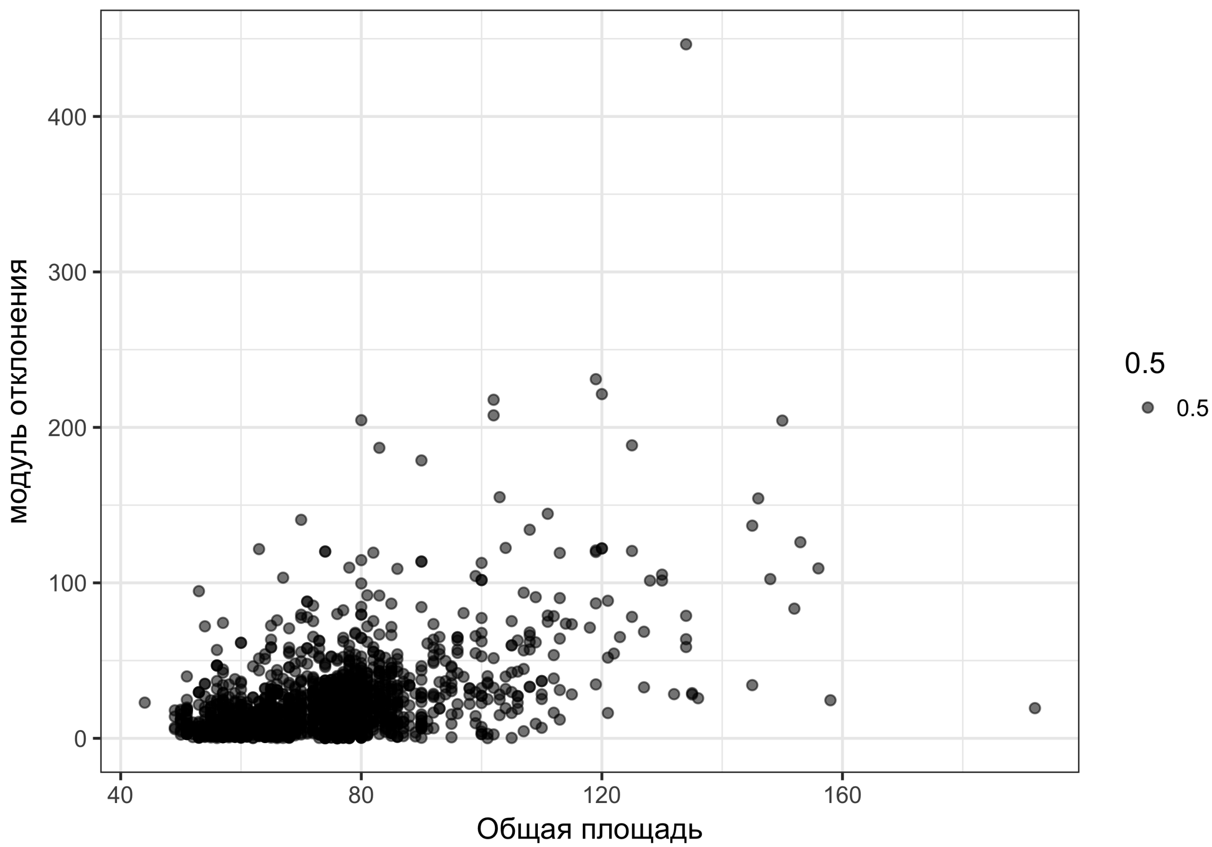
$$\widehat{u}_i^* = \frac{\widehat{u}_i^*}{\sqrt{1 - h_{ii}}}$$

В данном случае $h_{ii}$ это диагональные элементы матрицы H

$$\Rightarrow Var(\widehat{u}_i^*) = \sigma^2$$

Следовательно, в случае в случае гомоскедастичности дисперсия будет постоянной величиной.

В случае же гетероскедастичности график будет выглядить следующим образом (возьмём тот же пример с квартирами):




В даном случае строится график зависимости $\widehat{u}_i^{*2}$ или $|\widehat{u}_i^*|$ от переменной, преположительно виновной в гетероскедастичности.

3. Тесты

Как делаются большинство тестов:

Шаг №1

Оценим $y_i = \beta_i + \beta_2 x_i + u_i$

Отсюда получим $\widehat{u}_i$

Шаг №1.5

$$\widehat{u}_i^* = \frac{\widehat{u}_i}{\sqrt{1 - h_{ii}}}$$

Шаг 2: вспомогательная регрессия

Строим регрессию «размера остатка», это может быть $\widehat{u}_i^2$, $|\widehat{u}_i|$ или $\log \widehat{u}_i$ на объясняющие переменные:

$$\text{размер остатка}_i = \gamma_1 + \gamma_2 z_i + \gamma_3 w_i + \ldots + \gamma_h r_i + \nu_i$$

- В случае $\hat{u}_i^2$ используется тест Бройша — Пагана (про него можно подробно почитать на википедии)

Далее проверем гипотезу о гетероскедастичности:

$$H_0 : \gamma_1 = \gamma_2 = ... = \gamma_h = 0, \text{гомоскедастичность}$$

$H_a$: хотя бы один из коэффициентов $\gamma_1, \gamma_2, \ldots, \gamma_h$ отличен от нуля.

Далее для проверки можно взять одну из тестовых статистик:

F-test

$$F_{test} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/n - k_{UR}}$$

где

- $RSS_R$ - это сумма квадратов остатков модели с ограничениями
- $RSS_{UR}$ - это сумма квадратов остатков модели без ограничений
- $q$ - это количество ограничений
- $n$ - это объём выборки
- $k_{UR}$ - это количество параметров модели без ограничений

Тест множителей Лагранжа

$$LM = nR^2_{step2} \sim \chi^2_{n-1}$$

где

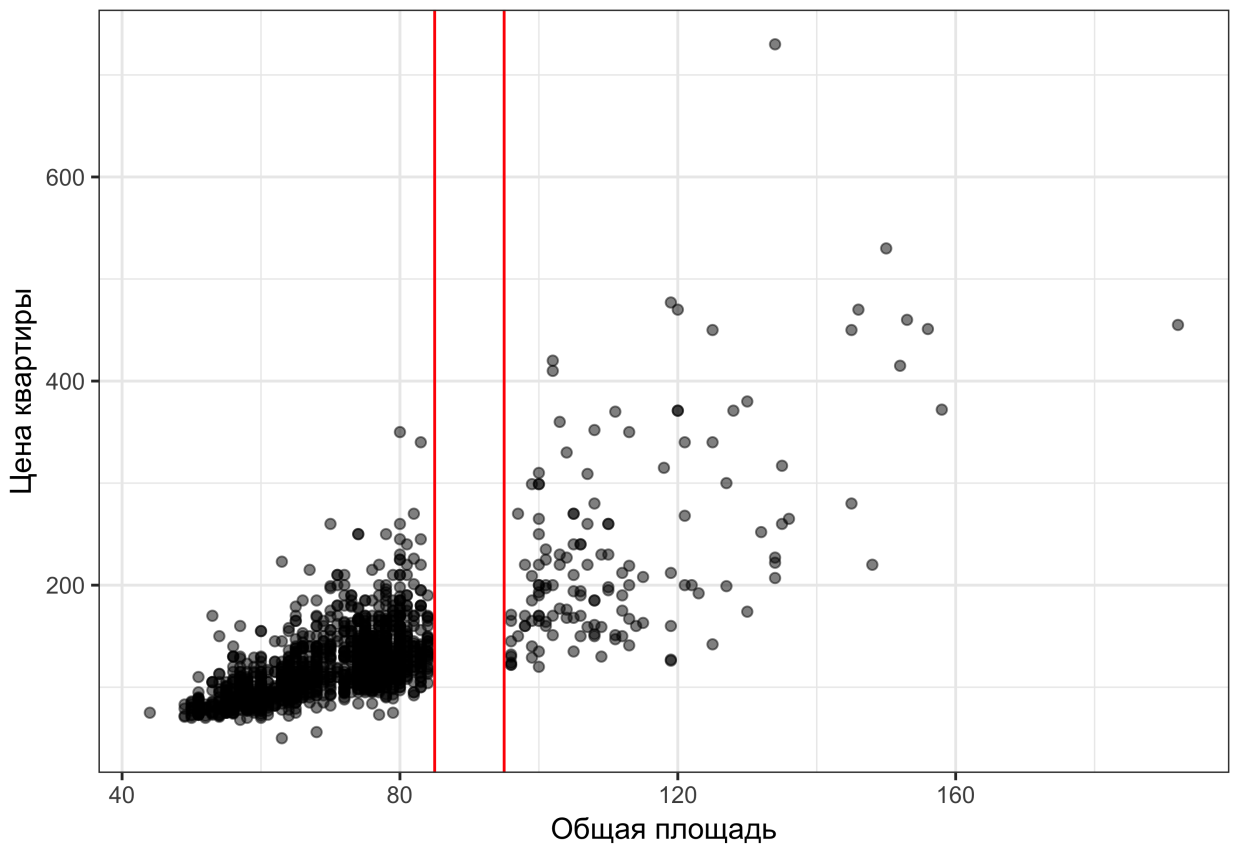
- $R$ - коэффициент детерминации
- $n$ - количество ограниченй

Тест Голдфельда-Квандта (Goldfeld-Quandt)

$$\begin{cases} H_0 : Var(u_i|x_i) = \sigma^2 \\ H_a : Var(u_i|x_i) = f(x_i) \end{cases}$$

Далее наблюдения сортируются по подозреваемому росту (по дисперсии). Из сортированной выборки выкидываются наблюдения с средней дисперсией (как правило, избавляются примерно от 20%). Далее берётся F-статистика.

Выходит примерно следующая картина:




$$F = \frac{RSS2/(n_2 - k)}{RSS_1/(n_1 - k)} \sim F(n_2 - k, n_1 - k)$$

где

- $RSS_1$ - сумма квадратов остатков части выборки с низкой дисперсией
- $RSS_2$ - сумма квадратов остатков части выборки с высокой дисперсией
- $n_1$ - количество наблюдений с низкой дисперсией
- $n_2$ - количество наблюдений с высокой дисперсией
- $k$ - количество ограничений

Примечание: если $n_1 = n_2$, тогда статистика равна:

$$F = \frac{RSS_2}{RSS_1}$$

Если F большое $\Rightarrow H_0$ отвергается (т.е. наблюдается гетероскедастичность).

### 12.0.3 Так делать нехорошо:

1. Протестировать наличие гетероскедастичности

2. Если гипотеза о гомоскедаксичности отвергнута, использовать робастные стандартные ошибки($se_{hco}(\widehat{\beta}_i)$)

3. Если гипотеза о гомоскедастичности отвергнута, использовать обычные стандартные ошибки ($se(\widehat{\beta}_i)$)

### 12.0.4 Примечание

Как делают нормальные люди:

RSS - Residiual Sum of sq $= \widehat{u}_i^2$

ESS - Explained Sum of sq $= \sum(\widehat{y}_i^2 - \bar{y}_i)^2$

Как делают гопники:

ESS - Error Sum of sq $= \widehat{u}_i^2$

RSS - Regression Sum of sq $= \sum(\widehat{y}_i^2 - \bar{y}_i)^2$

### 12.0.5 Пакеты для R:

- tidyvers
- lmtest
- memisc
- sandwich

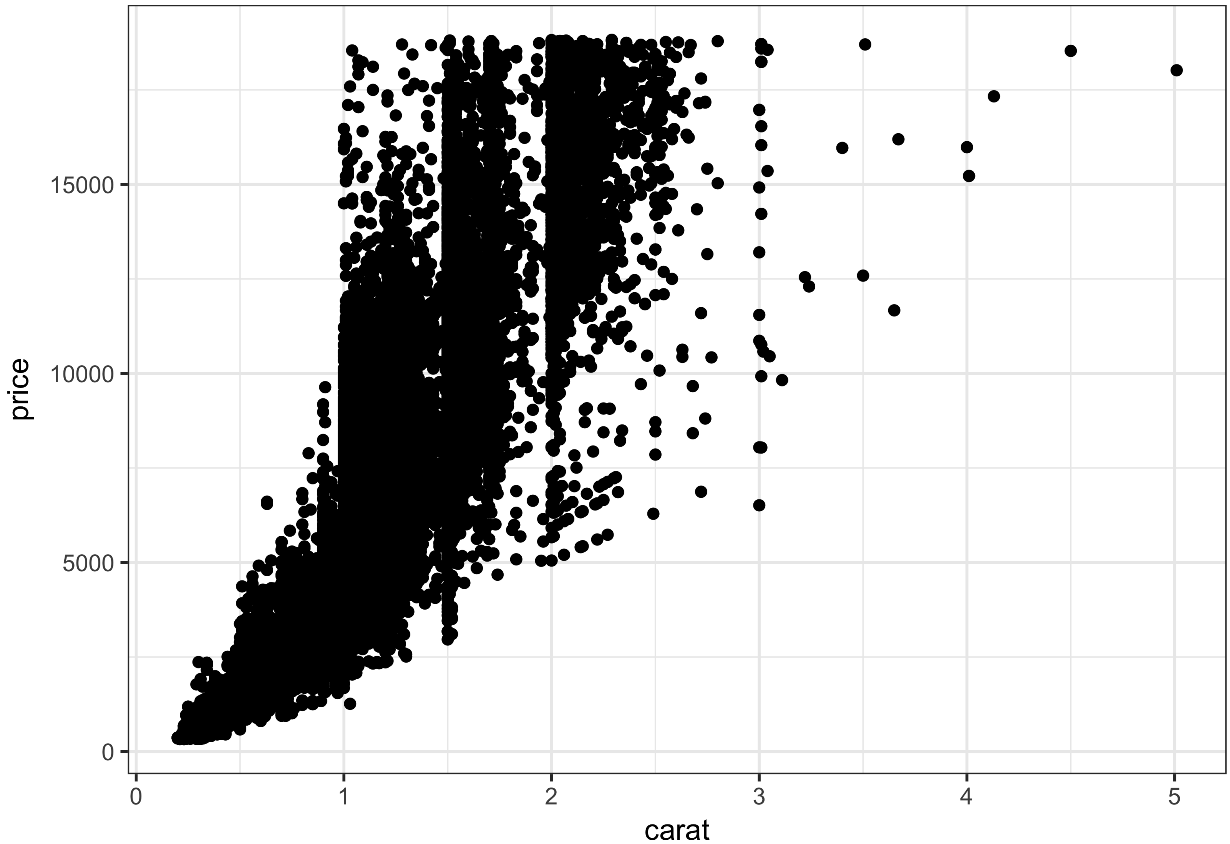
### 12.0.6 Немного R

```
glimpse(diamonds)
```

```
## Observations: 53,940
## Variables: 10
## $ carat   <dbl> 0.23, 0.21, 0.23, 0.29, 0.31, 0.24, 0.24, 0.26, 0.22, ...
## $ cut     <ord> Ideal, Premium, Good, Premium, Good, Very Good, Very G...
## $ color   <ord> E, E, E, I, J, J, I, H, E, H, J, J, F, J, E, E, I, J, ...
## $ clarity <ord> SI2, SI1, VS1, VS2, SI2, VVS2, VVS1, SI1, VS2, VS1, SI...
## $ depth   <dbl> 61.5, 59.8, 56.9, 62.4, 63.3, 62.8, 62.3, 61.9, 65.1, ...
## $ table   <dbl> 55, 61, 65, 58, 58, 57, 57, 55, 61, 61, 55, 56, 61, 54...
## $ price   <int> 326, 326, 327, 334, 335, 336, 336, 337, 337, 338, 339,...
## $ x       <dbl> 3.95, 3.89, 4.05, 4.20, 4.34, 3.94, 3.95, 4.07, 3.87, ...
## $ y       <dbl> 3.98, 3.84, 4.07, 4.23, 4.35, 3.96, 3.98, 4.11, 3.78, ...
## $ z       <dbl> 2.43, 2.31, 2.31, 2.63, 2.75, 2.48, 2.47, 2.53, 2.49, ...
```

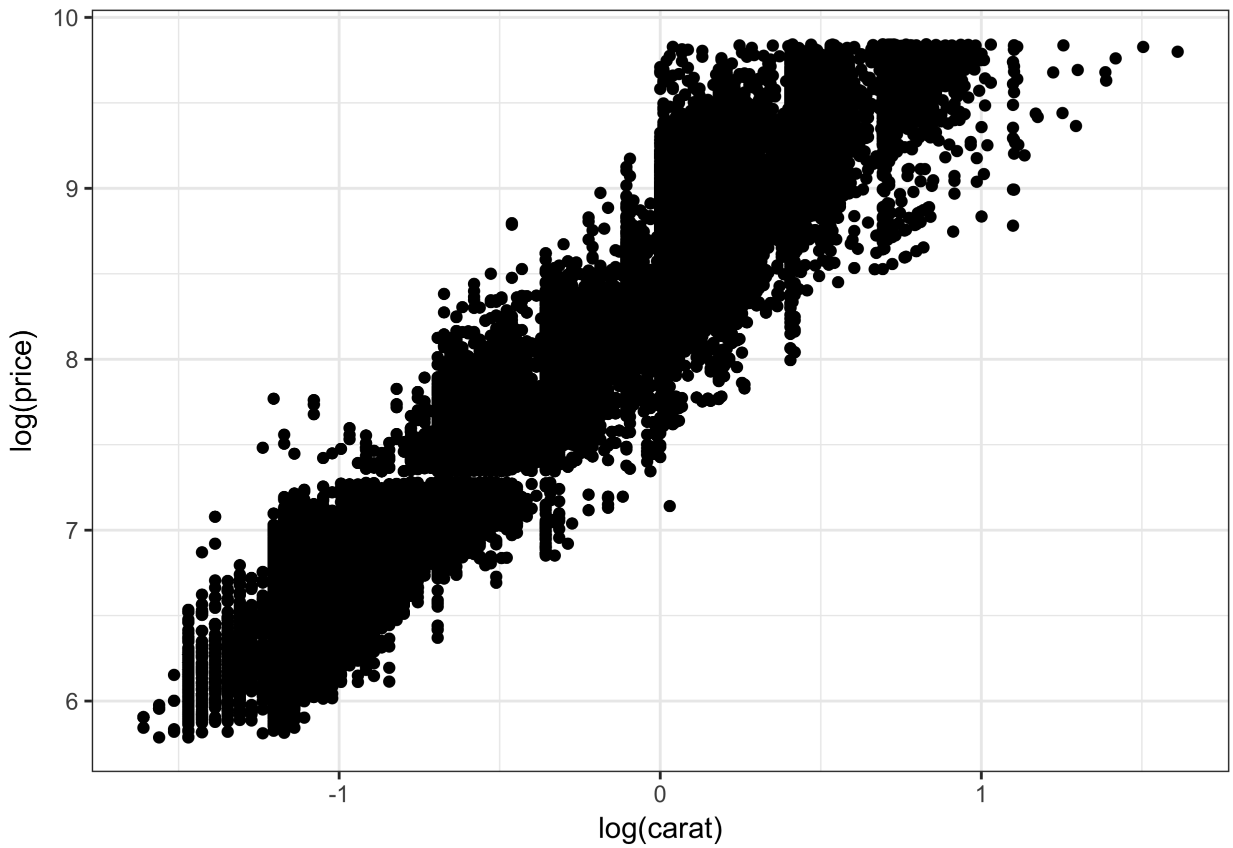
Посмотрим на график зависимости цены и карата бриллианта

```
qplot(data = diamonds, x = carat, y = price)
```




Видна гетероскедастичность Поробуем прологарифмировать.

```
qplot(data = diamonds, x = log(carat), y = log(price))
```




Гетероскедастичность осталась.

Построим модель регрессии, где карат - объясняющая переменнаяб цена - зависимая переменная.

```
model <- lm(data = diamonds,
            log(price) ~ log(carat))
coeftest(model)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 8.4486607  0.0013647 6190.90 < 2.2e-16 ***
## log(carat)  1.6758167  0.0019338  866.59 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(model, vcov. = vcovHC)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 8.4486607  0.0014750  5728.1 < 2.2e-16 ***
```

```
## log(carat)  1.6758167  0.0020278   826.4 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

тест Breush-Pagan (White)

```
bptest(model, varformula = ~ log(table) + log(carat), data = diamonds)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 449.18, df = 2, p-value < 2.2e-16
```

тест Goldfeld-Quandt

Чтобы отловить какая величина виновата в гетероскедастичности

```
gqtest(model, fraction = 0.2, order.by = diamonds$table)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model
## GQ = 1.004, df1 = 21574, df2 = 21574, p-value = 0.3834
```

```
gqtest(model, fraction = 0.2, order.by = diamonds$carat)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model
## GQ = 1.3634, df1 = 21574, df2 = 21574, p-value < 2.2e-16
```

### 12.0.7 Полезные ссылочки

Ссылки на одну свободную энциклопедию:

- Собственно гетероскедастичность Википедия
- F-тест Википедия
- Тест множителей Лагранжа Википедия
- Тест Бройша — Пагана Википедия
- Тест Голдфелда — Куандта Википедия

Прочие ссылки:

- Компьютерная гетероскедастичность от занимательного исследователя Гетероскедастичность в R
- Еще один рример определения гетероскедастичности в R R-bloggers

# Глава 13

# Байесовский подход

конспект: Мария Салунина, Полина Лапшова

дата: 23 января 2017

## 13.1 Хинты к брутальной части (16 января 2017)

### 13.1.1 Первая задача

Дано:

- $\hat{y} = 3 + 7x$
- $R^2 = 0.6$
- $sVar(y) = 1000$

Задание 1: найти $sCov(y, x)$.

Мы знаем, что $sVar(y) = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}$.

Замечание: чем отличается $Var(y)$ от $sVar(y)$? $Var(y)$ - матрица из констант размера $n \times n$. $sVar(y)$ - вектор размера $1 \times 1$, случайная величина.

Что такое $R^2$?

1) Первая интерпретация: доля объяснённой дисперсии.

   $R^2 = \frac{sVar(\hat{y})}{sVar(y)} = 0.6$, следовательно, $sVar(\hat{y}) = 600$.

   $\hat{y} = 3 + 7x$, т. е. $\hat{y}$ в семь раз больше $x$, значит:

   $sVar(x) = \frac{600}{49}$

2) Вторая интерпретация: $R^2 = sCorr^2(y, \hat{y})$

   $sCorr(y, \hat{y}) = \sqrt{0.6}$

   $sCorr(y, x) = +\sqrt{0.6}$

Замечание: знак $"+"$, так как в регрессии перед $x$ стоит плюс.

$sCorr(y,x) = \frac{sCov(y,x)}{\sqrt{sVar(y)sVar(x)}} = +\sqrt{0.6}$

$sVar(y) = 1000$

$sVar(x) = \frac{600}{49}$

Следовательно, $sCov(y,x) = 85.7143$

Итог: индивидуальные регрессии позволяют найти любую выборочную корреляцию, любую выборочную ковариацию, любую выборочную дисперсию.

Задание 2: найти регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$

Как найти вектор $\hat{\beta}$?

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1}X'y$$

Фриша-Вау: проведём регрессию в два шага.

1) Отрегрессируем $x$, $z$, $y$ на 1 (чтобы избавиться от единичного столбца)

   Для $x$:

   $x_i = c + u_i$

   тогда: $\hat{c} = \overline{x}$,

   $\widetilde{x}_i = x_i - \overline{x}$ - остаток от регрессии $x$.

   ...

   Получим остатки от регрессий $\widetilde{x}, \widetilde{z}, \widetilde{y}$.

2) Чтобы оценить $\hat{\beta}_2$ и $\hat{\beta}_3$, необходимо построить следующую регрессию.

   $\widetilde{y}$ на $\widetilde{x}$, $\widetilde{z}$

   $\hat{\widetilde{y}}_i = \hat{\beta}_2 \widetilde{x}_i + \hat{\beta}_3 \widetilde{z}_i$

   Следовательно:

$$\begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y}$$

$$\widetilde{y} = \begin{pmatrix} y_1 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{pmatrix}$$

$$\widetilde{X} = \begin{pmatrix} x_1 - \overline{x} & z_1 - \overline{z} \\ \vdots & \vdots \\ x_n - \overline{x} & z_n - \overline{z} \end{pmatrix}$$

$$\widetilde{X}'\widetilde{X} = \begin{pmatrix} \sum_{i=1}^{n}(x_i-\overline{x})^2 & \sum_{i=1}^{n}(x_i-\overline{x})(z_i-\overline{z}) \\ \sum_{i=1}^{n}(x_i-\overline{x})(z_i-\overline{z}) & \sum_{i=1}^{n}(z_i-\overline{z})^2 \end{pmatrix}$$

$$\widetilde{X}'\widetilde{y} = \begin{pmatrix} \sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \\ \sum_{i=1}^{n}(z_i-\overline{z})(y_i-\overline{y}) \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(\frac{\widetilde{X}'\widetilde{X}}{n-1}\right)^{-1}\frac{\widetilde{X}'\widetilde{y}}{n-1}$$

Средние были даны, любая регрессия проходит через среднюю линию, можно восстановить $\hat{\beta}_1$.

### 13.1.2 Третья задача

Дано две модели:

$$\hat{y}_i = \hat{\beta}_1 + \underset{(3.2)}{\hat{\beta}_2}\, x_i$$

$$\hat{y}_i = \hat{\gamma}_1 + \hat{\gamma}_2 x_i + \underset{(7.8)}{10}\, z_i$$

$R^2 = 0.8$ (для второй модели)

Задание: проверить гипотезу $H_0 : \beta_2 = 0$.

Проверим сначала $\gamma_3$:

$$t = \frac{10}{7.8}$$

Если знаем $R^2$, то можем проверить значимость регрессии в целом.

$$t^2 = F = \left(\frac{10}{7.8}\right)^2$$

Одинаковы ли регрессии.

$$F = \frac{(R^2_{UR} - R^2_R)/1}{R^2_{UR}/(n-k_{UR})}$$

Пояснение: единичка в знаменателе числителя — одно ограничение на вторую регрессию ($\gamma_3 = 0$).

Следовательно, получаем $R^2_R$, можем проверить гипотезу о значимости.

Почему $t^2 = F$?

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

$$t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{sCov(x,y)}{sVar(x)} \frac{1}{\sqrt{\frac{RSS}{n-2}\frac{1}{\sum(x_i-\overline{x})^2}}}$$

$$RSS^2 = (1-R^2) * TSS = (1 - sCorr^2(x,y))sVar(y)(n-1)$$

$$t^2 = \frac{sCov^2(x,y)}{sVar(x)} \frac{(n-2)(n-1)}{(1-sCorr^2(x,y))sVar(y)(n-1)} = \frac{sCorr^2(x,y)/1}{(1-sCorr^2(x,y))/(n-2)}$$

$H_0 : \beta_2 = 0$ — одно ограничение. Проверяем одну и ту же гипотезу только с разных сторон.

$F = \frac{R^2}{(1-R^2)/(n-2)}$ — вуаля! то же самое! Ч.Т.Д.

## 13.2 Отличие классического подхода от байесовского

### 13.2.1 Классический подход

$\beta$ vs $\hat{\beta}$

1) $\beta$ — неизвестный параметр $const$ (не знаю и никогда не узнаю). $P(\beta > 0)$ — бессмысленный вопрос (данная вероятность равна либо 0, либо 1).

2) Модель: как данные $y_i$ зависят от $\beta$.

3) Применяется метод (метод наименьших квадратов, метод моментов, метод максимального правдоподобия и т.д.); на основании модели получаем $\hat{\beta} = f(y_1, y_2, \ldots, y_n)$.

4) ЦПТ, ЗБЧ, $\Delta$-метод при $n \to \infty$

   $\hat{\beta} \sim \mathbb{N}(.,.)$.

Итог: пункт 3 — точечная оценка $\hat{\beta}$; пункт 4 — доверительный интервал $\beta \in [\hat{\beta} - \ldots; \hat{\beta} + \ldots]$.

### 13.2.2 Байесовский подход

Есть только $\beta$ — всё наше знание о неизвестном параметре сформулировано в виде закона распределения.

1) Априорное мнение о $\beta$ до получения наблюдения. Существует $P(\beta > 0)$.

2) Модель: как данные $y_i$ зависят от $\beta$.

3) Используется формула условной вероятности. По формуле $f(\beta \mid y_1, y_2, \ldots, y_n)$ получается апостериорное распределение $\beta$.

   $P(\beta > 0 \mid y_1, y_2, \ldots, y_n)$ — можно найти.

4) Можно получить точечную и интервальную оценку.

### 13.2.3 Упражнение №1 (про шапку с монетками)

Условие: в шапке есть серебряные и золотые монетки; $\beta$ — доля золотых монеток.

1) $0 \leqslant \beta \leqslant 1$

   Предположим, $\beta$ распределена равномерно: $\beta \sim \mathbb{U}[0, 1]$

2) $y_i$ — наблюдения.

$$y_i = \begin{cases} 0, & \text{если монетка серебряная;} \\ 1, & \text{если монетка золотая.} \end{cases}$$

Достаём, смотрим, какая монетка, отмечаем наблюдение, кидаем монетку обратно в шапку, перемешиваем.

$y_i \mid \beta$ — независимы

|$y_i \mid \beta$|0|1|
|-----------|---------|---------|
|*вероятность*|$1-\beta$| $\beta$ |

`![*Функция плотности "до"*](images/16_1.png)`

$y_1=1$

$y_2=1$

$f(\beta\mid y_1=1, y_2=1)=\frac{f(\beta, y_1=1, y_2=1)}{P(y_1=1,y_2=1)} \propto f(\beta, y_1=1, y_2=1)

$$P(y_1 = 1, y_2 = 1 \mid\mid \beta) f(\beta) = \begin{cases} \beta^2, & \beta \in [0;1], \\ 0, & \text{иначе.} \end{cases}$$

$$\int\limits_0^1 \beta^2\, d\beta = \left.\frac{\beta^3}{3}\right|_0^1 = \frac{1}{3}$$

Следовательно, необходимо домножить на 3.

\[
f(\beta\mid y_1=1, y_2=1) =
\begin{cases}
3\beta^2, & \beta \in [0;1], \\
0, & \text{иначе.}
\end{cases}
\]

`![*Функция плотности "после"*](images/16_2.png)`

До наблюдений: $P(\beta>\frac{1}{2})=\frac{1}{2}$.

После наблюдений: $P(\beta>\frac{1}{2}\mid y_1=1, y_2=1)=\int\limits_\frac{1}{2}^1 3\beta^2\,d\beta = \f

Точечная оценка: апостериорное мат. ожидание или медиана.

$\mathbb{E}(\beta\mid y_1=1, y_2=1) = \int\limits_0^1 \beta \cdot f(\beta\mid y_1=1, y_2=1)\,d\beta$

### 13.2.4 Упражнение №2 (про Машу и мышек)

Задача: где-то на числовой прямой, в точке $m$, Маша прячет от мышек сыр. При этом известно, что её любимое число — 13.

1) $m \sim \mathbb{N}(13, 100)$
2) Прибежали две мышки и пропищали, где, по их мнению, наиболее сильно пахнет сыром: $y_1 = 9$, $y_2 = 18$.
   - $y_i \mid m$ — независимы;
   - $y_i = m + u_i$;
   - $y_i \mid m \sim \mathbb{N}(m, 25)$.
3) $f(m) \propto e^{-(m-13)^2/(2*100)}$ — априорная функция плотности.

$$f(m \mid y_1 = 9, y_2 = 18) \propto f(m, y_1 = 9, y_2 = 18) = f(y_1 = 9, y_2 = 18 \mid m)f(m) \propto e^{-(9-m)^2/(2*25)}e^{-(18-m)^2/(2*25)}e^{-(m-13)^2/(2*100)} \propto e^{-\frac{1}{2}*\frac{1}{100}*(9m^2-2m(13+9*4+18*4))} \propto e^{-\frac{1}{2}*\frac{1}{100/9}*(m-\frac{121}{9})^2}$$ — функция плотности.

$$m \sim \mathbb{N}(\tfrac{121}{9}, \tfrac{100}{9})$$

Магия: $\frac{121}{9} = \frac{4}{9} * 9 + \frac{4}{9} * 18 + \frac{4}{9} * 13$

#### 13.2.4.1 Домашнее задание

Установить STAN bayesian.

# Глава 14

# Мультиномиальное распределение

Конспект: Молдованов Александр, Умрихин Александр

Дата: 08.02.2017

Можно почитать о мультиномиальном распределении на википедии

## 14.1 Задание 1

Дано: Вы решили сходить в лес 100 раз. Вероятность встретить там волка составляет 0,3, Бабу-Ягу - 0,1, а Хариса - 0,6.

Пусть

$$\begin{pmatrix} \tilde{o}_1 \\ \tilde{o}_2 \\ \tilde{o}_3 \end{pmatrix},$$

где $\tilde{o}_1$ - количество раз, когда вы встретили волка, $\tilde{o}_2$ - Бабу-Ягу, $\tilde{o}_3$ - Хариса.

Этот вектор имеет мультиномиальное распределение с параметрами (n=100, p(0,3; 0,1; 0,6)), т. е.

$$\begin{pmatrix} \tilde{o}_1 \\ \tilde{o}_2 \\ \tilde{o}_3 \end{pmatrix} \sim multi(n = 100, p(0,3; 0,1; 0,6))$$

Вопрос: Какова вероятность встретить волка 10 раз, Бабу-Ягу - 20, а Хариса - 70? p($\tilde{o}_1$=10, $\tilde{o}_2$=20, $\tilde{o}_3$=70)-?

Решение:

$$\left.\begin{array}{l} 0,3^{10} - \text{вероятность встретить 10 волков} \\ 0,1^{20} - \text{вероятность встретить 20 раз Бабу-Ягу} \\ 0,6^{70} - \text{вероятность встретить 70 раз заблудившегося Хариса} \end{array}\right\} =>$$

$$\text{вероятность встретить их всех за 100 раз} = 0,3^{10} \times 0,1^{20} \times 0,6^{70}$$

Важен коэффициент перед этим всем: !!!!!! обосновать !!!!!

$C_{100}^{10} \times C_{90}^{20} = \frac{100!}{10! \times 90!} \times \frac{90!}{20! \times 70!} = \frac{100!}{10! \times 20! \times 70!} \backslash$

Ответ: $\frac{100!}{10! \times 20! \times 70!} \times 0,3^{10} \times 0,1^{20} \times 0,6^{70}$

## 14.2 Задание 2

Дано:

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix},$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim mult(n, (p_1, p_2, p_3))$$

Вопрос: $E(X), Var(X)$ - ?

Решение:

$$E(X) = \begin{pmatrix} np_1 \\ np_2 \\ np_3 \end{pmatrix}$$

$$\begin{aligned} Cov(x_1; x_2) = Cov(I_{1,1} + I_{1,2} + ... + I_{1,100}; I_{2,1} + I_{2,2} + ... + I_{2,100}) = \\ = \sum Cov(I_{1,i}; I_{2,i}) = nCov(I_{1,1}; I_{2,1}) = n(E(I_{1,1}I_{2,1}) - E(I_{1,1})E(I_{2,1})) = -np_1p_2 \Rightarrow \\ Var(X) = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & -np_1p_3 \\ -np_1p_2 & np_2(1-p_2) & -np_2p_3 \\ -np_1p_3 & -np_2p_3 & np_2(1-p_2) \end{pmatrix} = n\left( \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{pmatrix} - pp^T \right) \end{aligned} \quad (14.1)$$

## 14.3 Определение симплекса (n-мерного)

Опять же можно почитать на википедии

$$\Delta^n = \{x | x_1, ..., x_n \geq 0, x_1 + ... + x_n = 1\}$$

## 14.4 Переходим в STAN

Для Байевского подхода нужна модель, которая описывает данные:

$$\begin{pmatrix} \tilde{o}_1 \\ \tilde{o}_2 \\ \tilde{o}_3 \end{pmatrix} \sim multi(n = 100, p(0,3; 0,1; 0,6))$$

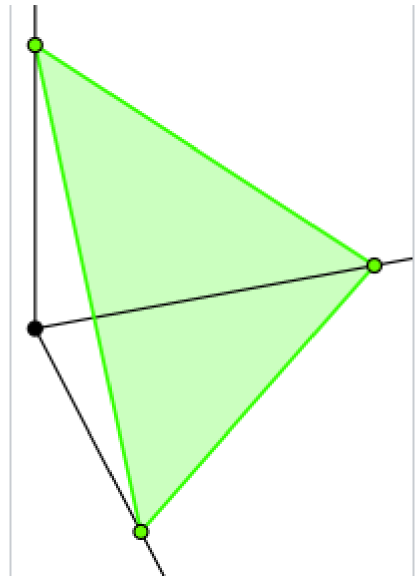


Рис. 14.1: Рисунок для n = 3

Давайте уточним вводные данные:

Мы были в гостях у бабушки. Она нам сказала, что волки водятся в лесу также часто, как и Баб-Яжки.

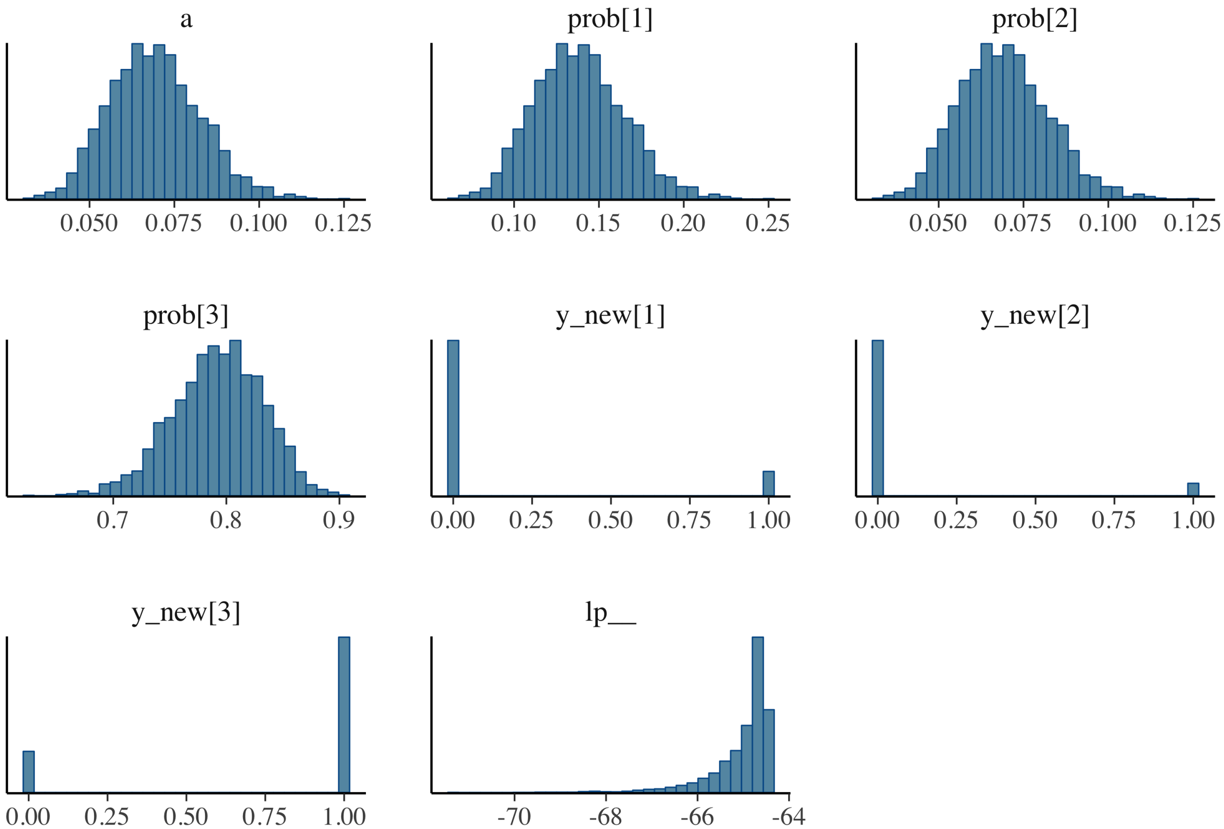
Следовательно у вероятность встретить волка равна $\alpha$, Бабу-Ягу тоже $\alpha$, а Хариса: $1 - 2\alpha$

Зададим априорное распределение на $\alpha$:

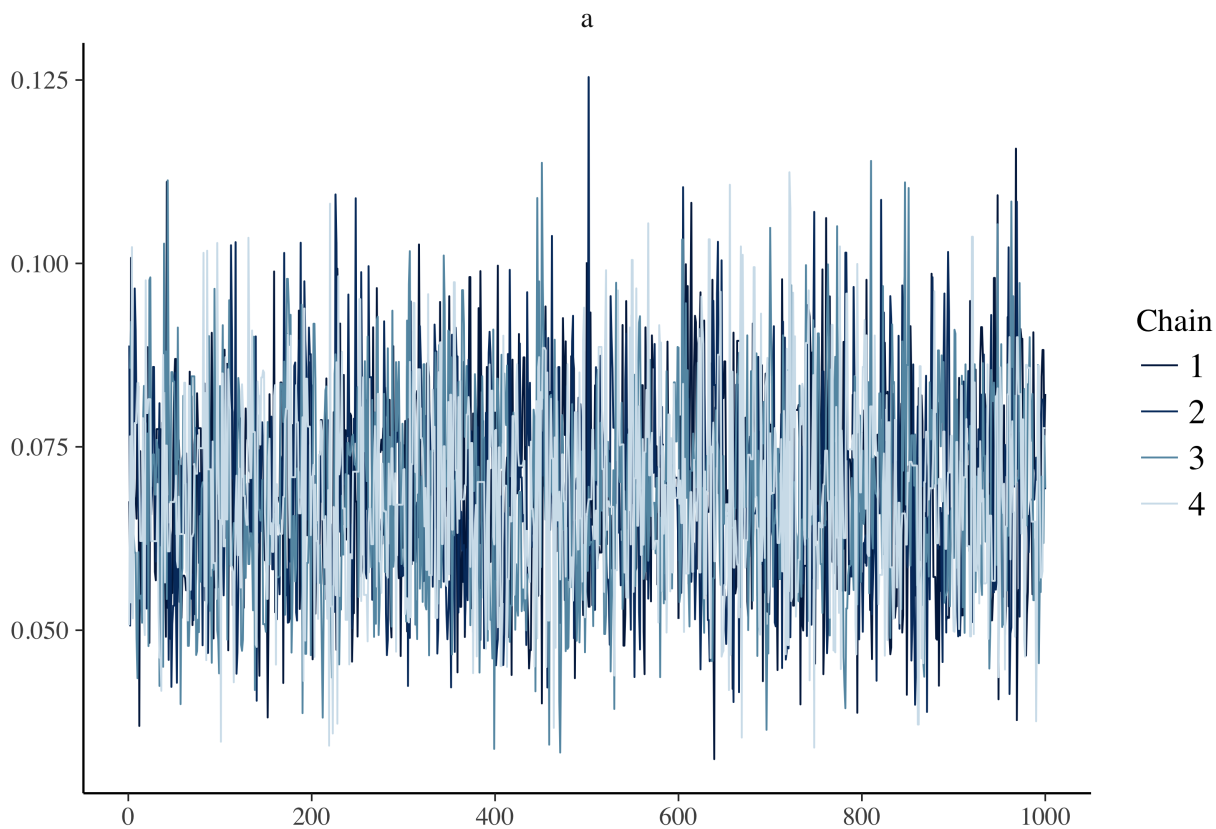
$\alpha \sim U[0, 0.5]$, где $U$ - равномерное распределение

Следовательно, наша задача сводится к тому, чтобы получить апостериорное распределение $\alpha$: f($\alpha$(x))

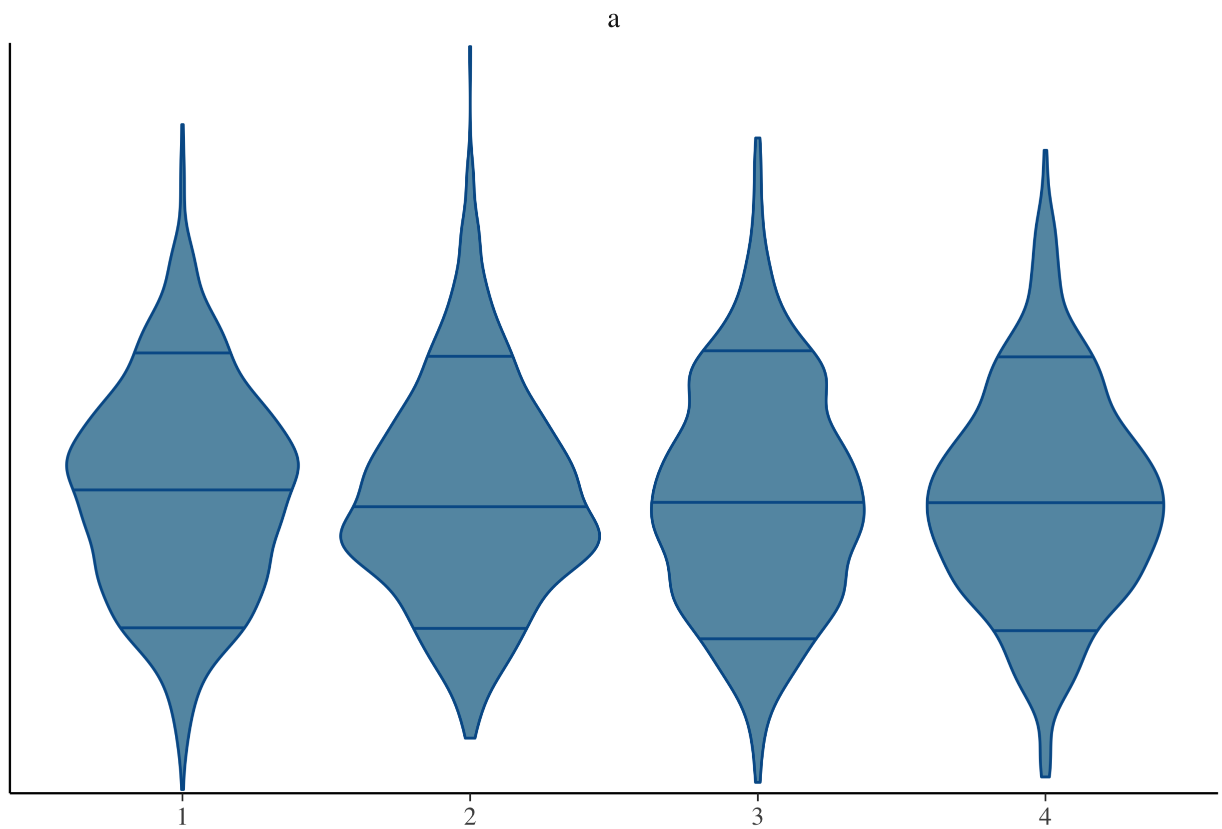
```
library(rstan)
library(bayesplot)

data <- list(y = c(15, 5, 80))

model <- stan_model(file = 'multinomial.stan')

fit <- sampling(model, data = data, seed = 42)

fit

fit_array <- as.array(fit)
mcmc_hist(fit_array)
```



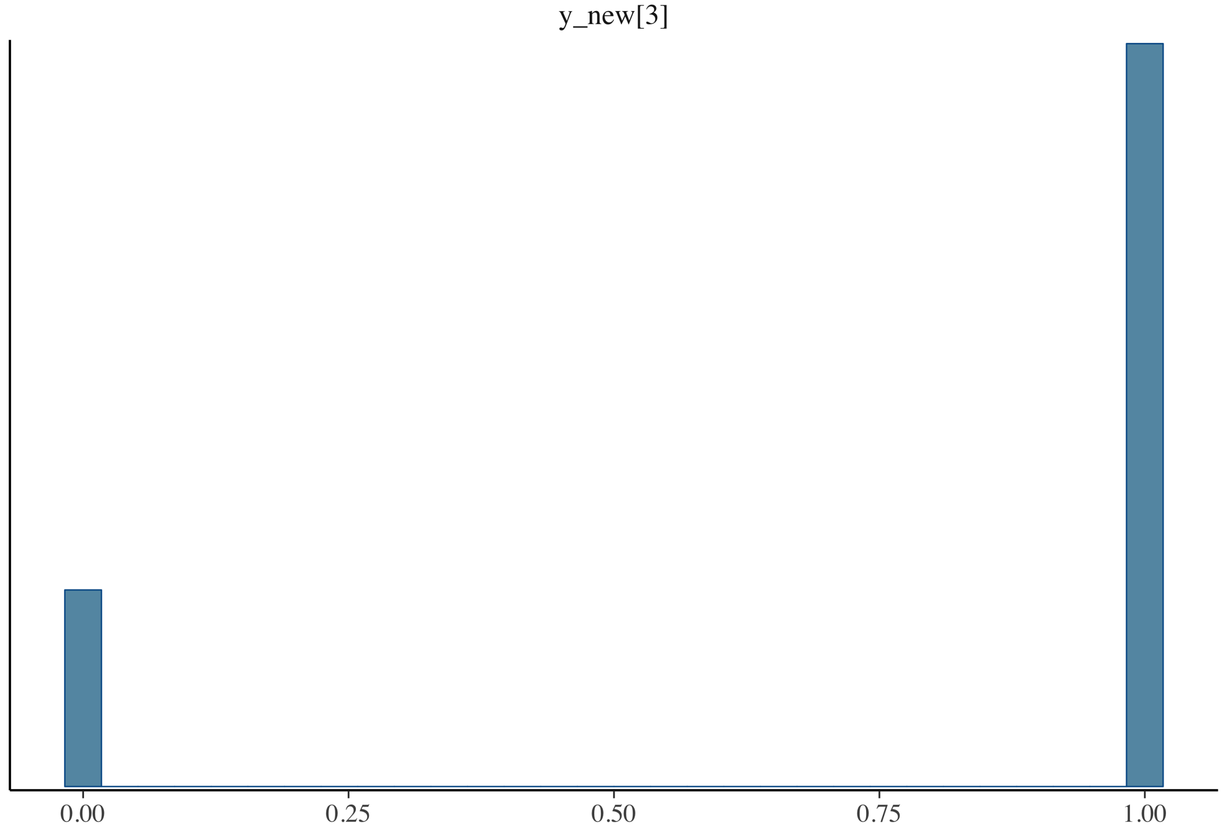

```
mcmc_trace(fit_array, pars = 'a')
```




```
mcmc_violin(fit_array, pars = 'a')
```




```
mcmc_hist(fit_array, pars = 'y_new[3]')
```




```
fit_array[ , 2, 1]
```