# Домашнее задание 1. RLMS + спецификация модели

22 ноября 2012 г.

## 1 Часть 1

1. Прочитайте про RLMS, http://www.hse.ru/rlms/

   Посмотрите описание проекта. Пролистайте вестник RLMS, чтобы иметь представление о том, какие исследования можно строить на основе RLMS.

2. Скачайте любую волну RLMS по своему выбору. Скачайте описание переменных.

   Пролистайте описание переменных. Там их больше тысячи. Попадаются довольно прикольные. Мне нравится pc9.6.5a, «У Вас есть GPRS навигатор?»

3. Загрузите данные в R.

   Данные RLMS выложены на сайте в формате SPSS. SPSS это потихоньку погибающий статистический пакет для домохозяек. Для чтения формата .sav в таблицу данных R можно сделать так

```
library(foreign)
file.name <- "/home/boris/downloads/r20hall23c.sav"
h <- read.spss(file.name, to.data.frame = TRUE)

## re-encoding from CP1251
```

   Первая команда, library(foreign), подгружает библиотеку R, в которой содержатся команды для чтения вражеских форматов, spss, stata, etc

   Описания переменных при этом также загружаются в таблицу данных. Можно их выделить в отдельный вектор и прочитать, например, про переменную pc9.631a.

```
var.labels <- attr(h, "variable.labels")
var.labels["pc9.631a"]

##                          pc9.631a
## "У Вас есть цифровая видеокамера?"
```

4. Выберите любую количественную переменную в качестве зависимой и несколько переменных в качестве объясняющей.

   Цель этой домашки скорее ознакомится с наличием мониторинга RLMS, поэтому можно не сильно заморачиваться с этим этапом. Хотя в реальности тут-то всё самое интересное и начинается. За оригинальные гипотезы будут плюшки.

5. Опишите выбранные переменные.

   Постройте симпатичные графики. Посчитайте описательные статистики. Много ли пропущенных наблюдений? Есть ли что-нибудь интересненькое?

6. Постройте регрессию зависимой переменной на объясняющие.

   Проверьте гипотезу о значимости каждого полученного коэффициента. Проверьте гипотезу о значимости регрессии в целом. Для нескольких коэффициентов (двух достаточно) постройте 95%-ый доверительный интервал.

7. Напишите свои пожелания и комментарии о курсе эконометрики

   Любые пожелания! Что не ясно в курсе эконометрики? Содержательные комментарии позволяют получить бонус. Искусная лесть оценивается :)

## 2 Часть 2.

Сравните распределение оценки коэффициента $\hat{\beta}_2$ графически и сделайте выводы в следующих восьми случаях, сделав 1000 экспериментов для каждого случая.

Во всех моделях:

1. 300 наблюдений
2. $x \sim N(0,1)$, $\varepsilon \sim N(0,1)$, $w \sim N(0,1)$ и независимы
3. $z = ax + (1-a)w$

Случаи:

1. В модель включена «лишняя» переменная, $a = 0$

   Истинная модель имеет вид $y = 2 + 3x + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$

2. В модель включена «лишняя» переменная, $a = 0.3$

   Истинная модель имеет вид $y = 2 + 3x + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$

3. В модель включена «лишняя» переменная, $a = 0.6$

   Истинная модель имеет вид $y = 2 + 3x + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$

4. В модель включена «лишняя» переменная, $a = 0.9$

   Истинная модель имеет вид $y = 2 + 3x + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x + \hat{\beta}_3 z$

5. В модели пропущена «нужная» переменная, $a = 0$

   Истинная модель имеет вид $y = 2 + 3x - 2z + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

6. В модели пропущена «нужная» переменная, $a = 0.3$

   Истинная модель имеет вид $y = 2 + 3x - 2z + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

7. В модели пропущена «нужная» переменная, $a = 0.6$

   Истинная модель имеет вид $y = 2 + 3x - 2z + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

(b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

8. В модели пропущена «нужная» переменная, $a = 0.9$

   Истинная модель имеет вид $y = 2 + 3x - 2z + \varepsilon$.

   С помощью графиков сравните распределение оценки $\hat{\beta}_2$ если

   (a) Оценивается истинная модель

   (b) Оценивается модель $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

## 3 Примеры в помощь

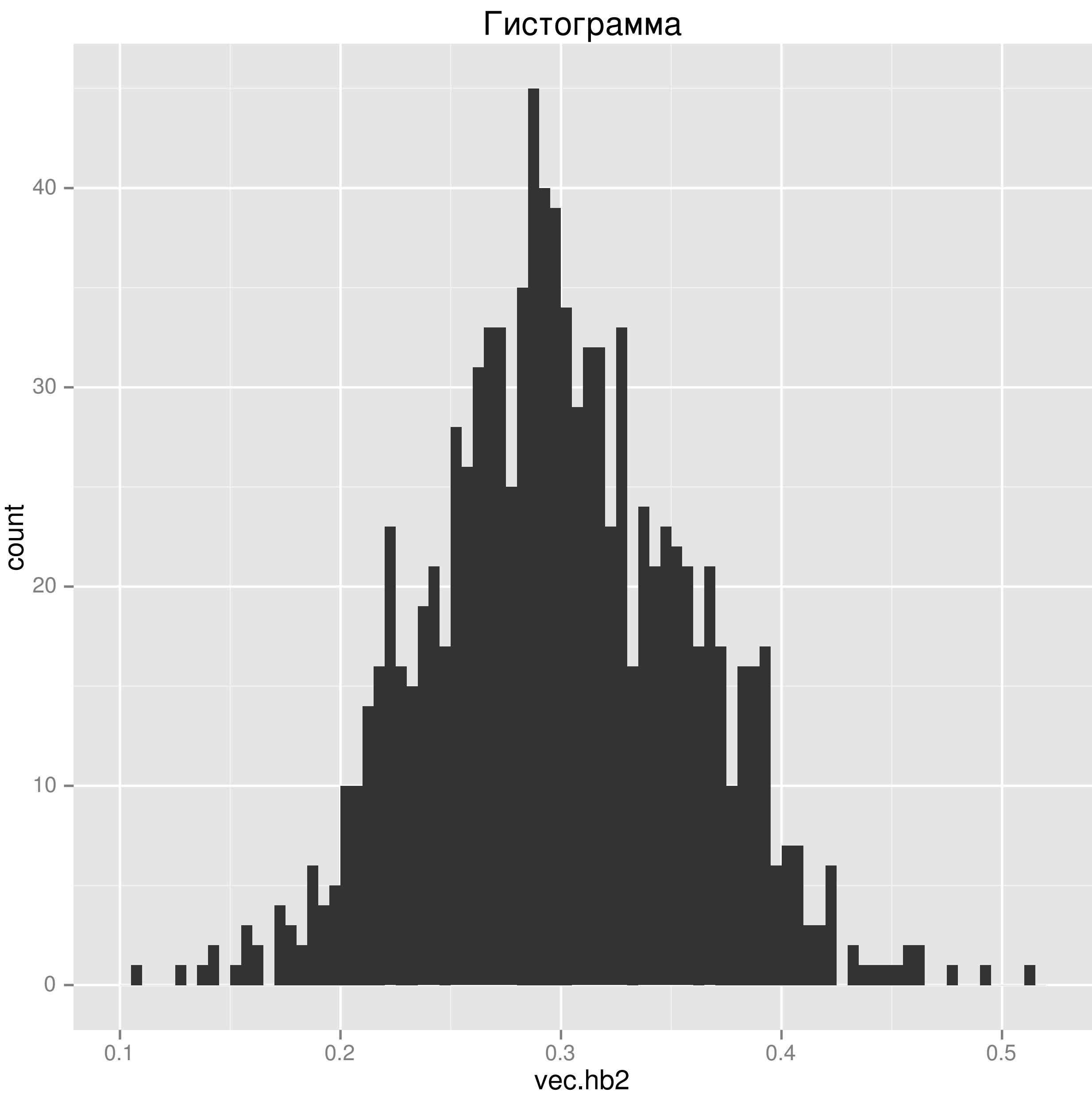
```
library(ggplot2)
# предварительно пакет ggplot2 надо скачать install.packages('ggplot2')
```

1. Пример 1. Оцениваем истинную модель

```
n.obs <- 300  # задаём число наблюдений

# задаём число экспериментов по оценке модели
n.exp <- 1000

vec.hb2 <- rep(0, n.exp)  # вектор из n.exp нулей

# n.exp раз будем делать то, что в фигурных скобках
for (i in 1:n.exp) {

    x <- rnorm(n.obs, mean = 0, sd = 1)
    epsilon <- rnorm(n.obs, mean = 0, sd = 1)
    y <- 2 + 0.3 * x + epsilon

    m1 <- lm(y ~ x)

    # запомним в переменную vec.hb2[i] второй коэффициент
    vec.hb2[i] <- m1$coefficients[2]
}
```

Гистограмма оценок $\hat{\beta}_2$:

```
qplot(vec.hb2, binwidth = 0.005, main = "Гистограмма")
```




2. Пример 2. Для сравнения законов распределения удобно их нарисовать на одном графике. Например, можно использовать один из трёх предложенных графиков. Можно использовать и другие графики! Эти три даны для примера, а не для ограничения свободы выбора.

```
# генерим данные
x1 <- rnorm(n.obs, mean = 0, sd = 1)
x2 <- rnorm(n.obs, mean = 1, sd = 4)

# из двух иксов создаём табличку, в первом столбце они будут идти друг за
# другом во втором столбце будет стоять название переменной
x.num <- factor(rep(c("X1", "X2"), each = n.obs))
x.combined <- data.frame(x.data = c(x1, x2), x.num)
```

- Две наложенные гистограммы

```
p1 <- ggplot(x.combined, aes(x = x.data, fill = x.num)) + geom_histogram(binwidth = 0.5,
    alpha = 0.5, position = "identity")
```

- Две чередующиеся гистограммы

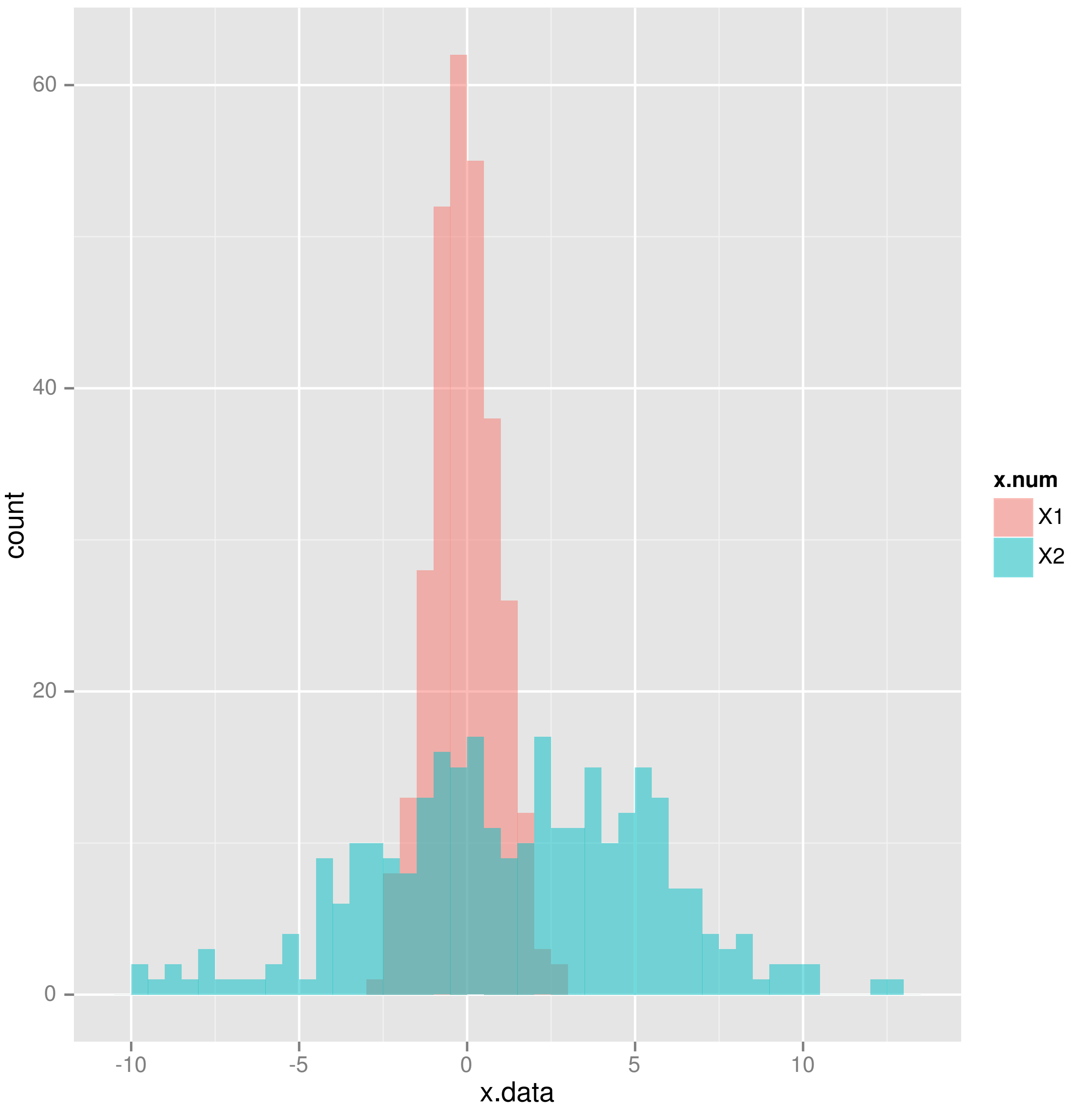
```
p2 <- ggplot(x.combined, aes(x = x.data, fill = x.num)) + geom_histogram(binwidth = 0.5,
    position = "dodge")
```

- Две сглаженные функции плотности

```
p3 <- ggplot(x.combined, aes(x = x.data, fill = x.num)) + geom_density(alpha = 0.3)
```

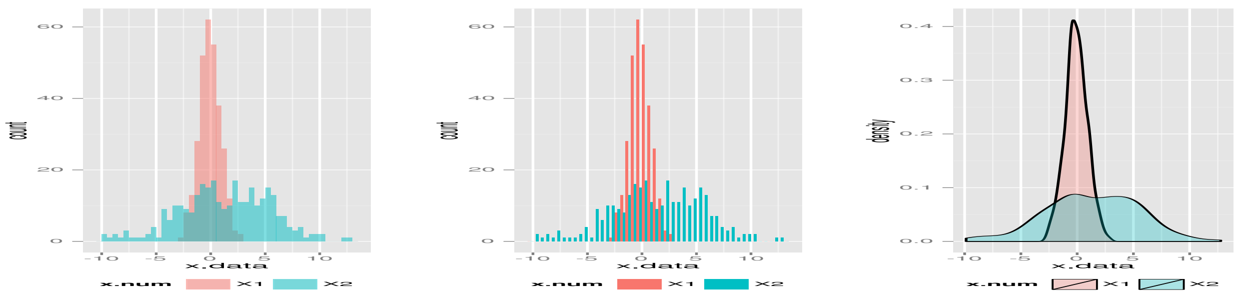
Выводим один график:

```
p1
```




Выводим три графика:

```
library(gridExtra)
bot.legend <- theme(legend.position = "bottom")

all.graphs <- arrangeGrob(p1 + bot.legend, p2 + bot.legend, p3 + bot.legend,
    nrow = 1)
all.graphs
```



count
x.data
x.num
X1
X2
density