

Язык R – наглядная статистика!

Зачем нужен R?

- Быстрая статистическая обработка данных
- Построение красивых графиков
- Бесплатный, удобный, быстрый для изучения язык!!!

Структура курса

- Базовый синтаксис
- Графики
- Генерация данных
- Инструменты для статистического анализа данных
- Анализ биоинформатических данных

Обучение и контроль

- Wiki
- <https://kodomo.fbb.msu.ru/wiki/2009/8/R#preview>
- Домашние задания (30% зачетных баллов)
- Контрольные работы

Для работы в компьютерном классе:

- Логин Class пароль Class
- Создать для работы директорию со своей фамилией
- Wifi: FBB

Среда разработки RStudio

- Достоинства:
- Бесплатная удобная среда разработки
- Где скачать: <http://www.r-studio.com/>

Помощь

- Форумы: *Stackoverflow, R mailing list, ...* (да поможет нам Google :-))

- Документация (<http://www.r-project.org>, *help(...)*)

- Полезная шпаргалка:

<http://r-language.ru/functions>

- Спросить преподавателя (ну в самом крайнем случае!)

R – векторизованный язык

- Основной тип данных – вектор (упорядоченный набор чисел)
- Идея – работать с набором данных как с одним числом (параллельно обрабатывать все значения набора)
- Это позволяет обходиться (в ряде случаев) без циклов

Вектор

```
> x<-1:5 ; y<-6:10
```

```
> x
```

```
[1] 1 2 3 4 5
```

```
> y
```

```
[1] 6 7 8 9 10
```

```
> x+y
```

```
[1] 7 9 11 13 15
```

```
> x*2
```

```
[1] 2 4 6 8 10
```

```
> x>4
```

```
[1] FALSE FALSE FALSE FALSE TRUE
```

```
> y==7
```

```
[1] FALSE TRUE FALSE FALSE FALSE
```

```
> x*y
```

```
[1] 6 14 24 36 50
```

Как можно создать вектор?

Оператор `c()`

```
> c(1, 2, 3)
```

```
[1] 1 2 3
```

Последовательности

```
> 1:10
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

```
> seq(from=1, to=8, by=2)
```

```
[1] 1 3 5 7
```

```
> seq(1, 10, 2)
```

```
[1] 1 3 5 7 9
```

Как можно создать вектор?

Объединение нескольких векторов

```
> x<-c(1, 2, 3)
```

```
> x<-c(x, 1:3); x
```

```
[1] 1 2 3 1 2 3
```

Повторы

```
> rep(0.5, 6)
```

```
[1] 0.5 0.5 0.5 0.5 0.5 0.5
```

Для целых чисел (работает быстрее)

```
> rep.int(1, 5)
```

```
[1] 1 1 1 1 1
```

Как можно создать вектор?

Распределение

- ✓ Нормальное распределение:
- ✓ `dnorm(x, mean = 0, sd = 1, log = FALSE)` – плотность распределения
- ✓ `pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)` – функция распределения
- ✓ `qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)` – квантильная функция
- ✓ `rnorm(n, mean = 0, sd = 1)` – случайная генерация

Случайная генерация из распределения:

```
> rnorm(5)
```

```
[1] 1.1568405 -0.8248219 0.1428891 -0.4784408 0.7561443
```

Равномерное

```
runif(n, min=0, max=1)
```

```
> runif(5, 0, 1)
```

```
[1] 0.1972687 0.3090867 0.2865924 0.1409635  
0.3441481
```

Биномиальное

```
rbinom(n, size, prob)
```

```
> rbinom(10, 100, 0.5)
```

```
[1] 54 47 55 50 47 45 52 45 58 52
```

Пуассона

```
rpois(n, lambda)
```

```
> rpois(10, 4)
```

```
[1] 2 3 2 4 10 3 2 3 5 6
```

Срезы

```
> x<-c(1, 5, 7, 9, 15, 3)
```

```
> x[1]
```

```
[1] 1
```

```
> x[2:4]
```

```
[1] 5 7 9
```

```
> x[c(2, 5)]
```

```
[1] 5 15
```

```
> x[-1]
```

```
[1] 5 7 9 15 3
```

```
> x[-(1:3)]
```

```
[1] 9 15 3
```

```
> x[x>5]
```

```
[1] 7 9 15
```

```
> x[x>5 & x<10]
```

```
[1] 7 9
```

Простейший статистический анализ

```
> x=rnorm(100)
```

Среднее

```
> mean(x)
```

```
[1] -0.04029328
```

Дисперсия

```
> var(x)
```

```
[1] 1.037552
```

Простейший статистический анализ

Минимальное и максимальное значения

```
> min(x)
```

```
[1] -2.605444
```

```
> max(x)
```

```
[1] 2.51254
```

Медиана

```
> median(x)
```

```
[1] -0.1039548
```

Квантили

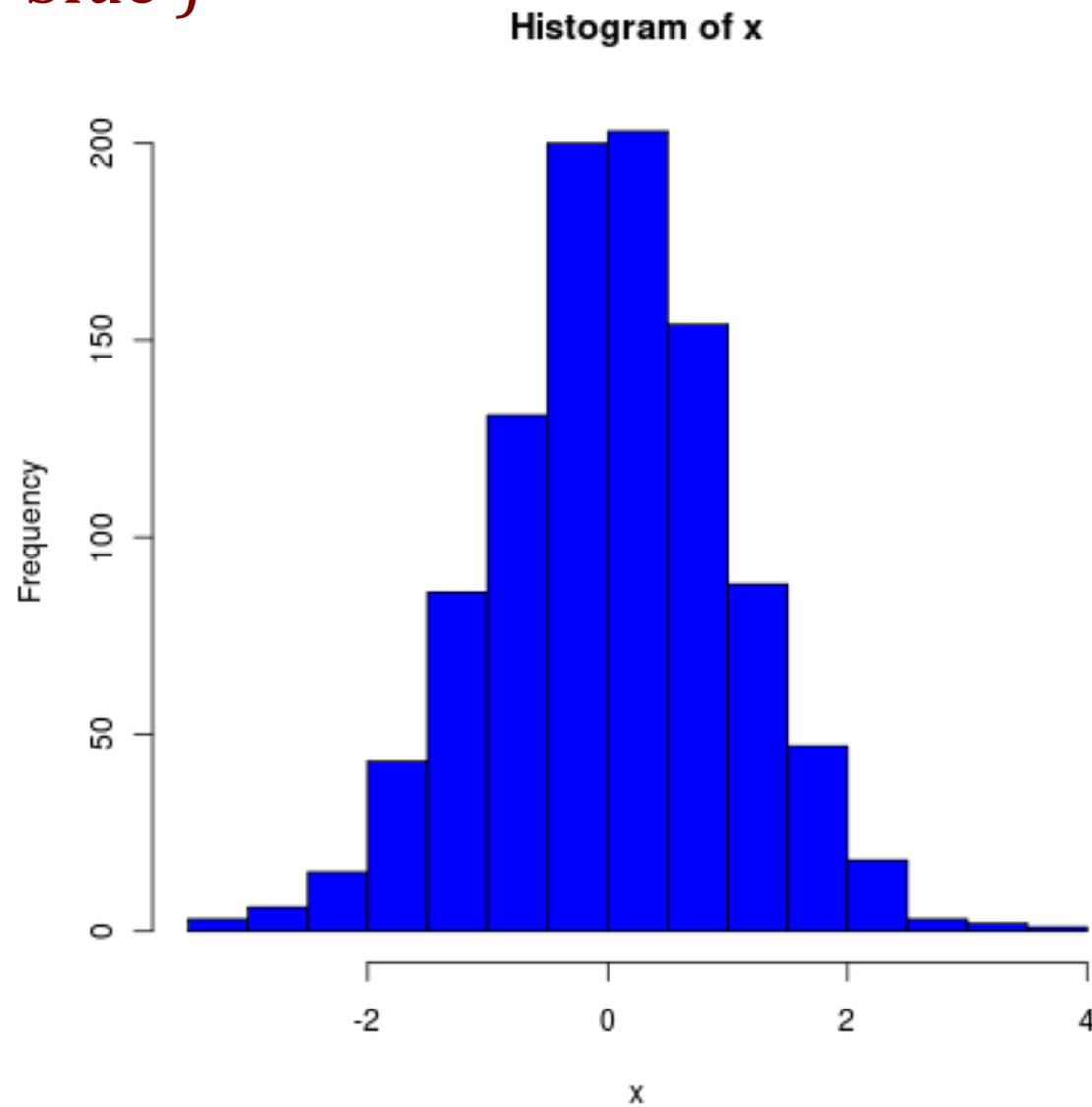
```
> quantile(x)
```

0%	25%	50%	75%	100%
-2.6054443	-0.6321819	-0.1039548	0.4765935	2.5125400

А еще данные удобно анализировать
с помощью графиков!

Гистограммы

```
> x=rnorm(1000)  
> hist(x, col='blue')
```



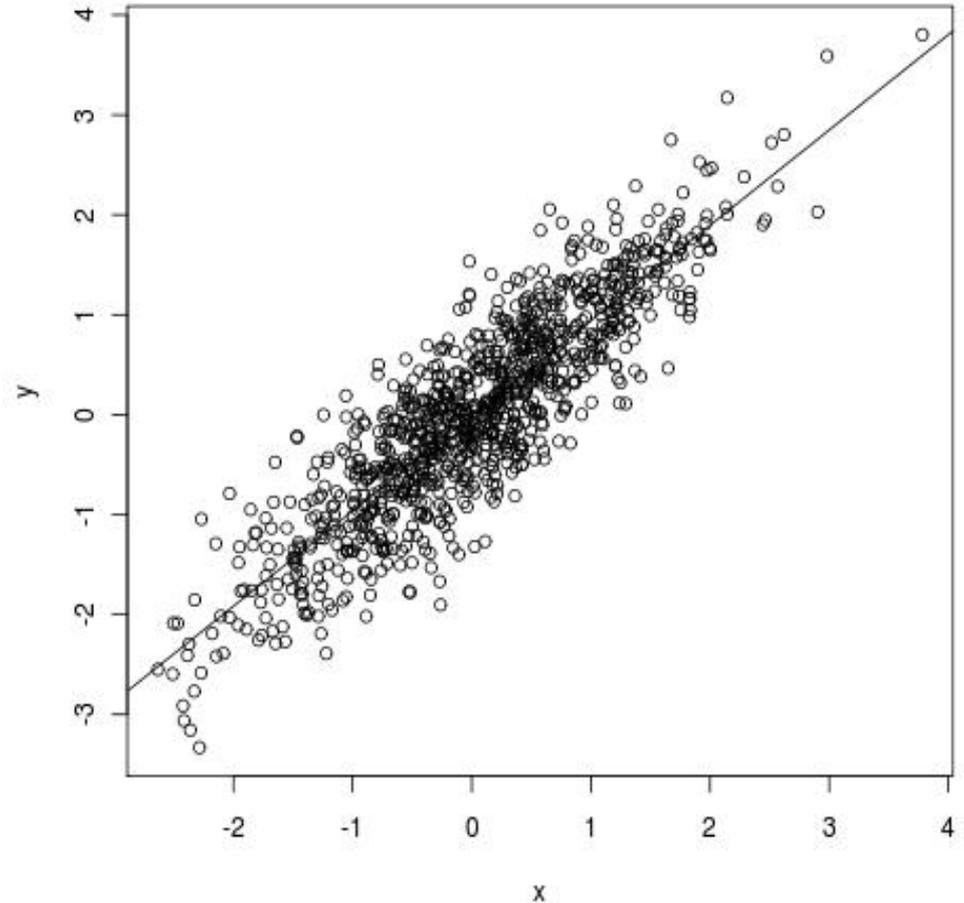
Точки

```
> x<-rnorm(1000)
```

```
> y<-x+rnorm(1000, sd=0.5)
```

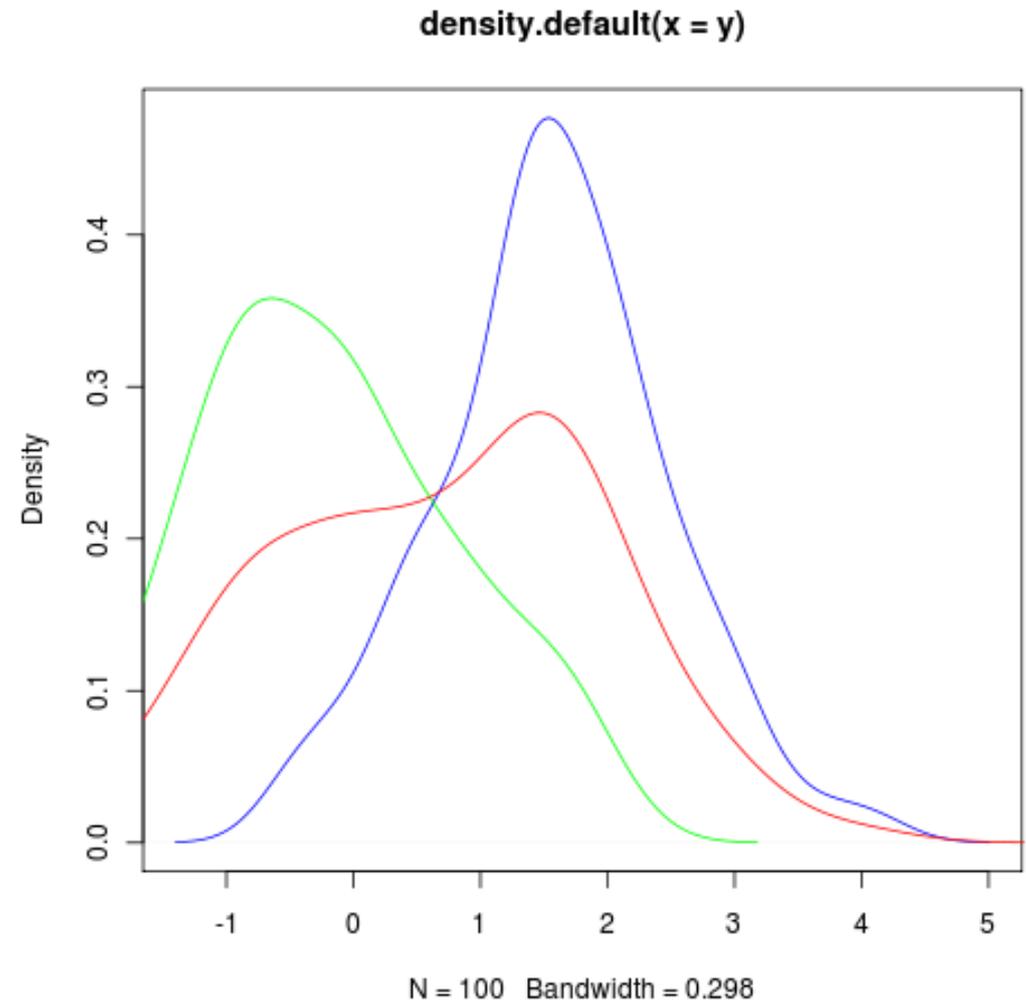
```
> plot(x, y)
```

```
> abline(lm(y~x))
```



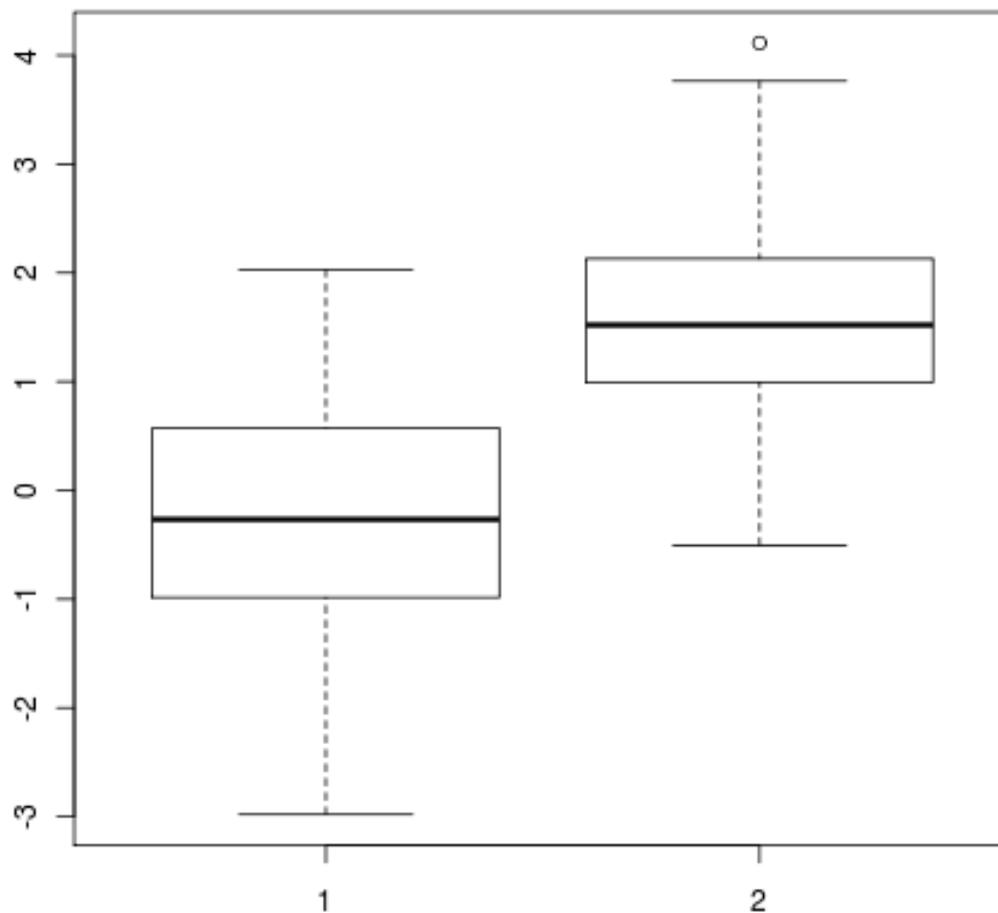
Линии

```
> x<-rnorm(100)
> y<-rnorm(100, mean = 1.5)
> plot(density(y), col = 'blue')
> lines(density(x), col = 'green')
> lines(density(c(x, y)), col = 'red')
```



Ящик с усами

`> boxplot(x,y)`



Сохранение графика в файл

```
> png(file="Pictures/boxplot.png")
```

```
> boxplot(x,y)
```

```
> dev.off()
```

Домашнее задание

- ✓ Задание Вы найдете здесь:
<https://kodomo.fbb.msu.ru/wiki/2009/8/R#preview>
- ✓ Ответы отправляйте по адресу fbbrcourse@gmail.com, в теле письма ответы приводите форме "номер задания [пробел] вариант ответа»:

1 1

2 1

3 3

5 2

- ✓ Можно отвечать не на все вопросы (в приведенном примере отсутствует ответ на четвертый вопрос).
- ✓ Если есть какие-то комментарии/вопросы по заданию, пишите их после ответов на вопросы (кстати, вопросы и пожелания также отправлять и без ответов на домашнее задание).
- ✓ Мы постараемся создать более удобную систему для учета ответов на домашнее задание уже скоро!