Векторы, датафреймы  
  
[**seq**](http://inside-r.org/r-doc/base/seq)(from = 1, to = 5, [**by**](http://inside-r.org/r-doc/base/by) = 0.5)

[**rep**](http://inside-r.org/r-doc/base/rep)("test", 5)

[**c**](http://inside-r.org/r-doc/base/c)("a", "b", "c", 1, 2)

y <- [**c**](http://inside-r.org/r-doc/base/c)(5, 3, 2, 6, 1)

y[3:5]

y[-с(1, 4)]

y[y>2]

emp.data <- data.frame(

emp\_id = c (1:5),

emp\_name = c("Rick","Dan","Michelle","Ryan","Gary"),

salary = c(623.3,515.2,611.0,729.0,843.25),

start\_date = as.Date(c("2012-01-01", "2013-09-23", "2014-11-15", "2014-05-11",

"2015-03-27")),

stringsAsFactors = FALSE

)

data.frame(emp.data$emp\_name,emp.data$salary)

emp.data[1:2,]

emp.data[c(3,5),c(2,4)]

emp.data$dept <- c("IT","Operations","IT","HR","Finance")

emp.newdata <- data.frame(

emp\_id = c (6:8),

emp\_name = c("Rasmi","Pranab","Tusar"),

salary = c(578.0,722.5,632.8),

start\_date = as.Date(c("2013-05-21","2013-07-30","2014-06-17")),

dept = c("IT","Operations","Fianance"),

stringsAsFactors = FALSE

)

# Bind the two data frames.

emp.finaldata <- rbind(emp.data,emp.newdata)

df <- data.frame(var1=c(11,21,31), var2=c(12,22,32), var3=c(13,23,33), var4=c(14,24,34), row.names=c("case1", "case2", "case3"))

df[,1, drop=FALSE]

Числовые характеристики выборок

Mean(iris)

population <- rnorm(3, 2, n = 1000)

hist(population, breaks = 20, col = "tomato", freq = TRUE,

main ="Распределение выборочных средних",

xlab = "средние значения по выборкам" ,

ylab = "частоты")

sd(iris)

str(iris)

summary(iris)

*# строим примерно такую же гистограмму, как и до этого*

*# но добавляем аргумент xlim, который отвечает за то, в каких границах по оси X*

*# будет построен график*

*# нужно для того, чтобы последующие гистограммы построить одинаково,*

*# иначе R их растянет по-своему, и сравнивать их будет неудобно*

hist(sample\_means, breaks = 20, col = "tomato", freq = TRUE,

main ="Распределение выборочных средних",

sub = "n = 2",

xlab = "средние значения по выборкам" ,

ylab = "частоты",

xlim = c(-3, 9))

*# добавляем линию для среднего генеральной совокупности*

*# v = 3 -- вертикальная прямая x = 3*

*# lwd -- толщина линии (от line width)*

*# lty -- тип линии, 2 отвечает за пунктир (от line type)*

abline(v = 3, col = "blue", lwd = 2, lty = 2)

*# гистограмма, как и раньше, но есть важное отличие:*

*# freq = FALSE -- хотим, чтобы по оси Y отмечались не частоты, а значения плотности*

*# здесь это нужно, так как дальше на гистограмму будем наносить график функции плотности*

hist(sample\_means\_un, breaks = 20, col = "tomato", freq = FALSE,

main ="Распределение выборочных средних",

xlab = "средние значения по выборкам" ,

ylab = "плотность", ylim = c(0, 5))

*# график плотности нормального распределения (dnorm)*

*# нормальное распределение со средним и ст.отклонением*

*# add = TRUE, чтобы добавить к предыдущему графику (гистограмме)*

curve(dnorm(x, mean = mean(sample\_means\_un), sd = sd(sample\_means\_un)),

add = TRUE, col = "navy", lwd = 2)