

#Удаляем ранее созданные объекты

```
rm(list =ls() )
```

#Загружаем таблицу

```
df <- read.csv2("~/ДОКИ/Кафедра/ВГУ/Перспективные инф технологии/К ПЗ множ  
регр.csv")
```

Показать таблицу

```
View(df)
```

#Структура таблицы

```
str(df)
```

#Построим гистограмму по переменной Y

```
hist(df$Y,main="Гистограмма частот", xlab="Y")
```

Справка по функции hist

```
?hist
```

Расчет количества интервалов

```
k<-round(1+3.322*log10(length(df$Y)))
```

#шаг

```
h<-(max(df$Y)-min(df$Y))/k
```

1 способ

```
hist(df$Y, main="Гистограмма",xlab="Y", breaks = k)
```

2 способ

```
hist(df$Y, main="Гистограмма", breaks=seq(min(df$Y), max(df$Y), length.out=k+1),  
col="lightgreen")
```

#3 способ

```
library("ggplot2")
```

```
ggplot(df,aes(Y))+geom_histogram(binwidth =h, color="black", fill="lightgreen")
```

#Описательные статистики

Квантили для всех показателей в таблице df

```
summary(df)
```

#Выборочное среднее

```
mean(df$Y)
```

#Медиана

```
median(df$Y)
```

Квантили

```
quantile(df$Y)
```

#Квантили от 0 до 1 с шагом 0,20

```
quantile(df$Y, seq(from = 0, to = 1, by = 0.20))
```

Дополнительные пакеты для нахождения описательных статистик

#1 пакет: pastecs

```
install.packages("pastecs")
```

```
library("pastecs")
```

```
stat.desc(df)
```

#В результате по каждой переменной будут найдены значения следующих параметров

#nbr.val – число значений переменной,

#nbr.null – число нулевых значений,

#nbr.na – число пропущенных значений,

#min – мин значение переменной,

#max – максимальное значение переменной,

#range – размах переменной (разница между макс и мин),

#sum – сумма,

#median – медиана,

#mean – среднее значение,

#SE.mean – стандартная ошибка для среднего,

#CI.mean – длина 95% доверительного интервала,

#Var – дисперсия,

#std.dev – стандартное отклонение,

#coef.var – коэффициент вариации.

#2 пакет: psych

install.packages("psych")

library(psych)

describe(df\$Y)

#В результате по каждой переменной будут найдены значения следующих параметров:

vars – это сколько переменных участвуют в вычислениях

#(команда describe(df\$Y) прописывает вычисления лишь для

переменной Y),

n – число наблюдений,

mean – среднее арифметическое,

sd – среднее квадратическое отклонение,

median – медиана,

trimmed – усечённое среднее (рассчитывается следующим образом: значения переменной упорядочиваются по возрастанию, далее убираются 5% значений (наблюдений) слева и 5% справа, по оставшимся значениям рассчитывается простое среднее арифметическое. Используется, когда в выборке много нетипичных значений (выбросов)),

mad – медиана,

min, max – минимальное и максимальное значение,

range – размах,

skew – коэффициент асимметрии,

kurtosis – коэффициент эксцесса,

se – стандартная ошибка среднего.

3 пакет: stargazer

```
install.packages("stargazer")
```

```
library(stargazer)
```

```
stargazer(df, type="text", title="Descriptive statistics")
```

#Другие графики

par() в данном случае делит область на одну строку и две колонки,

куда будут помещаться графики

```
par(mfrow=c(1,2))
```

ящик с усами

```
boxplot(df$Y, main = "Ящик с усами для Y",
```

```
      ylab = "Y")
```

```
boxplot(df$X1, main = "Ящик с усами для X1",
```

```
      ylab = "X1")
```

```
par(mfrow=c(1,1))
```

```
#Горизонтальный ящик с усами
```

```
boxplot(df$Y, main = "Ящик с усами для Y",  
        ylab = "Y", horizontal = TRUE)
```

```
#Добавляет точки к ящику, соответствующие значениям Y
```

```
stripchart(df$Y, method = "jitter", pch = 19, add = TRUE, col = "blue")
```

```
#Дисперсия
```

```
var(df$Y)
```

```
#Среднее квадратическое отклонение
```

```
sd(df$Y)
```

```
#Диаграмма рассеяния
```

```
#1 способ
```

```
plot(x = df$X1, y = df$Y, main = "Диаграмма рассеяния", xlab = "X1", ylab = "Y",  
     col="#1b98e0", pch = 16)
```

```
#2 способ
```

```
# Создаем последовательность от 100 до 280 с шагом 50, чтобы потом
```

```
# использовать для разметки оси абсцисс
```

```
gr<-seq(from = 100, to = 280, by = 50)

ggplot(df,aes(x=X1,y=Y))+geom_point(fill="black")+

  scale_x_continuous(breaks=gr)
```

```
#Изобразим все возможные диаграммы рассеяния

pairs(df)
```

Ковариация Y и X1

```
cov(df$Y, df$X1)
```

#Коэффициент корреляции

```
cor(df$Y,df$X1)
```

#Корреляционная матрица

```
cor(df)
```

Тепловая карта

```
heatmap(cor(df), Rowv = NA, Colv = NA)
```

#Доверительный интервал для среднего

```
install.packages("DescTools")
```

```
library("DescTools")
```

```
MeanCI(df$Y, conf.level = 0.99)
```