

Московский государственный университет имени М. В. Ломоносова
Факультет Вычислительной математики и кибернетики

Никонов Максим Викторович
316 группа

2020

Задание 1: Ознакомиться с процедурой бутстреп (bootstrap) процедурой.

Функция для получения R-Squared из данных:

```
rsq <- function(formula, data, indices)
{
  d <- data[indices,]
  fit <- lm(formula, data = d)
  return(summary(fit)$r.square)
}
```

Bootstrapping с 1000 replications

```
bootobject <-
boot(data = data,
statistic=rsq,
R = 1000, formula =
Volume~Date+Close)
```

Результат

```
bootobject
plot(bootobject)
```

95% доверительный интервал

```
boot.ci(bootobject,
type="bca")
```

boot.ci(bootobject, conf=, type=) where

parameter	description
bootobject	The object returned by the boot function
conf	The desired confidence interval (default: conf=0.95)
type	The type of confidence interval returned. Possible values are "norm", "basic", "stud", "perc", "bca" and "all" (default: type="all")

Используемые packages. dplyr, boot, sm

Тестирование. Не требует

Неразрешенные вопросы. Нет

Новые функции. boot.ci, boot

Статус компиляции. НЕ ОК. Зависает. Данные из протокола:

```
> bootobject <- boot(data = data, statistic=rsq, ~
+ R = 1000, formula = Volume~Date+Close) ~
^C^M
^D^N^H^M
^M
x^M
x=5^M
^D^N^H^D^N^H^C^M
Выполнение остановлено^M
^I[1m^[[7m%^[[27m^[[1m^[[0m
^M ^M^[[7;file://MBP-
Nikon.Dlink/Users/Nikon/Desktop/CMC%20MSU/MC/5%20sem/R/tz6/1^G^M^[[0m^[[27m^[[24m^[[
JNikon@MBP-Nikon 1 % ^[[K^[[?2004he^Hexit^[[?2004l^M^M
```

Задание 2-3: Сгенерировать данные из нормального распределения с различными параметрами и провести анализ с помощью графиков квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности, рассмотренных на семинаре (6 тестов). Рассмотреть выборки малого (не более 50-100 элементов) и умеренного (1000-5000 наблюдений) объемов.

Сгенерировать данные из комбинаций, реализованных в R распределений, а затем провести анализ с помощью графиков квантилей, метода огибающих, а также стандартных процедур проверки гипотез о нормальности. Рассмотреть выборки малого и умеренного объемов. Сравнить эффективность методов

```
small_data = data.frame(
  Factor = rep(c("1", "2"), each = 50),
  Variable = c(rnorm(50, 5, 2),
               rnorm(50, 4, 3))
)
big_data = data.frame(
  Factor = rep(c("1", "2"), each = 2500),
  Variable = c(rnorm(2500, 5),
               rnorm(2500, 4))
)
```

Создаем 2 frame
нормального распределения
разных размерностей

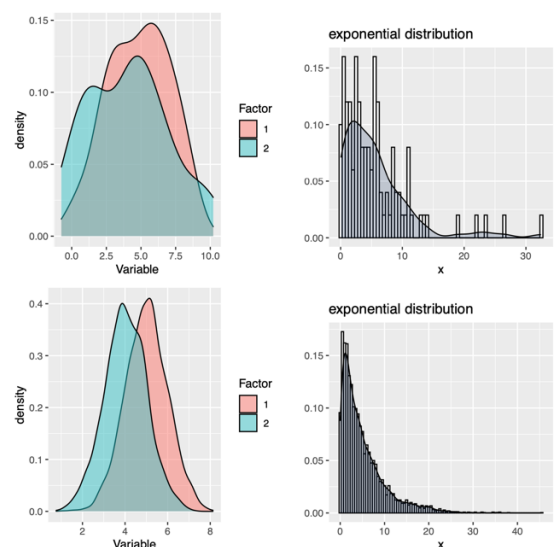
Чередуются 1,2,1,2,...и
соот-ие значения

```
set.seed(123)
lambda <- 0.2
n <- 5000
data_expBig <- data.frame(x = rexp(n, lambda))
q <- 100
data_expSmall <- data.frame(x = rexp(q, lambda))
```

Экспонен-ое распределение

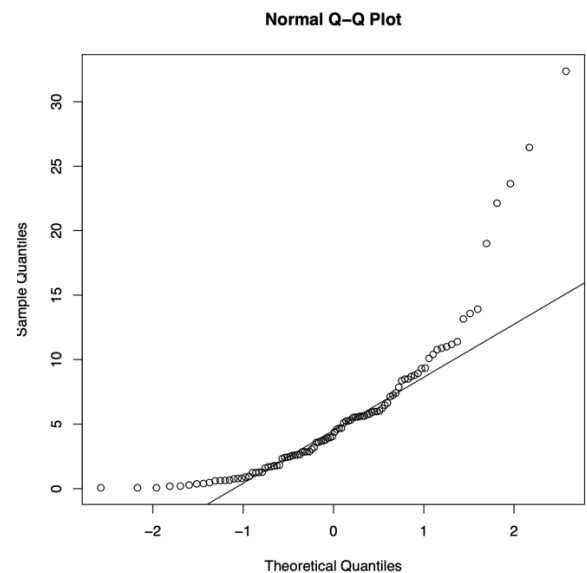
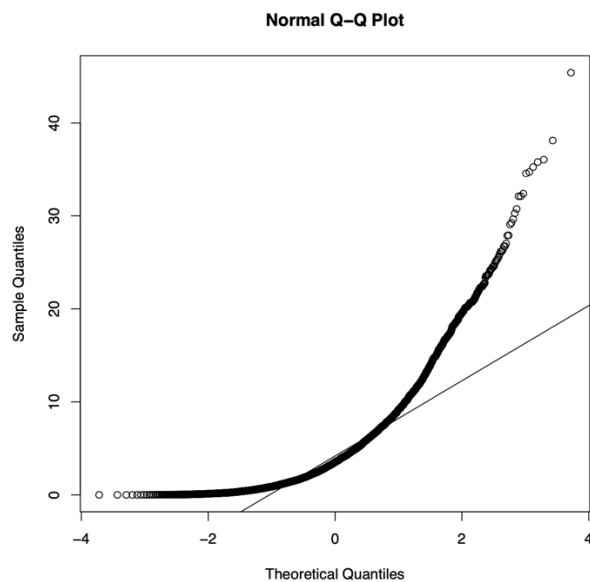
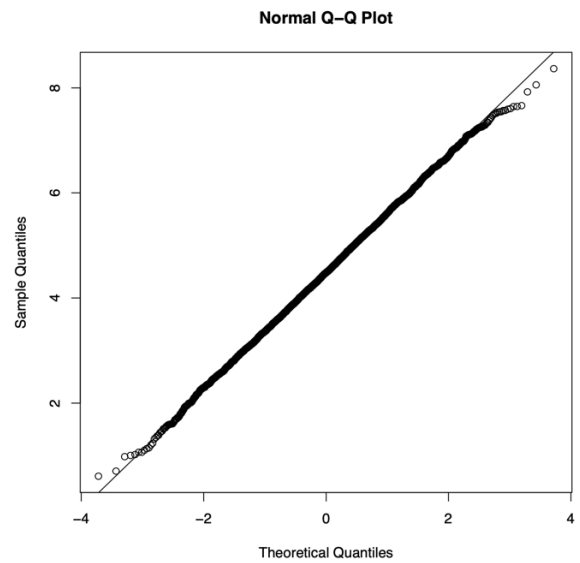
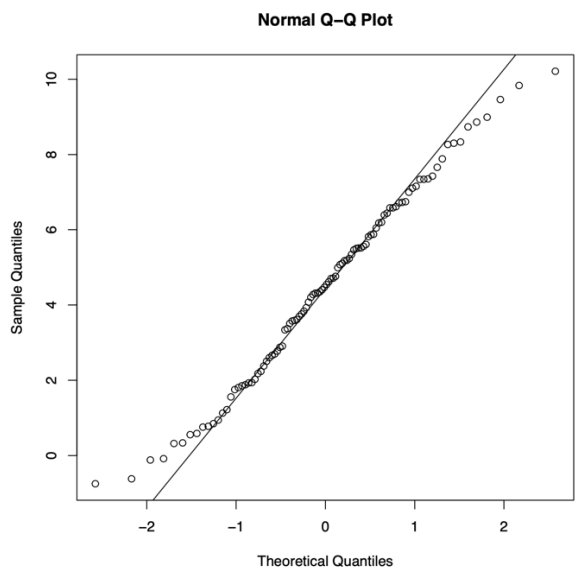
```
graf_small <- ggplot(small_data,
  aes(Variable, group = Factor, fill =
  Factor)) + geom_density(alpha = 1/2)
graf_big <- ggplot(big_data,
  aes(Variable, group = Factor, fill =
  Factor)) + geom_density(alpha = 1/2)
ExponentialBig <- (ggplot(data_expBig,
  aes(x = x)) + ylab("") +
  geom_histogram(aes(y=..density..),
  binwidth=0.5, color="black",
  fill="white") +
  ggtitle("
  exponential distribution") +
  geom_density(alpha = 0.2, fill =
  "#102C60"))
ExponentialSmall <-
  (ggplot(data_expSmall, aes(x = x)) +
  ylab("") +
  geom_histogram(aes(y=..density..),
  binwidth=0.5, color="black",
  fill="white") + ggtitle("
  exponential distribution") +
  geom_density(alpha = 0.2, fill =
  "#102C60"))
graf_small + facet_grid()
graf_big + facet_grid()
plot(ExponentialBig)
plot(ExponentialSmall)
grid.arrange(graf_small,
  ExponentialSmall, graf_big,
  ExponentialBig, ncol = 2)
```

Визуализация отдельно и вместе



```
qqnorm(small_data$Variable)
qqline(small_data$Variable)
qqnorm(big_data$Variable)
qqline(big_data$Variable)
qqnorm(data_expBig$x)
qqline(data_expBig$x)
qqnorm(data_expSmall$x)
qqline(data_expSmall$x)
```

Графики квантилей:



Проверка на нормальность методом огибающих :

```
x <- sort(rweibull(100, 2, (1 + 1.21*rbinom(100, 1, 0.05)) ))
c(mean(x), sd(x))
SmallV <- sort(rweibull(100, 2, (1 + 1.21*rbinom(100, 1, 0.05)) ))
c(mean(small_data$Variable), sd(small_data$Variable))
```

Данные из протокола:

```
> x <- sort(rweibull(100, 2, (1 + 1.21*rbinom(100, 1, 0.05)) ))
> c(mean(x), sd(x))
[1] 0.9337549 0.4389801
> SmallV <- sort(rweibull(100, 2, (1 + 1.21*rbinom(100, 1, 0.05)) ))
> c(mean(small_data$Variable), sd(small_data$Variable))
[1] 4.700931 2.290999
```

Проверка на нормальность:

```
shapiro.test(small_data$Variable)
ad.test(small_data$Variable)
cvm.test(small_data$Variable)
lillie.test(small_data$Variable)
sf.test(small_data$Variable)
shapiro.test(big_data$Variable)
ad.test(big_data$Variable)
cvm.test(big_data$Variable)
lillie.test(big_data$Variable)
sf.test(big_data$Variable)
shapiro.test(data_expBig$x)
ad.test(data_expBig$x)
cvm.test(data_expBig$x)
lillie.test(data_expSmall$x)
sf.test(data_expSmall$x)
shapiro.test(data_expBig$x)
ad.test(data_expBig$x)
cvm.test(data_expBig$x)
lillie.test(data_expBig$x)
sf.test(data_expBig$x)
```

Данные из протокола:

См. След страницу =)

```
> shapiro.test(small_data$Variable)

      Shapiro-Wilk normality test

data:  small_data$Variable
W = 0.98139, p-value = 0.1704

> ad.test(small_data$Variable)

      Anderson-Darling normality test

data:  small_data$Variable
A = 0.66736, p-value = 0.07908

> cvm.test(small_data$Variable)

      Cramer-von Mises normality test

data:  small_data$Variable
W = 0.12176, p-value = 0.05575

> lillie.test(small_data$Variable)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  small_data$Variable
D = 0.078397, p-value = 0.1367

> sf.test(small_data$Variable)

      Shapiro-Francia normality test

data:  small_data$Variable
W = 0.98436, p-value = 0.2426

> shapiro.test(big_data$Variable)

      Shapiro-Wilk normality test

data:  big_data$Variable
W = 0.99969, p-value = 0.6849

> ad.test(big_data$Variable)

      Anderson-Darling normality test

data:  big_data$Variable
A = 0.19817, p-value = 0.8868

> cvm.test(big_data$Variable)

      Cramer-von Mises normality test

data:  big_data$Variable
W = 0.027485, p-value = 0.8791

> lillie.test(big_data$Variable)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  big_data$Variable
D = 0.0062156, p-value = 0.9112

> sf.test(big_data$Variable)

      Shapiro-Francia normality test

data:  big_data$Variable
W = 0.99975, p-value = 0.7855

> shapiro.test(data_expBig$x)

      Shapiro-Wilk normality test

data:  data_expBig$x
W = 0.81615, p-value < 2.2e-16

> ad.test(data_expBig$x)

      Anderson-Darling normality test

data:  data_expBig$x
A = 239.87, p-value < 2.2e-16

> cvm.test(data_expBig$x)

      Cramer-von Mises normality test

data:  data_expBig$x
W = 40.582, p-value = 7.37e-10

Предупреждение:
В cvm.test(data_expBig$x) :
  p-value is smaller than 7.37e-10, cannot be computed more accurately
> lillie.test(data_expSmall$x)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  data_expSmall$x
D = 0.16956, p-value = 1.712e-07

> sf.test(data_expSmall$x)

      Shapiro-Francia normality test

data:  data_expSmall$x
W = 0.77763, p-value = 1.39e-09

> shapiro.test(data_expBig$x)

      Shapiro-Wilk normality test

data:  data_expBig$x
W = 0.81615, p-value < 2.2e-16

> ad.test(data_expBig$x)

      Anderson-Darling normality test

data:  data_expBig$x
A = 239.87, p-value < 2.2e-16

> cvm.test(data_expBig$x)

      Cramer-von Mises normality test

data:  data_expBig$x
W = 40.582, p-value = 7.37e-10

Предупреждение:
В cvm.test(data_expBig$x) :
  p-value is smaller than 7.37e-10, cannot be computed more accurately
> lillie.test(data_expBig$x)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  data_expBig$x
D = 0.15853, p-value < 2.2e-16

> sf.test(data_expBig$x)

      Shapiro-Francia normality test

data:  data_expBig$x
W = 0.81595, p-value < 2.2e-16

>
```

Используемые packages. ggplot2, gridExtra, nortest

Тестирование. Не требует

Неразрешенные вопросы. Нет

Новые функции. ggplot, facet_grid(), qqnorm, qqline, rweibull, shapiro.test, ad.test, cvm.test, lillie.test, sf.test

Статус компиляции. ОК. Данные из протокола:

```
^M
> ^M
^[[1m^[[7m%^[[27m^[[1m^[[0m
^M ^M^[[17;file://MBP-Nikon.Dlink/Users/Nikon/Desktop/CMC%20MSU/MC/5%20sem/R/tz6/2-
3^G^M^[[0m^[[27m^[[24m^[[JNikon@MBP-Nikon 2-3 % ^[[K^[[?2004he^Hexit^[[?2004l^M^M

Script done on Tue Oct 6 18:57:59 2020
```