

Практикум на ЭВМ
Роор Даниил, группа 316

Задание 5

1. С помощью функций пакета outliers с помощью формальных критериев для идентификации выбросов проверить в данных, является ли максимальное наблюдение выбросом.

```
df = read.csv('COVID-19-in-Italy.csv')  
  
infected = df[df$region == 'Abruzzo', ]$new_positive  
  
library(mitools)  
library(pan)  
library(mix)  
library(outliers)  
library(Hmisc)  
  
grubbs.test(infected)
```

Output:

Grubbs test for one outlier

```
data: infected  
G = 4.34008, U = 0.87848, p-value = 0.0006022  
alternative hypothesis: highest value 160 is an outlier
```

pvalue мало (< 0.05), принимаем гипотезу о том, что выброса нет

Для теста Диксона требуется выборка малого размера (не более 30 наблюдений)

```
dixon.test(infected[1:30])
```

Output:

Dixon test for outliers

```
data: infected[1:30]  
Q = 0.37705, p-value = 0.09823  
alternative hypothesis: highest value 122 is an outlier
```

pvalue велико (> 0.05), отвергаем гипотезу о том, что выброса нет, в пользу альтернативы: максимальное наблюдение является выбросом

```
chisq.out.test(infected)
```

Output:

chi-squared test for outlier

```
data: infected  
X-squared = 18.836, p-value = 1.424e-05  
alternative hypothesis: highest value 160 is an outlier
```

pvalue мало (< 0.05), принимаем гипотезу о том, что выброса нет