

R: Сходимость-ЗГЧ-проверка за норм. разпр. данни (2020-11-23)

qqplot сравнява квантила на данните ни с квантила на нормалното разпределение (или на експоненциалното и на др. разпределения)

Ако раздвигим средното - линията ще се премести нагоре или надолу;

Ако раздвигим стандартното отклонения - линията ще се завърти в центъра и опашките ще започнат да се раздаличат от нея, тъй като при сравняването разпределение ще е с по-голяма вероятност да имаме по-далечно наблюдение от средното.

```
qqnorm(X); qqline(X)
```

```
qqplot.das(X, "norm") # "exp", "unif", etc.
```

```
simple.eda(homedata$y1970)
```

```
simple.eda(log(homedata$2000[homedata$y2000>0]))
```

Тестове за нормално разпределение

Jarque-Bera test:

Тества само за нормално разпределение.

Проверява дисперсията на изместеността (skewness) и изостреността (kurtosis) на наблюденията спрямо сравняваните наблюдения.

```
install.packages("tseries")  
library(tseries)
```

```
X=rnorm(n=1000, mean=3, sd=4)  
jb=jarque.bera.test(X)  
if(jb$p.value>0.5) print("normal") else print("not normal")
```

Shapiro-Wilk test:

Тества само за нормално разпределение.

```
sh=shapiro.test(X)
```

Anderson-Darling test:

Тества само за нормално разпределение.

```
install.packages("nortest")  
library(nortest)
```

```
ad=ad.test(X)
```

Kolmogorov-Smirnov test:

При този тест се взима най-голямата разлика при сравняването на разпределението на нашите данни с тези спрямо които ги сравняваме (сравняваме разликите в CDF – кумулативната функция на разпределението)

```
ks.test(X, "pnorm", mean(X), sd(X))
```

```
X=rexp(n=1000, rate=10)
```

```
fit=fitdistr(X, "exponential")
```

```
# "Poisson", "geometric", "chi-squared", "cauchy", "negative-binomial", "gamma", etc.
```

```
ks=ks.test(X, "pexp", fit$estimate)
```

```
if(ks$p.value>0.05) print("exponential") else print("not exponential")
```

```
# [1] "exponential"
```