УВОД У СТАТИСТИКУ час 10

3. мај ’17.

Пример 1

У извештају локалног риболовачког друштва наведени су резултати студије у којој је испитивана контаминација живом популације пастрмског гргеча (рибе „бас“).

Прикупљен је узорак рибе из 53 језера на локалу и измерена концентрација живе у мишићном ткиву (у ppm).

Сматрајући да је дисперзија посматраног обележја коначна одредити апроксимативни 95% интервал поверења за мат. очекивање m тог обележја.

Подаци из узорка су:

```
 [1] 1.230 0.490 0.490 1.080 0.590 0.280 0.180 0.100 0.940 1.330 0.190 1.160 0.980 0.340 0.340 0.190 0.210 0.400 0.040
[20] 0.830 0.050 0.630 0.340 0.750 0.040 0.860 0.430 0.044 0.810 0.150 0.560 0.840 0.870 0.490 0.520 0.250 1.200 0.710
[39] 0.190 0.410 0.500 0.560 1.100 0.650 0.270 0.270 0.500 0.770 0.730 0.340 0.170 0.160 0.270
```

Решење:

Реализоване вредности дескриптивних статистика:

$\bar{x}_n$	m_e	$\tilde{s}_n$	min	max	q_1	q_3
0.5250	0.4900	0.3486	0.0400	1.3300	0.2500	0.7700

Оба графика (десно) на којима су приказани подаци из узорка − дијаграм „грана-лишће“ и Q-Q график указују да обележје од интереса нема нормалну расподелу и да је расподела обележја позитивно асиметрична.

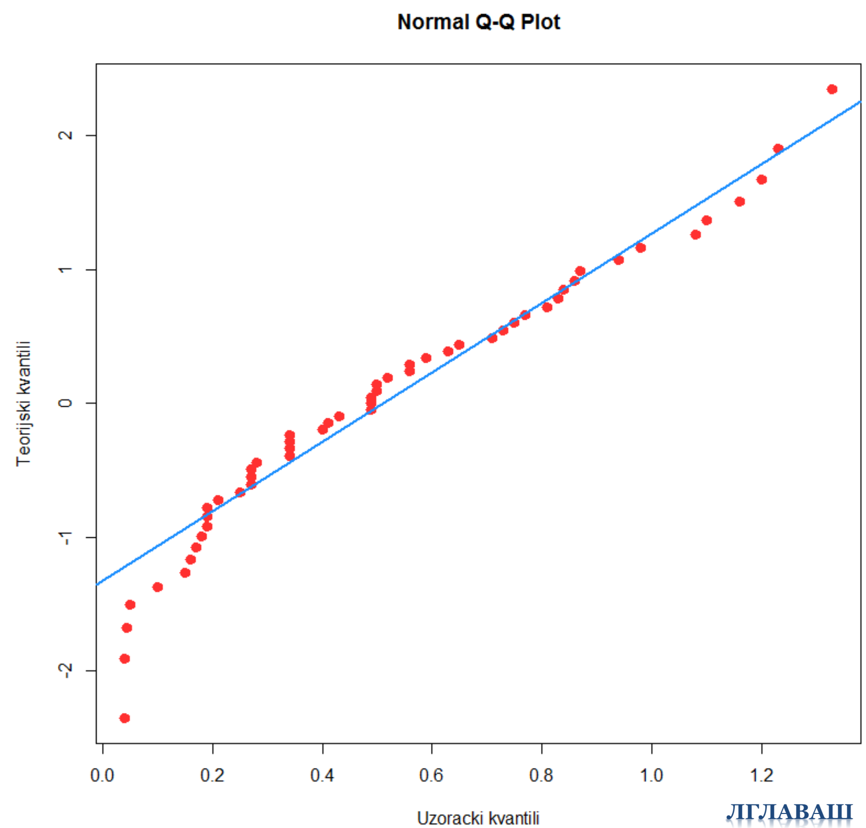
Напомена: Q-Q график је вероватносни график, који служи као графички метод за поређење двеју расподела вероватноћа (у смислу њиховог слагања). Овде се врши поређење узорачке расподеле са нормалном расподелом.
x -координате тачака на графику су реализоване вредности статистика поретка $x_{(j)}$, а y-координате су квантили стандардне нормалне расподеле z_j за које важи:

$$\Phi(z_j) = \frac{j - 0.5}{n},$$

где је n обим узорка.
Ако претпостављена расподела адекватно описује податке из узорка нацртане тачке су концентрисане приближно око праве, а ако оне значајно одступају од праве претпостављена расподела није одговарајућа.

```
The decimal point is 1 digit(s) to the left of the |

 0 | 4445
 1 | 05678999
 2 | 157778
 3 | 4444
 4 | 013999
 5 | 002669
 6 | 35
 7 | 1357
 8 | 13467
 9 | 48
10 | 8
11 | 06
12 | 03
13 | 3
```



Решење: (наставак)

С обзиром да је обим датог узорка велики ($n > 40$), не претпостављајући нормалну расподељеност обележја, ипак се може користити статистика

$$\frac{\bar{X}_n - m}{\tilde{S}_n} \cdot \sqrt{n}$$

за формирање апроксимативног интервала поверења за мат. очекивање m са задатим нивоом поверења (на основу важења ЦГТ). Ова статистика има приближно $N(0,1)$ расподелу.

Апроксимативни двострани 95% интервал поверења за m је:

$$[0.4311, 0.6189]. \triangle$$

Пример 2

У извесном истраживању, из области технологије материјала, испитивана је затезна чврстоћа одређене легуре.

Испитивање је спроведено на укупно 22 узорка легуре и добијене су следеће вредности затезне чврстоће (у MPa) у тренутку разарања узорка:

```
[1] 19.8 10.1 14.9  7.5 15.4 15.4 15.4 18.5  7.9 12.7 11.9 11.4 11.4 14.1 17.6 16.7 15.8 19.5  8.8 13.6 11.9 11.4
```

Одредити 95% интервал поверења за мат. очекивање m овог обележја.

Решење:

Реализоване вредности дескриптивних статистика:

$\bar{x}_n$	m_e	$\tilde{s}_n$	min	max	q_1	q_3
13.71	13.85	3.55	7.50	19.80	11.40	15.70

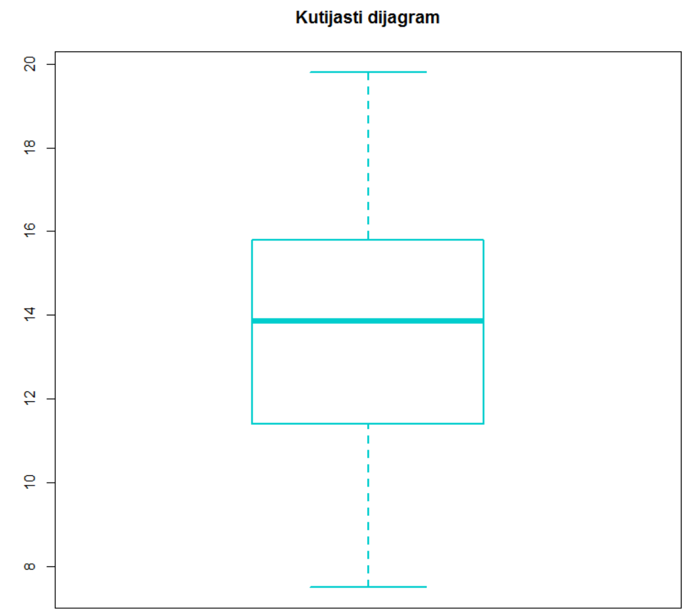
Кутијасти дијаграм указује на то да нема изражене асиметрије нити аутлајера.
Заправо ни овај график нити Q-Q график не даје основа за веровање да би претпоставка о нормалној расподели обележја од интереса била погрешна.

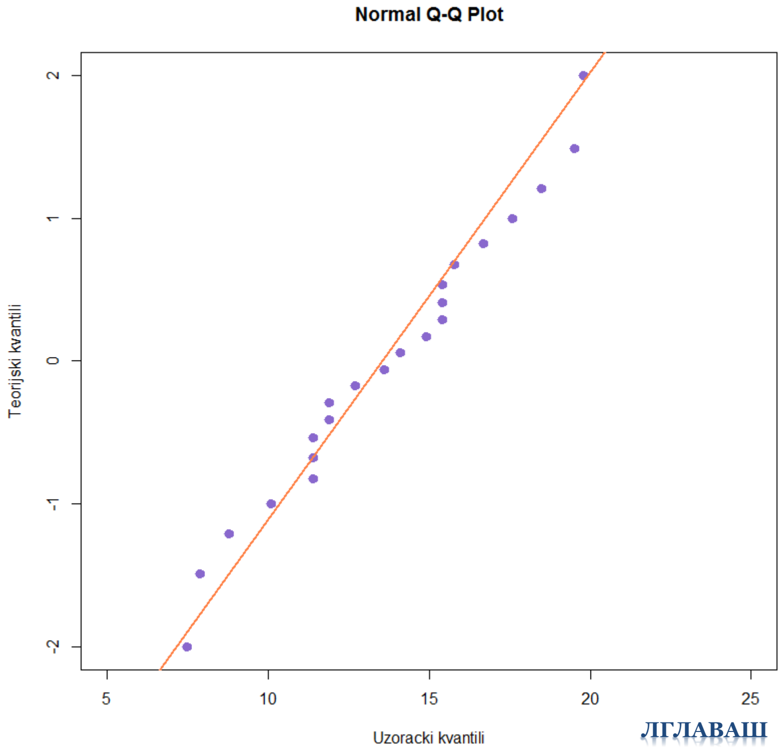
Ако се, дакле, претпостави да је расподела обележја заиста нормална, са непознатим параметрима m и σ^2, користи се статистика

$$\frac{\bar{X}_n - m}{\tilde{S}_n} \cdot \sqrt{n}$$

за формирање интервала поверења за мат. очекивање m са задатим нивоом поверења. Ова статистика има Студентову t_{n-1} расподелу.

Тражени двострани 95% интервал поверења за m је:
$[12.14, 15.28]$. △





ВАЖНО!

При одређивању интервала поверења за мат. очекивање m нормалне расподеле, на основу узорка малог обима n, замена непознате вредности σ^2 (а овај параметар је, у применама, најчешће непознат) њеном непристрасном оценом $\tilde{S}_n^2$ (формираном на основу узорка) има великог утицаја на закључак о m. Наиме, тада постоје значајне вероватносне разлике између статистике

$$\frac{\bar{X}_n - m}{\tilde{S}_n} \cdot \sqrt{n}$$

и статистике $\frac{\bar{X}_n - m}{\sigma} \cdot \sqrt{n}$, на којој се базира закључивање о m када је σ^2 познато. Зато се у тим ситуацијама увек користи горња статистика која има Студентову (а не стандардну нормалну) расподелу.

Шта се дешава код „великих“ узорака??

Пример 3

У фабрици се користи аутоматска машина за пуњење боца течним детерџентом. На простом случајном узорку обима 20 боца израчуната је поправљена узорачка дисперзија запремине пуњења: $\tilde{s}_n^2 = 13.3824\,(ml)^2$.

Ако је дисперзија запремине пуњења превелика, удео боца које су препуњене односно које су недовољно напуњене биће неприхватљив за произвођача. При томе, претпоставља се да је запремина пуњења боце нормално расподељена.

Одредити једнострани горњи 95% интервал поверења за непознати параметар σ^2 расподеле посматраног обележја.

Решење:

Тражени једнострани горњи 95% интервал поверења за непознату дисперзију σ^2 је:

$$(0, 25.1325].$$

Једнострани горњи 95% интервал поверења за непознато стандардно одступање σ је:

$$(0\,ml, 5.0132\,ml]. \triangle$$

Пример 4

Применом стандардизоване процедуре у процесу производње пешкира очекује се постизање врло малог одступања у њиховој дебљини.

Под претпоставком да је дебљина пешкира нормално расподељена случајна величина са непознатим мат. очекивањем m и дисперзијом σ^2 одредити најмањи обим узорка n пешкира тако да очекивана дужина двостраног 95% интервала поверења за σ^2 не буде већа од σ^2.

Решење:

Дужина двостраног интервала поверења за σ^2 са нивоом поверења β је случајна величина $(n-1)\tilde{S}_n^2\left(\frac{1}{\chi^2_{n-1;\frac{1+\beta}{2}}}-\frac{1}{\chi^2_{n-1;\frac{1-\beta}{2}}}\right)$, а њено мат. очекивање једнако је $(n-1)\sigma^2\left(\frac{1}{\chi^2_{n-1;\frac{1+\beta}{2}}}-\frac{1}{\chi^2_{n-1;\frac{1-\beta}{2}}}\right)$.

Оно овде мора бити $\leq \sigma^2$, па n мора бити одабрано тако да је $(n-1)\left(\frac{1}{\chi^2_{n-1;0.975}}-\frac{1}{\chi^2_{n-1;0.025}}\right)\leq 1$.

Пример 4 (наставак)

Коришћењем таблица за χ^2 расподелу или неког статистичког пакета може се добити таблица следећег облика:

n	$\chi^2_{n-1;0.025}$	$\chi^2_{n-1;0.975}$	$(n-1)\left(\frac{1}{\chi^2_{n-1;0.975}} - \frac{1}{\chi^2_{n-1;0.025}}\right)$
15	26.11895	5.62873	1.9512
20	32.85233	8.90652	1.5549
30	45.72229	16.04707	1.1729
…			
38	55.66797	22.10563	1.0091
39	56.89552	22.87848	0.9931

На основу последње врсте горње таблице може се закључити да ће узорак обима (најмање) $n = 39$ пешкира резултирати двостраним 95% интервалом поверења за σ^2 просечне дужине мање од σ^2. △