

# Štatistické spracovanie experimentálnych dát

Štatistická analýza veľkých výberov  
Štatistická analýza malých výberov podľa Horna  
Štatistické testovanie  
Analýza rozptylu

Dátum: 12. máj 2008

Vypracoval: Ing. Ľudmila Sýkorová

<b>Štatistická analýza veľkých výberov.....</b>	<b>4</b>
Zadanie .....	4
Dáta .....	4
<b>Exploratívna analýza dát – EDA (QC Expert).....</b>	<b>4</b>
Riešenie EDA .....	4
Záver EDA.....	9
<b>Základné predpoklady o výbere (QC Expert).....</b>	<b>10</b>
Riešenie .....	10
Záver základných predpokladov výberu.....	10
<b>Transformácia dát (QC Expert) .....</b>	<b>11</b>
Exponenciálna transformácia.....	11
BOX-COXOVA transformácia .....	13
Záver transformácie dát.....	14
<b>Odhady parametrov (QC Expert) .....</b>	<b>15</b>
Záver.....	15
<b>Štatistická analýza malých výberov podľa Horna .....</b>	<b>16</b>
Zadanie .....	16
Dáta .....	16
Riešenie .....	16
Záver analýzy podľa Hornovho postupu .....	17
<b>Štatistické testovanie.....</b>	<b>18</b>
<b>Test správnosti (QC Expert) .....</b>	<b>18</b>
Zadanie .....	18
Dáta .....	18
Riešenie .....	18
Záver.....	20
<b>Test zhodnosti (QC Expert) .....</b>	<b>21</b>
Zadanie .....	21
Dáta .....	21
Riešenie .....	21
Záver.....	24
<b>Párový test (QC Expert).....</b>	<b>25</b>
Zadanie .....	25
Dáta .....	25
Riešenie .....	25
Záver.....	25
<b>Analýza rozptylu – ANOVA .....</b>	<b>26</b>
<b>Jednofaktorová ANOVA (QC Expert).....</b>	<b>26</b>
Zadanie .....	26
Dáta .....	26
Riešenie .....	26
Záver.....	29
<b>Dvojfaktorová ANOVA bez opakovania (ADSTAT).....</b>	<b>30</b>
Zadanie .....	30
Dáta .....	30
Riešenie .....	30
Záver.....	31

<b>Dvojfaktorová ANOVA s opakovaním (ADSTAT).....</b>	<b>32</b>
Zadanie .....	32
Dáta .....	32
Riešenie .....	32
Záver.....	34

# Štatistická analýza veľkých výberov

## Zadanie

Množina dát obsahuje výsledky vyšetrenia krvi Trombocyty označované PLT. Dátový súbor obsahuje 79 výsledkov stanovených pre skupinu mužov. Posúďte symetriu, zošikmenie dát a stanovte odhad parametrov. V prípade potreby vykonajte transformáciu.

## Dáta

Hodnoty pre premennú PLT:

97	183	57	57	238	97	330	122	180	181	142	248	60
247	49	124	837	98	155	79	170	207	139	217	119	238
212	401	178	172	290	472	171	281	190	350	368	61	59
48	272	146	207	200	88	107	284	464	184	295	182	391
149	9	294	88	211	758	276	198	185	212	324	131	183
153	323	168	191	125	268	177	414	188	184	226	189	190
107												

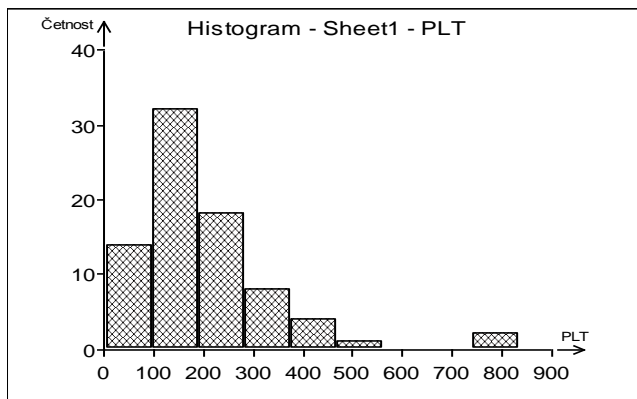
## Štatistická analýza:

- Exploratívna analýza dát
- Základné predpoklady o výbere
- Transformácia dát
- Odhady parametrov

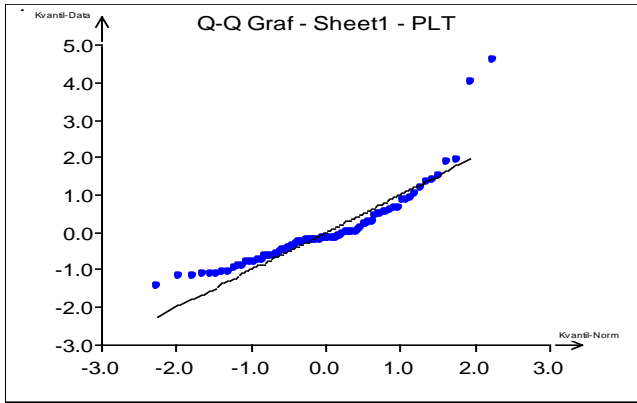
## Exploratívna analýza dát – EDA (QC Expert)

Prieskumová analýza dát slúži na prešetrenie štatistických zvláštností dát, ako je lokálna koncentrácia dát, tvarové zvláštnosti ich rozdelenia a prítomnosť podozrivých hodnôt. Odhaľuje anomálie a odchýlky rozdelenia výberu od normálneho rozdelenia (Gaussovho). Zistené skutočnosti určujú ďalší postup spracovania dát. Interpretácia výsledkov zohľadňuje charakter dát, vizuálne posúdenie grafického zobrazenia výsledkov a zhodnotenie vypočítaných štatistických charakteristík.

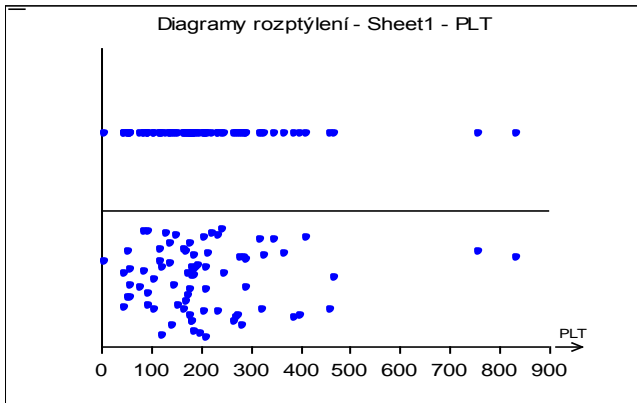
## Riešenie EDA



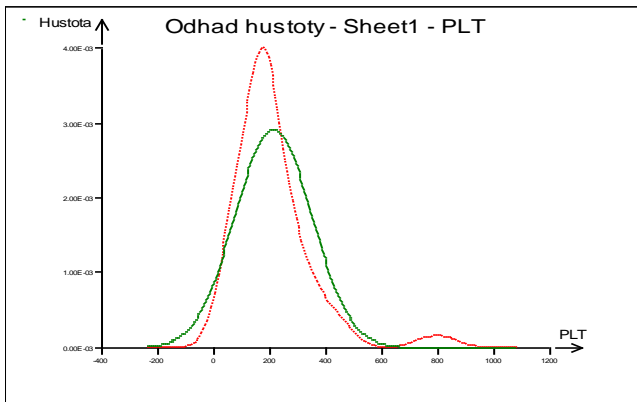
Obr. 1 Histogram PLT



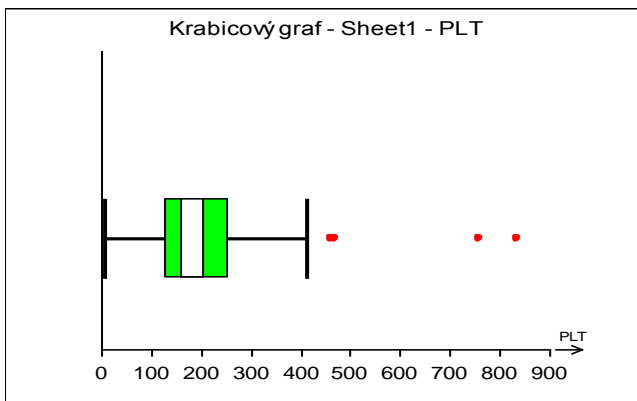
Obr. 2 Q-Q graf PLT



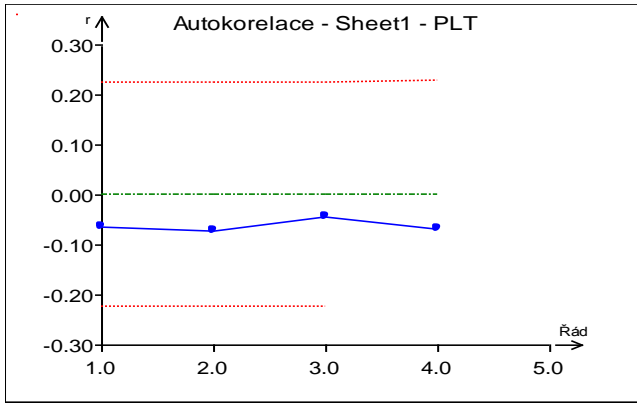
Obr. 3 Diagram rozptýlenia PLT



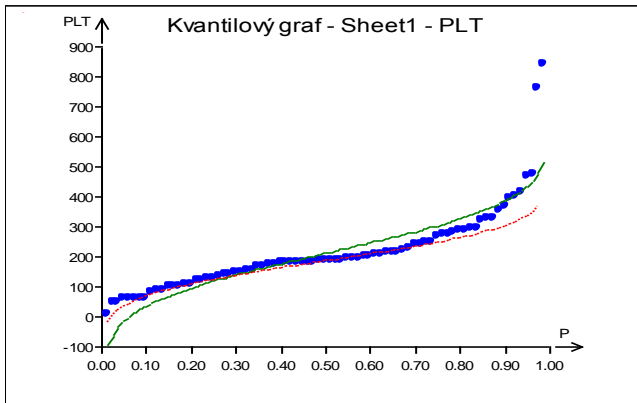
Obr. 4 Odhad hustoty PLT



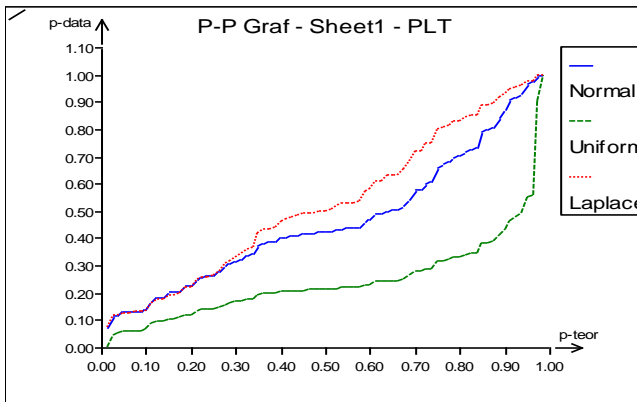
Obr. 5 Krabicový graf PLT



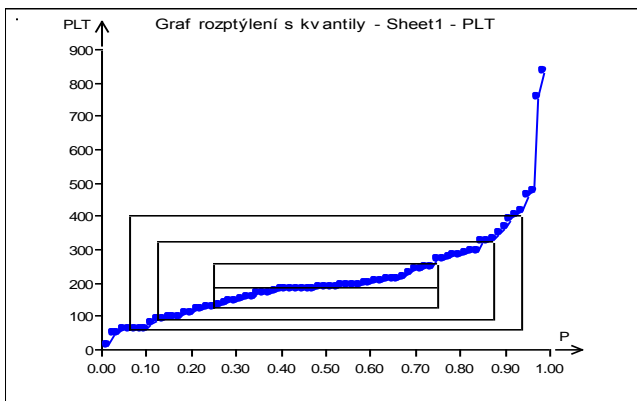
Obr. 6 Autokorelácia PLT



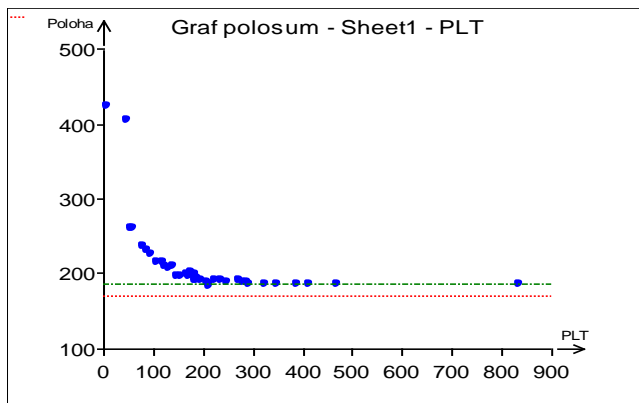
Obr. 7 Kvantilový graf PLT



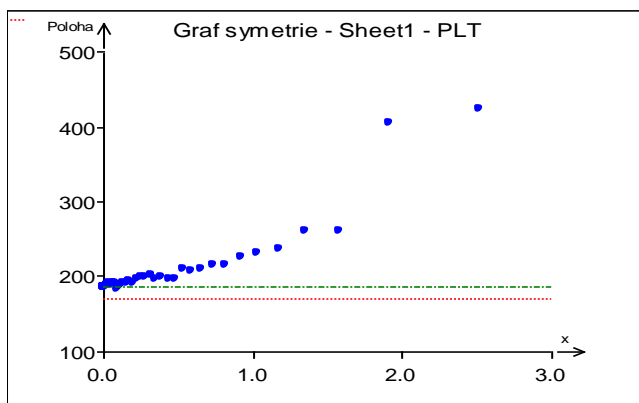
Obr. 8 P-P graf PLT



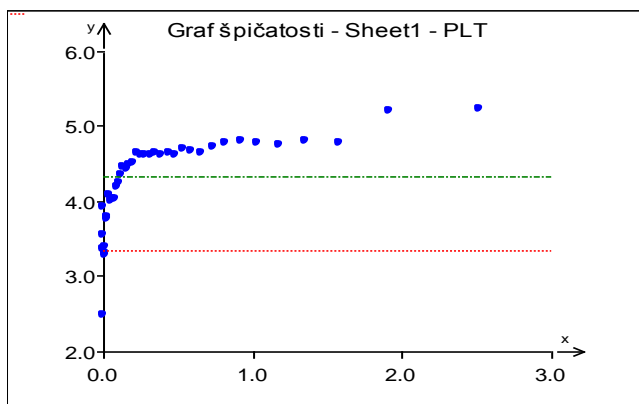
Obr. 9 Graf rozptýlenia s kvantilami PLT



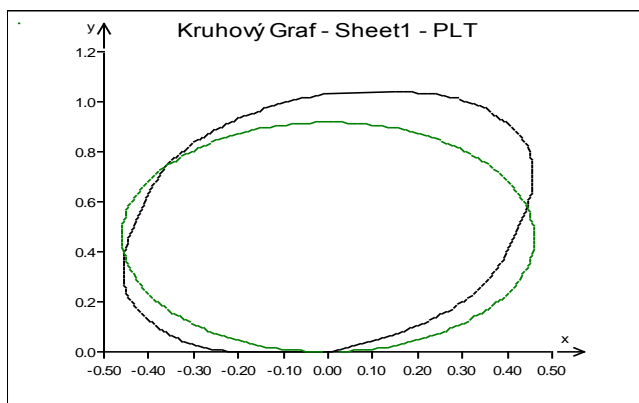
Obr. 10 Graf polosum PLT



Obr. 11 Graf symetrie PLT



Obr. 12 Graf špičatosti PLT



Obr. 13 Kruhový graf PLT

### **Obr. 1 Histogram PLT**

Graf zobrazuje početnosti (os y: úmerná hustote pravdepodobnosti) pre jednotlivé triedy premennej PLT (os x: premenná PLT). Triedy majú konštantnú šírku. Optimálny počet tried je stanovený automaticky, s ohľadom na počet pozorovaní vo výberovom súbore. Rozdelenie nie je symetrické, vykazuje zošikmenie vpravo s dvoma bodmi s výrazne odľahlou hodnotou.

### **Obr. 2 Q-Q graf PLT**

Graf zobrazuje odchýlku od teoretického normálneho rozdelenia v oblasti koncov. Posudzuje zhodu výberového rozdelenia  $Q_E(P_i)$  s kvantilovou funkciou teoretického rozdelenia  $Q_T(P_i)$ . Body neležia na priamke. Maximálne odchýlky vykazujú 2 body v oblasti horných kvantilov.

Vztýčením kolmice na os X v bode 0 získame priesečník s priamkou (aritmetický priemer) a priesečník s bodmi (medián). Priesečníky nie sú totožné, čo svedčí o miernej asymetrii.

### **Obr. 3 Diagram rozptýlenia PLT**

Na os x sú vynesené hodnoty x a os y reprezentuje ľubovoľnú úroveň. Graf reprezentuje jednorozmernú projekciu kvantilového grafu do osy x. Aby bolo možné lepšie posúdiť rozloženie dát sú v spodnom grafe zobrazené body rozptýlené v smere osy y. Výsledkom tohto spôsobu zobrazenia je, že nedochádza k splývaniu zhodných dát alebo dát s blízkou hodnotou. V hornej časti sú identifikované 2 odľahlé body. Body nevytvárajú separátne zhluky.

### **Obr. 4 Odhad hustoty PLT**

Graf slúži pre porovnanie priebehu hustoty pravdepodobnosti normálneho rozdelenia (zelená čiara) a jadrového odhadu hustoty, ktorý je vypočítaný z dát - červená čiara. Empirická a Gaussova krivka sa odlišujú, to značí, že nie sú totožné.

Vrchol Gaussovej krivky znázorňuje aritmetický priemer. Vrchol empirickej krivky udáva medián. Hodnoty nie sú totožné, čo potvrdzuje miernu asymetriu so zvýraznením odľahlých bodov. Empirická krivka má väčšiu špicatosť ako Gaussova. Os x reprezentuje hodnoty PLT a os y hustotu pravdepodobnosti  $f(x)$ . Nehomogenitu dát spôsobenú zhlukmi vyjadrujú maximá na tejto krivke. Hladkosť krivky je daná parametrom vyhladenia hustoty, keď sa pri jeho nízkej hodnote objavia maximá pre akékoľvek dáta. Pre PLT sú odľahlé body znázornené ako zhluk dát.

### **Obr. 5 Krabicový graf PLT**

Graf sumarizuje dáta a znázorňuje robustný odhad polohy (medián M), posúdenie symetrie a identifikáciu odľahlých bodov. Rozdelenie hodnôt je mierne zošikmené vpravo. Asymetriu vytvárajú 4 odľahlé body identifikované vpravo.

### **Obr. 6 Autokorelácia PLT**

Koeficienty autokorelácie 1. až 4. rádu dosahujú hodnoty blízke nule. Hodnoty v spracovávanom súbore nie sú na sebe závislé.

### **Obr. 7 Kvantilový graf PLT**

Kvantilový graf zvýrazňuje 2 odľahlé body a miernu asymetriu rozdelenia.

### **Obr. 8 P-P graf PLT**

Pravdepodobnostný graf porovnáva distribučnú funkciu výberu vyjadrenú cez poradovú pravdepodobnosť so štandardizovanou distribučnou funkciou zvoleného teoretického rozdelenia. Lineárnemu grafu (priamke) s jednotkovou smernicou a nulovým úsekom je najbližšie graf Laplaceovho rozdelenia.

### **Obr. 9 Graf rozptýlenia s kvantilami PLT**

Miernu asymetriu s dvoma výraznými odľahlými bodmi.

### **Obr. 10 Graf polosum PLT**



Graf je citlivým indikátorom asymetrie rozdelenia. Prostredná horizontálna krivka, na ktorej leží posledný bod je medián a červené prerušované priamky reprezentujú medze jeho intervalu spoľahlivosti. Os x obsahuje poriadkové štatistiky  $x_i$ , os y hodnotu  $Z_i = 0.5(x_{(n+1-i)} + x_{(i)})$ . Prvé dva body na grafe vľavo sa výrazne odchyľujú od horizontálnej priamky. V zobrazení sa objavuje nenáhodný trend.

**Obr. 11 Graf symetrie PLT**

Graf symetrie má podobný význam ako graf polosúm. Os x obsahuje hodnoty  $M - x_{(i)}$  a os y hodnoty  $x_{(n+1-i)} - M$ . Symetrické rozdelenie je charakteristické priamkou  $y = M$ . Smernica priamky vytvorenej z bodov v grafe je nenulová, čo je prejavom asymetrického rozdelenia.

**Obr. 12 Graf špicatosti PLT**

Smernica nenáhodného trendu, ktorý je tvorený bodmi zodpovedá parametru špicatosti.

**Obr. 13 Kruhový graf PLT**

Kruhový graf slúži k overeniu, že výber pochádza zo symetrického (najčastejšie Gaussovho) rozdelenia. Odchýlky od kružnice poukazujú na iné ako symetrické rozdelenie výberu. Elipsovité tvar zobrazený na obrázku charakterizuje asymetrické rozdelenie.

## **Záver EDA**

Dáta získané meraním trombocytov PLT sú asymetrické. Obsahujú dve odľahlé hodnoty. Vzhľadom na charakter dát, je vylúčenie týchto dvoch hodnôt nemožné. Dôsledkom by bola strata informácií. Podľa PP grafu dáta preukazujú podobnosť s Laplaceovým rozdelením.

## Základné predpoklady o výbere (QC Expert)

### Riešenie

#### Klasické parametre :

Priemer :	210,9493671
Spodná hranica:	180,3040262
Horná hranica :	241,594708
Rozptyl :	18718,89484
Smer. odchýlka :	136,8170122
Šikmosť	2,145876321
Odchýlka od 0 :	<b>Významná</b>
Špicatosť :	9,833475514
Odchýlka od 3 :	<b>Významná</b>

#### Znamienkový test :

Záver : Dáta sú **nezávislé**

#### Rad autokorelácie

**1**

Korelačný koeficient : -0,06397564786

Pravdepodobnosť : 0,288946198

Záver : **Nevýznamný**

#### Test normality :

Názov stĺpca : PLT

Priemer : 210,9493671

Rozptyl : 18718,89484

Šikmosť 2,145876321

Špicatosť : 9,833475514

Normalita : **Zamietnutá**

Vypočítaný : 17,7995736

Teoretický : 5,991464547

Pravdepodobnosť : 0,000136418008

#### Vybočujúce body :

Homogenita : **Zamietnutá**

Počet vybočujúcich bodov : 2

### Záver základných predpokladov výberu

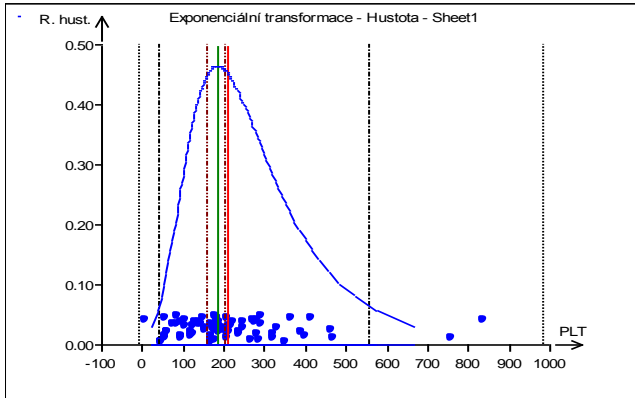
Test nezávislosti dát je prijatý na hladine významnosti 0.05.

Test normality s hladinou významnosti 0.05 predpoklad normality zamietol, čo potvrdilo výsledky z exploratívnej analýzy dát. Hodnota špicatosti sa výrazne líši od hodnoty 3 a hodnota šikmosti sa výrazne líši od hodnoty 0. Odhad parametrov vyžaduje transformáciu dát.

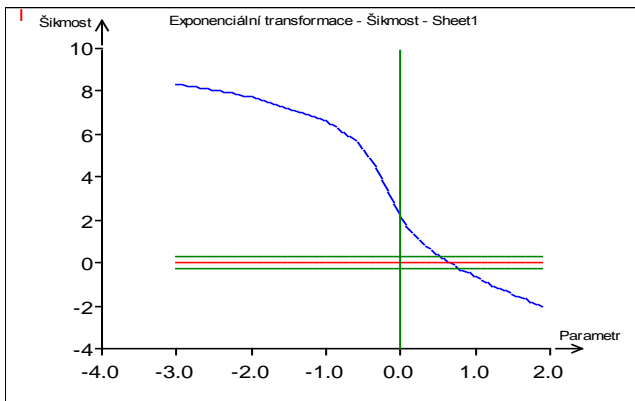
Súbor obsahuje dva vybočujúce body s maximálnymi hodnotami.

## Transformácia dát (QC Expert)

### Exponenciálna transformácia

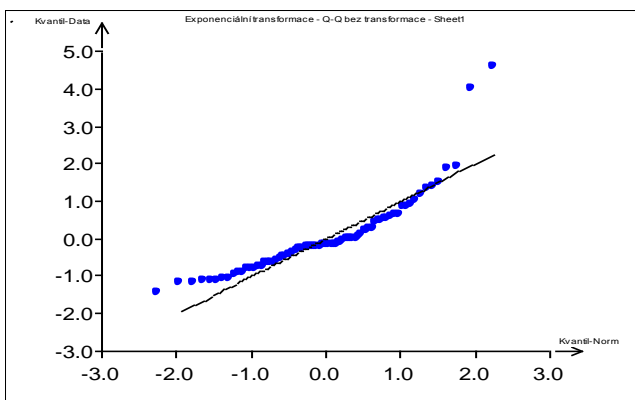


Obr. 14 Exponenciálna transformácia – Hustota PLT

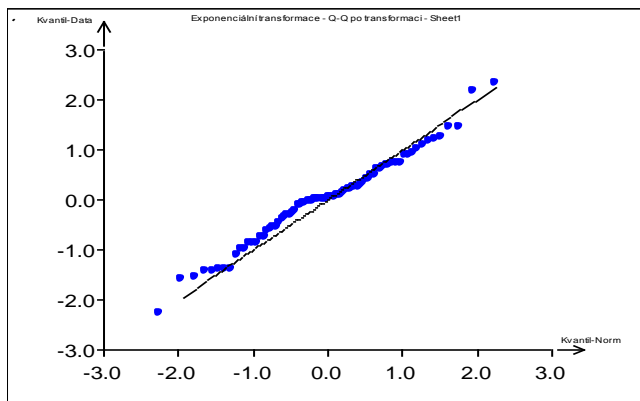


Obr. 15 Exponenciálna transformácia - Šikmost'

Šikmost' transformovaných dát je závislá na parametri exponenciálnej transformácie. Normálne rozdelenie požaduje nulovú šikmost', ktorá zodpovedá optimálnemu parametru. Graf na Obr. 15 slúži k stanoveniu parametra transformácie a k určeniu štatistickej významnosti transformácie. Ak priesečník zvislej zelenej priamky s krivkou leží mimo interval spoľahlivosti šikmosti (vodorovná zelená priamka), je transformácia opodstatnená.



Obr. 16 Exponenciálna transformácia – Q-Q graf pôvodných dát



**Obr. 17 Exponenciálna transformácia – Q-Q graf po transformácii**

Q-Q graf pôvodných dát (Obr. 16 zhodný s Q-Q grafom z EDA) nie je zhodný s Q-Q grafom po exponenciálnej transformácii (Obr. 17). Po vykonaní transformácie sú body bližšie k priamke ako v grafe pôvodných dát. Transformácia je úspešná. Oproti štatistikám Q-Q graf umožňuje vizuálne posúdenie, či je nelinearita (odchýlka od normality) spôsobená len niekoľkými bodmi alebo všetkými dátami.

Optimálny parameter : 0,6608848572

Zvolený parameter : 0,6608848572

Oprávenosť transformácie : **Áno**

Opravený priemer : **182,5472443**

Interval spoľahlivosti :

Spodný : 160,7909524

Horný : 206,8613934

LCL : -5,830292453

UCL : 984,5670493

LWL : 40,14554429

UWL : 559,9046969

Významné  
opravené  
kvantily p

spodný

horný

50 %

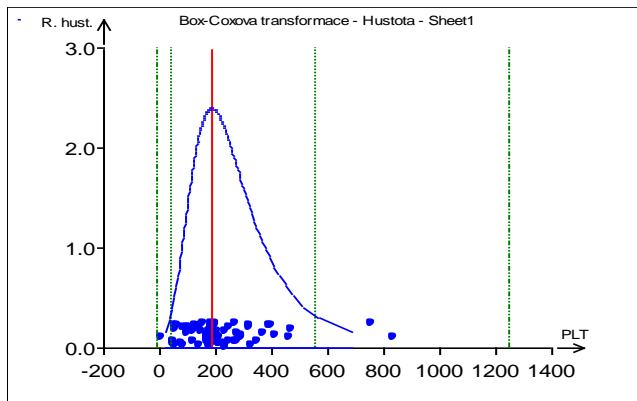
182,5472443

25 %

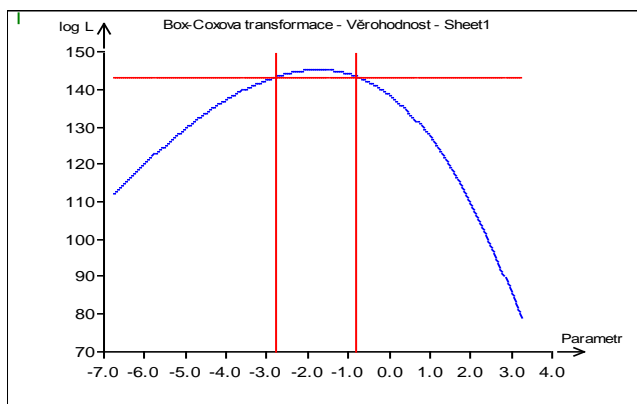
122,9358115

265,9762882

## BOX-COXOVA transformácia

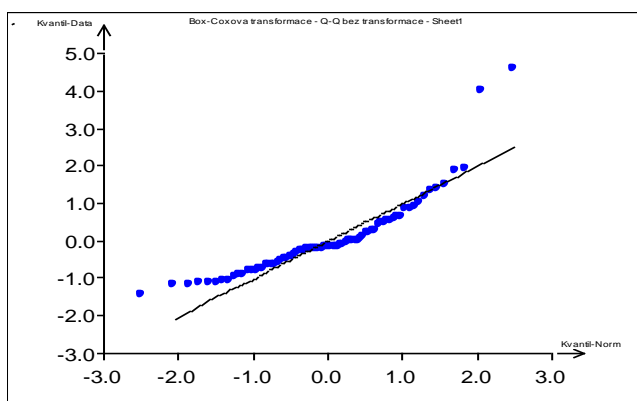


Obr. 18 Box Coxova transformácia - Hustota

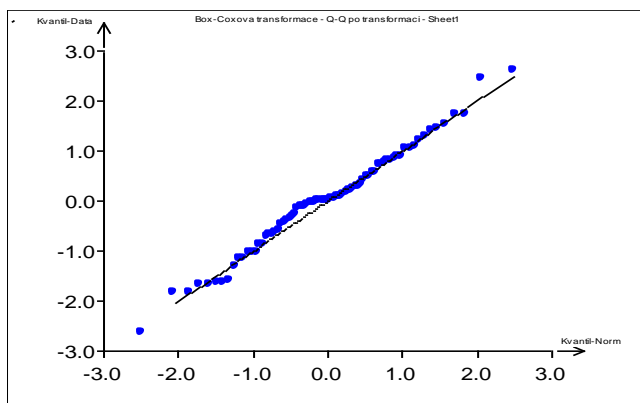


Obr. 19 Box Coxova transformácia - Věrohodnost'

Graf závislosti logaritmu vierohodnostnej funkcie na parametri je zobrazený v grafe na Obr. 19. Optimálna hodnota parametra zodpovedá maximu. Vodorovná priamka zodpovedá spodnej medzi 95%-ného intervalu spoľahlivosti maxima vierohodnosti a zvislé priamky intervalu reprezentujú spoľahlivosť odhadu. Ak tento interval obsahuje hodnotu 1, transformácia nie je potrebná. Ak interval túto hodnotu neobsahuje, transformácia je oprávnená.



Obr. 20 Box Coxova transformácia – Q – Q graf pôvodných dát



**Obr. 21** Box Coxova transformácia – Q – Q graf po transformácii

Optimálny parameter :	-1,751796341
Dolná medza parametra :	-2,781947117
Horná medza parametra :	-0,8221481329
Vierohodnosť bez transformácie :	127,1455951
Vierohodnosť s transformáciou :	145,1789231
Oprávenosť transformácie:	<b>Áno</b>
Pravdepodobnosť :	99,9999998090848 %
Zvolený parameter :	<b>-1,751796341</b>
Vierohodnosť :	145,1789231
Opravený priemer :	<b>183,7236172</b>
LCL :	-5,781498809
UCL :	1249,83284
LWL :	99,48707274
UWL :	314,0920127

Významné  
opravené

kvantily p	spodný	horný
50 %	<b>183,7236172</b>	
25 %	123,546849	264,2176885

## Záver transformácie dát

Na základe analýzy EDA a vykonaní testu normality sa potvrdilo, že rozdelenie výberu dát sa štatisticky významne odlišuje od normálneho rozdelenia.

Štatistickým testom sa potvrdila opodstatnenosť exponencionálnej transformácie. Exponencionálna transformácia s optimálnym parametrom 0.66 dosahuje opravený priemer 182,54. Interval spoľahlivosti je v rozpätí 160,79 až 206,86.

Oprávenosť Box-Coxovej transformácie bola preukázaná na základe grafu vierohodnosti Obr. 19. Optimálny parameter nadobúda hodnotu -1,75. Interval medzi dolnou a hornou medzou neobsahuje hodnotu 1, čo potvrdzuje vhodnosť použitia Box-Coxovej transformácie. Opravený priemer dosahuje hodnotu 183,72.

## Odhady parametrov (QC Expert)

### Klasické parametre :

Priemer :	210,9493671
Spodná hranica:	180,3040262
Horná hranica :	241,594708
Rozptyl :	18718,89484
Smer. odchýlka :	136,8170122
Šikmosť	2,145876321
Špicatosť :	9,833475514
Polosuma	423
Modus :	131,4487342

### Robustné parametre :

Medián :	184
IS spodný :	166,2242532
IS horný :	201,7757468
Mediánová smer. odchýlka :	8,928735432
Mediánový rozptyl :	79,72231641
10% Priemer :	196,7746479
10% IS spodný :	171,9362311
10% IS horný :	221,6130647
10% Smer. odchýlka :	80,56294516
10% Rozptyl :	6490,388132
20% Priemer :	193,015873
20% IS spodný :	168,8188688
20% IS horný :	217,2128772
20% Smer. odchýlka :	59,25431586
20% Rozptyl :	3511,073949
40% Priemer :	189,5531915
40% IS spodný :	166,1973983
40% IS horný :	212,9089847
40% Smer. odchýlka :	32,00264954
40% Rozptyl :	1024,169578

### Exponenciálna transformácia

Opravený priemer :	<b>182,5472443</b>
Interval spoľahlivosti spodný :	160,7909524
Interval spoľahlivosti horný :	206,8613934

### Záver

Použitie klasických parametrov pri asymetrickom rozdelení nie je možné. Všetky podmienky pre transformáciu dát sú splnené. Výsledok z exponenciálnej transformácie – retransformovaný priemer je použitý ako odhad strednej hodnoty pre trombocyty PLT v skupine mužov. Priemerná hodnota je 182,547 a s 95%-nou pravdepodobnosťou sa stredná hodnota nachádza v intervale 160,79 až 206,86. Vypočítaná hodnota nie je veľmi odlišná od retransformovaného priemeru získaného spätnou transformáciou priemernej hodnoty dát Box-Coxovej transformácie (183,72).

# Štatistická analýza malých výberov podľa Horna

## Zadanie

U jedného pacienta bol stanovený hemoglobín 8 krát. Hodnota je stanovená v jednotkách g/l. Úlohou je určiť priemernú hodnotu hemoglobínu v krvi pre pacienta. Interval spoľahlivosti je stanovený na 95%.

## Dáta

Hemoglobín [g/l]:

137 138 133 132 137 134 141 132

## Riešenie

Analýza malých výberov je vždy nepresná a zaťažená vysokou mierou štatistickej neistoty. Hornov postup je založený na poriadkových štatistikách. Vychádza z hĺbky pivotov zodpovedajúcich približne výberovým kvartilom (písmeno F).

### 1. Poriadkové štatistiky

<b>i</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
<b><math>x_i</math></b>	132	132	133	134	137	137	138	141

### 2. Hĺbka pivotu

$$H = \text{int} \frac{\frac{n+1}{2} + 1}{2} = 2$$

### 3. Dolný a horný pivot

$$x_d = x_{(H)} = x_{(2)} = 132$$

$$x_h = x_{(n+1-H)} = x_{(7)} = 138$$

### 4. Pivotova polosuma (miera polohy)

$$P_L = \frac{x_d + x_h}{2} = 135$$

### 5. Pivotové rozpätie (miera rozptýlenia)

$$R_L = x_h - x_d = 6$$

### 6. Kvantil $t_{L,0.975(8)} = 0.564$ je určený z tabuľky

### 7. Pre 95%-ný interval spoľahlivosti platí:

$$P_L - R_L * t_{L,0.975(8)} \leq \mu \leq P_L + R_L * t_{L,0.975(8)}$$

$$131.616 \leq \mu \leq 138.384$$



## **Záver analýzy podľa Hornovho postupu**

S 95%-nou štatistickou istotou môžeme tvrdiť, že odhad strednej hodnoty ( $\mu$ ) hemoglobínu pacienta sa nachádza v intervale  $L_d = 131.616$  a  $L_h = 138.384$ .

# Štatistické testovanie

## Test správnosti (QC Expert)

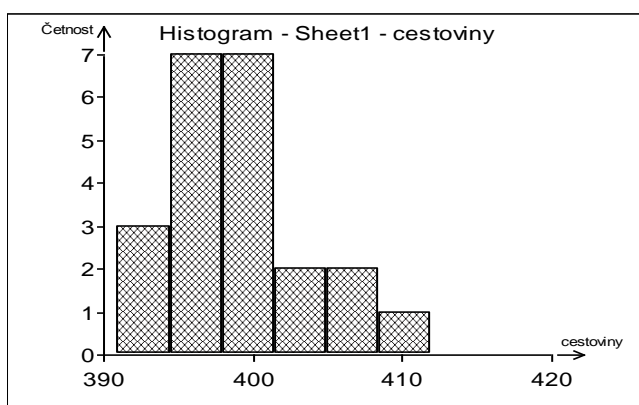
### Zadanie

Výrobca cestovín Trabel Gran Canaria uvádza hmotnosť výrobku 400g. Kontrola bola vykonaná na 20 baleniach vážením hmotnosti na digitálnej váhe. Vyhovujú namerané hodnoty hmotnosti deklarovanej výrobcom?

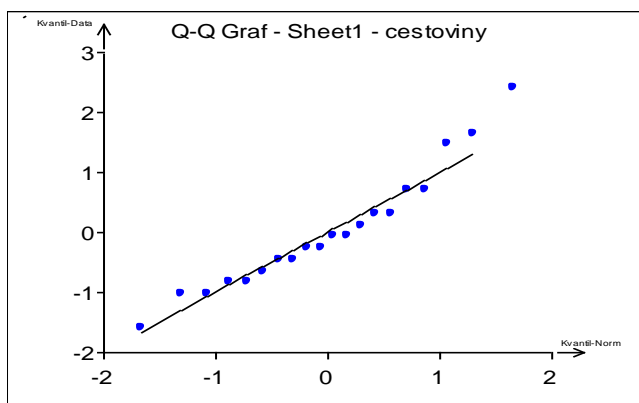
### Dáta

399	397	407	394	397	395	394	401	408	393
396	398	403	391	412	403	399	400	401	395

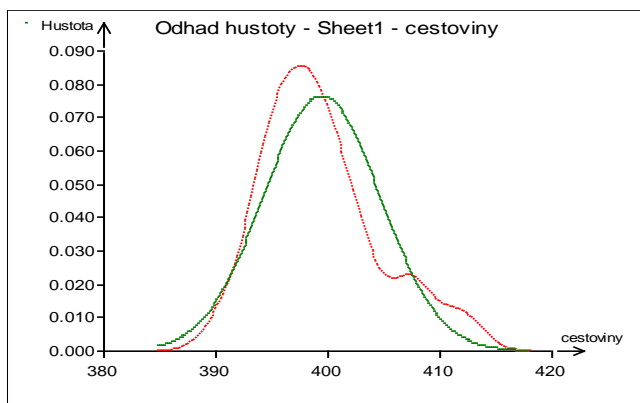
### Riešenie



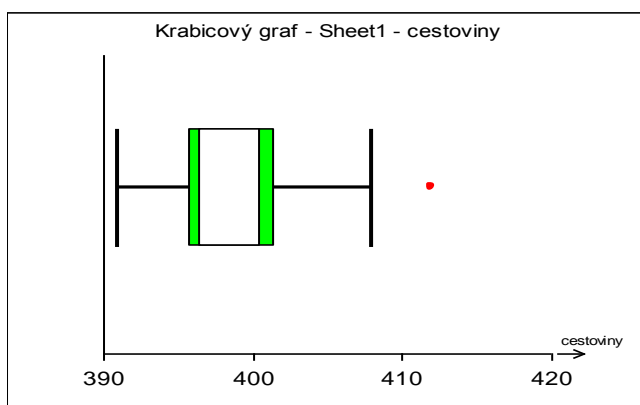
Obr. 22 Histogram



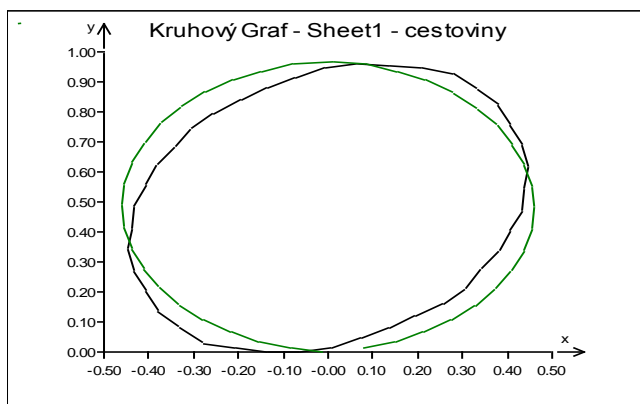
Obr. 23 Q-Q graf



Obr. 24 Odhad hustoty



Obr. 25 Krabicový graf



Obr. 26 Kruhový graf

**Klasické parametre :**

Názov stĺpca :	cestoviny
Priemer :	399,4
Spodná medza :	396,959
Horná medza :	401,84
Rozptyl :	27,2
Smer. odchýlka :	5,215
Šikmosť :	0,7714
Odchýlka od 0 :	Nevýznamná
Špicatosť :	3,1449
Odchýlka od 3 :	Nevýznamná

**t-test**

Testovaná hodnota :	400
Rozdiel :	<b>Nevýznamný</b>
Vypočítaný :	-0,5145
Teoretický :	2,093
Pravdepodobnosť :	0,3064191904
Konfidenčný interval ľavý:	397,383502
Konfidenčný interval pravý:	401,416498

**Robustné parametre :**

Medián :	398,5
IS spodný :	387,287
IS horný :	409,713
Mediánová smer. odchýlka :	5,357
Mediánový rozptyl :	28,70

**Znamienkový test :**

Záver :	<b>Dáta sú nezávislé</b>
---------	--------------------------

**Test normality :**

Priemer :	399,4
Rozptyl :	27,2
Šikmosť	0,771
Špicatosť :	3,144
Normalita :	<b>Prijatá</b>
Vypočítaný :	2,883
Teoretický :	5,991
Pravdepodobnosť :	0,236

**Vybočujúce body :**

Homogenita :	Prijatá
Počet vybočujúcich bodov :	0

**Záver**

Hmotnosť cestovín nameraná na 20 vzorkách vykazuje normálne rozdelenie, čo potvrdil test normality a grafy EDA (Obr. 22 - Obr. 26). Stredná hodnota je daná priemerom 399,4 g s intervalom spoľahlivosti 396,959 až 401,84. Hodnota daná výrobcom (400g) sa nachádza medzi dolnou a hornou medzou intervalu spoľahlivosti. Rovnaký výsledok potvrdil aj test strednej hodnoty s konštantou 400, kde hypotéza  $H_0$  je prijatá. Rozdiel je nevýznamný.

## Test zhodnosti (QC Expert)

### Zadanie

Dve množiny dát obsahujú hodnoty hematokrytu v krvi skupiny mužov a žien. Stanovte, či skupiny majú zhodné výsledky, resp. hodnoty hematokrytu sú pre mužov odlišné od hodnôt žien.

### Dáta

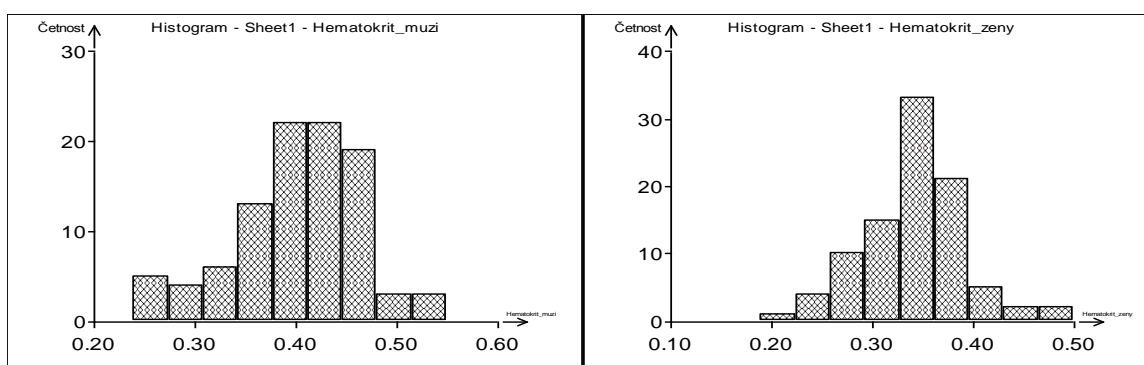
#### Muži:

0,44	0,281	0,45	0,463	0,48	0,536	0,47	0,36	0,48	0,41	0,36	0,4	0,41
0,36	0,43	0,45	0,46	0,47	0,35	0,42	0,45	0,39	0,3	0,48	0,39	0,28
0,43	0,34	0,42	0,25	0,43	0,31	0,45	0,36	0,44	0,39	0,42	0,34	0,44
0,37	0,47	0,36	0,44	0,3	0,25	0,42	0,32	0,44	0,44	0,44	0,38	0,27
0,25	0,39	0,48	0,31	0,35	0,43	0,42	0,45	0,45	0,51	0,4	0,41	0,42
0,35	0,41	0,34	0,4	0,37	0,38	0,43	0,4	0,39	0,43	0,37	0,35	0,41
0,43	0,52	0,41	0,49	0,45	0,38	0,48	0,45	0,41	0,41	0,4	0,55	0,36
0,43	0,4	0,46	0,43	0,51	0,24							

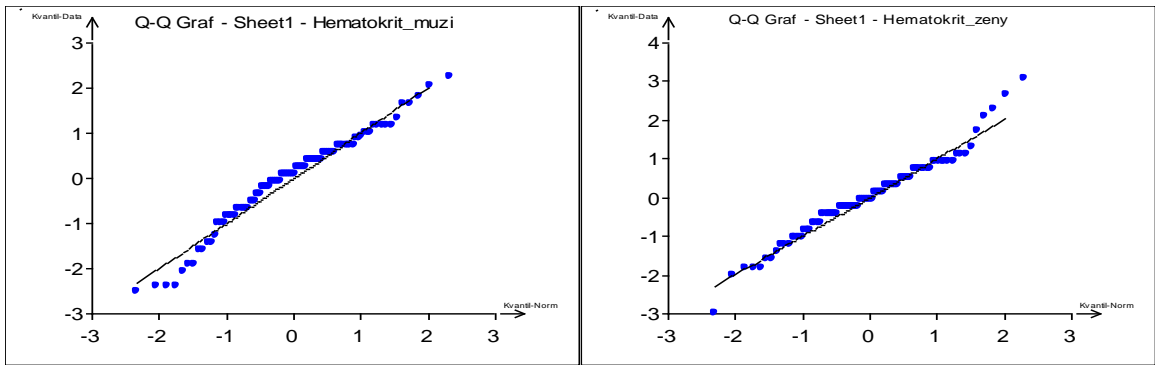
#### Ženy

0,34	0,38	0,35	0,33	0,45	0,19	0,31	0,25	0,36	0,39	0,37	0,39	0,28
0,36	0,36	0,34	0,29	0,32	0,25	0,27	0,24	0,38	0,35	0,32	0,3	0,33
0,36	0,34	0,38	0,35	0,32	0,4	0,39	0,34	0,31	0,31	0,31	0,38	0,29
0,26	0,38	0,33	0,36	0,32	0,37	0,4	0,38	0,28	0,36	0,39	0,29	0,32
0,35	0,38	0,33	0,34	0,33	0,29	0,36	0,34	0,3	0,38	0,28	0,4	0,26
0,35	0,33	0,32	0,33	0,37	0,37	0,25	0,36	0,34	0,39	0,34	0,37	0,46
0,48	0,33	0,5	0,39	0,3	0,32	0,39	0,35	0,33	0,33	0,43	0,32	0,37
0,41	0,33											

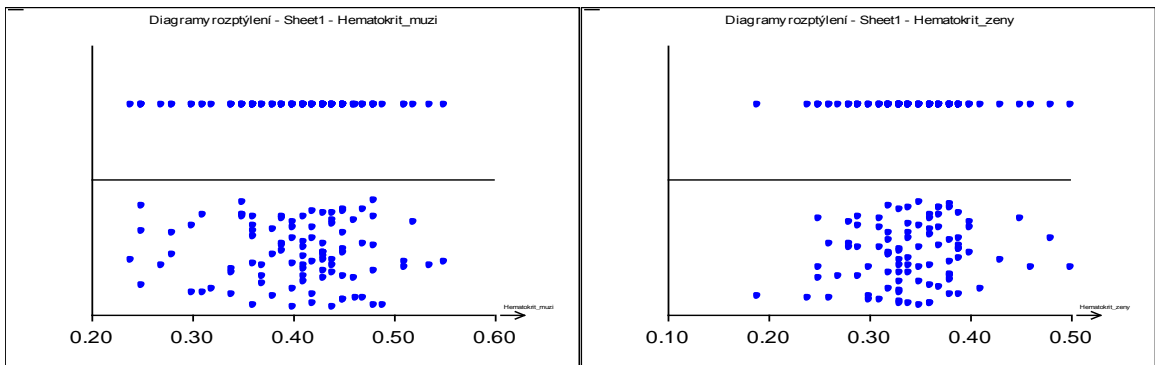
### Riešenie



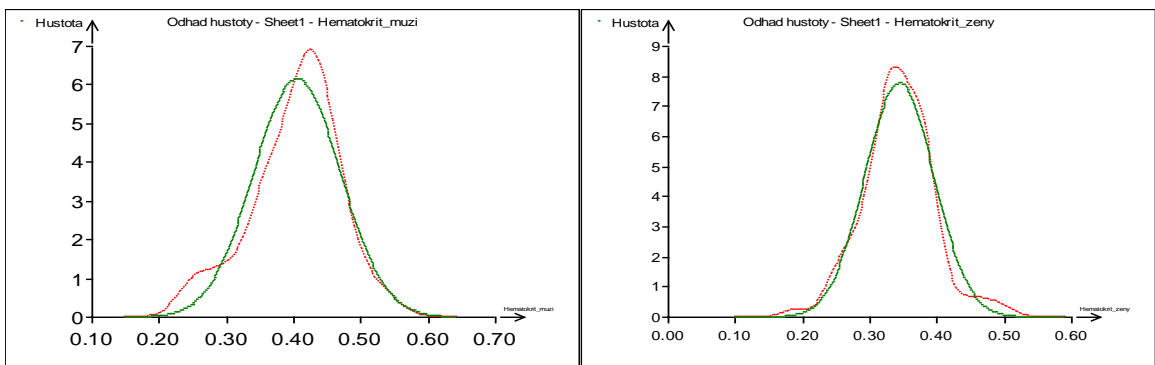
Obr. 27 Histogram



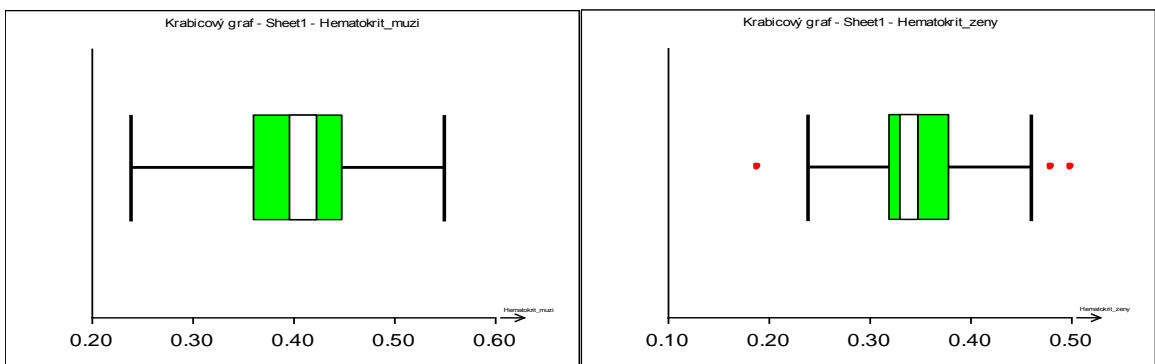
Obr. 28 Q-Q graf



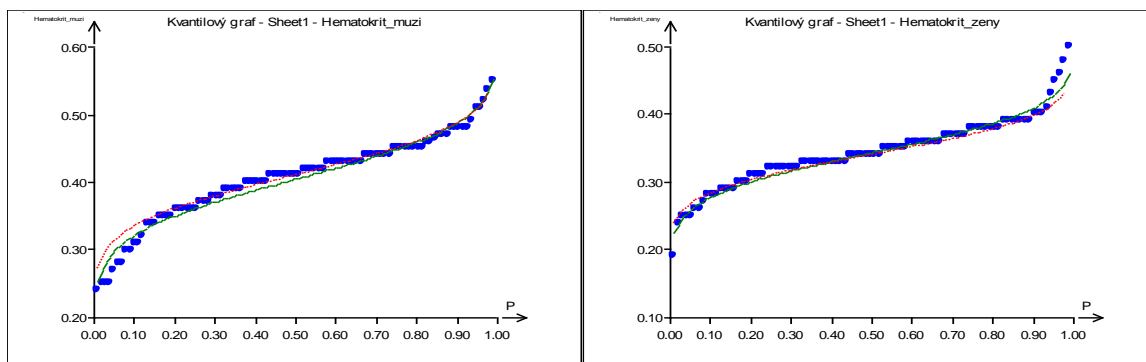
Obr. 29 Diagram rozptylenia



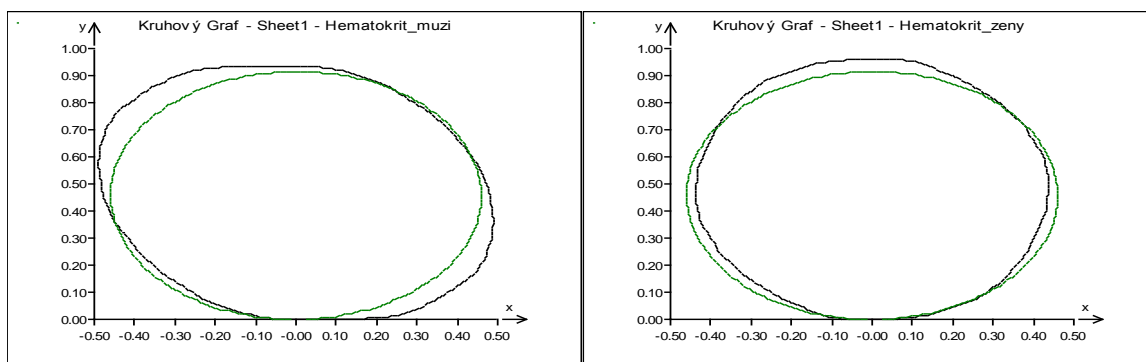
Obr. 30 Odhad hustoty



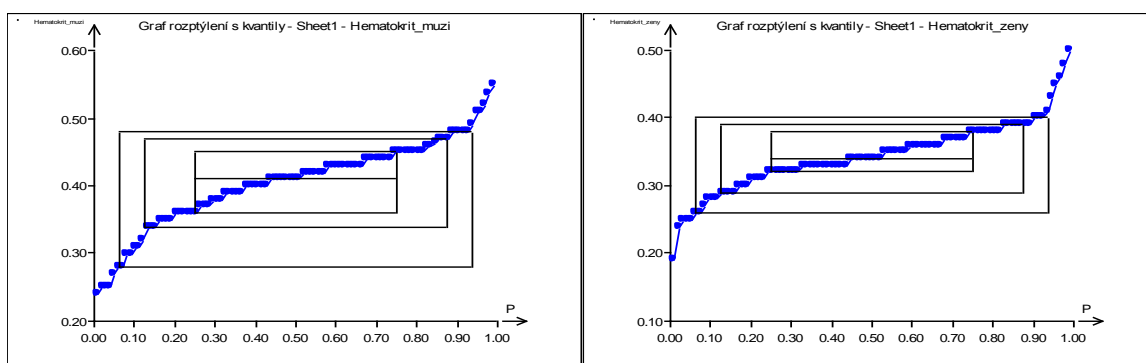
Obr. 31 Krabicový graf



Obr. 32 Kvantilový graf



Obr. 33 Kruhový graf



Obr. 34 Graf rozptýlenia s kvantilami

**Klasické parametre :**

Názov stĺpca :	Hematokrit_muži	Hematokrit_ženy
Priemer :	0,404	0,343
Spodná medza :	0,391	0,333
Horná medza :	0,417	0,354
Rozptyl :	0,00419	0,00263
Smer. odchýlka :	0,06474	0,05129
Šikmosť	-0,51707	0,11668
Odchýlka od 0 :	Významná	Nevýznamná
Špicatosť :	3,182	4,0383
Odchýlka od 3 :	Nevýznamná	Významná

**Robustné parametre :**

Názov stĺpca :	Hematokrit_muži	Hematokrit_ženy
Medián :	0,41	0,34

IS spodný :	0,3948	0,3248
IS horný :	0,425	0,355
Mediánová smer. odchýlka :	0,00765	0,00765
Mediánový rozptyl :	5,8571 E-005	5,857 E-005

### Znamienkový test :

Záver : **Dáta sú nezávislé**      **Dáta sú nezávislé**

### Test normality :

Názov stĺpca :	Hematokrit_muži	Hematokrit_ženy
Priemer :	0,404	0,3432
Rozptyl :	0,00419	0,002631
Šikmosť	-0,51707	0,11667
Špicatosť :	3,182	4,038
Normalita :	<b>Prijatá</b>	<b>Prijatá</b>
Vypočítané T :	4,235	0,339
Teoretické T :	5,9914	5,991
Pravdepodobnosť :	0,1203	0,84397

### Vybočujúce body :

Názov stĺpca :	Hematokrit_muži	Hematokrit_ženy
Homogenita :	Prijatá	Zamietnutá
Počet vybočujúcich bodov :	0	2

### Porovnanie dvoch výberov

Hladina významnosti : 0,05

#### Test zhody rozptylov

Pomer rozptylov :	1,59
Kritická hodnota :	1,3959
Záver :	<b>Rozptyly sú ROZDIELNE</b>
Pravdepodobnosť :	0,01168

#### Test zhody priemerov pre ROZDIELNE rozptyly

t - štatistika :	7,202
Redukované stupne voľnosti :	182
Kritická hodnota :	1,97
Záver :	<b>Priemery sú ROZDIELNE</b>
Pravdepodobnosť :	1,5149 E-011

### Záver

Pre skupinu mužov i žien, má obsah hematokrytu v krvi normálne rozdelenie, čo je potvrdené testom normality a vyhodnotením EDA grafov (Obr. 27 - Obr. 34). Pre porovnanie stredných hodnôt analyzovaných skupín test zhody rozptylov potvrdil ich rozdielnosť. Pre posúdenie stredných hodnôt bol vykonaný test zhody priemerov, ktorý deklaroval rozdielnosť obsahu hematokrytu v skupine mužov od hodnôt v skupine žien.



## Párový test (QC Expert)

### Zadanie

Pre pacientov s diagnózou D510 – anémia z nedostatku vitamínu B12 boli odobraté vzorky krvi pred nasadením liečby a pri prvej kontrole po liečbe. Úlohou je porovnať hodnoty hemoglobínu (HGB) a stanoviť či odlišnosť hodnôt po liečbe je štatisticky významná.

### Dáta

HGB_pred	HGB_po
138	148
89	116
115	118
134	133
164	154
111	123
129	140
140	131
130	136
144	141
115	123
136	140
143	145
134	145

### Riešenie

Hladina významnosti :	0,05	
Priemerná diferencia :	-5,0714	
Interval spoľahlivosti:	-10,56	0,42
Smer. odchýlka :	9,5	
Rozptyl :	90,38	
t-štatistika :	18,97	
Počet stupňov voľnosti :	13	
Kritická hodnota :	2,16	
Záver:	Rozdiely sú <b>NEVÝZNAMNÉ</b>	
Pravdepodobnosť :	0,0336	

### Záver

Hodnoty hemoglobínu po nasadení liečby u pacientov s diagnózou D510 nie sú odlišné od hodnôt pred zahájením liečby. Rozdiely sú štatisticky nevýznamné.

# Analýza rozptylu – ANOVA

## Jednofaktorová ANOVA (QC Expert)

### Zadanie

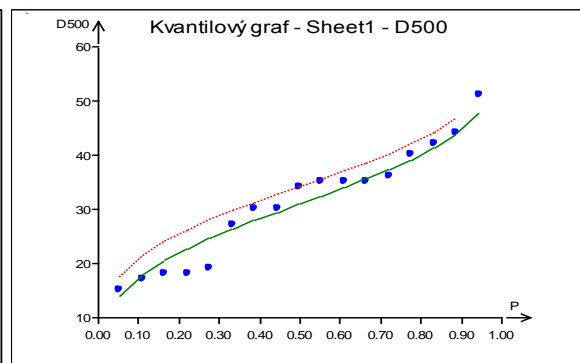
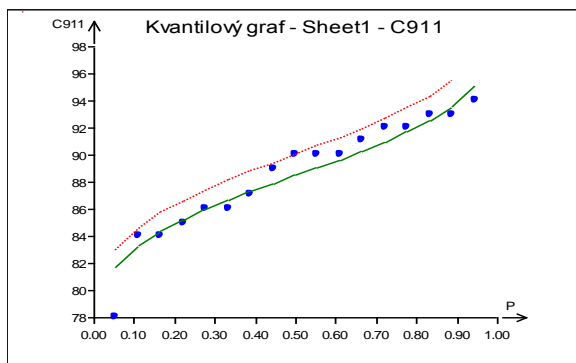
Pre 4 diagnózy C911, D500, D51, D693 bolo sledované množstvo leukocytov BMPV v krvi pacientov. Na hladine významnosti  $\alpha = 0.05$  vyšetrite, či je počet leukocytov rovnaký pre všetky diagnózy.

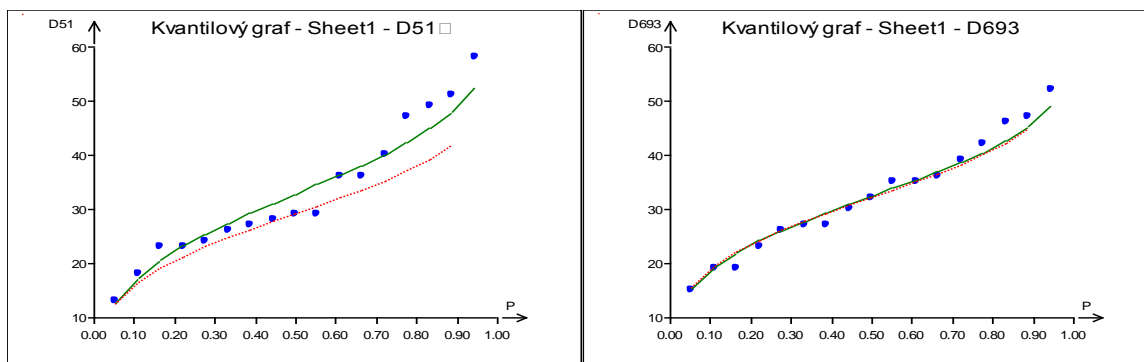
### Dáta

#### C911 D500 D51 D693

90	36	51	32
90	30	29	27
89	44	27	19
94	35	36	19
92	27	23	42
93	17	49	46
92	34	13	27
85	15	26	39
84	35	23	52
93	30	18	15
90	18	29	47
87	18	47	35
86	51	28	30
84	42	24	23
86	35	40	36
78	19	36	35
91	40	58	26

### Riešenie



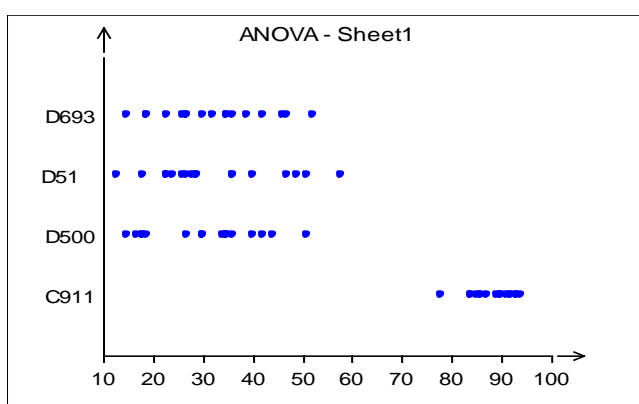


Obr. 35 Kvantilový graf lymfocytov BMPV pre diagnózu C911, D510, D51 a D693

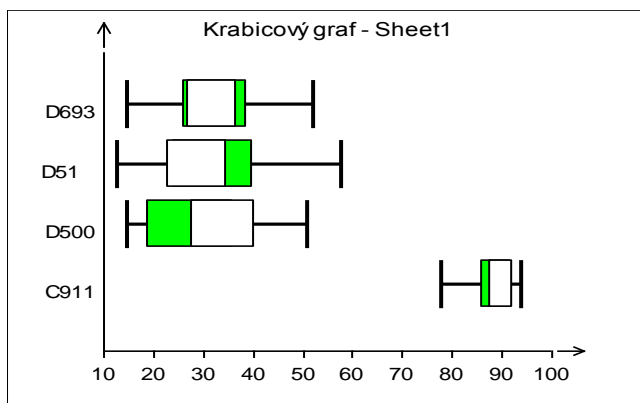
Klasické parametre :				
Diagnóza	<b>C911</b>	<b>D693</b>	<b>D500</b>	<b>D51</b>
Priemer :	88,47058824	30,94117647	32,76470588	32,35294118
Spodná medza :	86,29591705	25,48028513	26,31847165	26,899708
Horná medza :	90,64525942	36,40206781	39,21094011	37,80617435
Rozptyl :	17,88970588	112,8088235	157,1911765	112,4926471
Smer. odchýlka :	4,229622428	10,62114982	12,53759054	10,60625509
Šikmosť	-	-	-	-
Odchýlka od 0 :	Nevýznamná	Nevýznamná	Nevýznamná	Nevýznamná
Špicatosť :	3,151178337	2,052025321	2,305733489	2,129551372
Odchýlka od 3 :	Nevýznamná	Nevýznamná	Nevýznamná	Nevýznamná

Znamienkový test :				
Záver :	Dáta sú <b>nezavislé</b>	Dáta sú <b>nezavislé</b>	Dáta sú <b>nezavislé</b>	Dáta sú <b>nezavislé</b>

Normalita :	<b>Prijatá</b>	<b>Prijatá</b>	<b>Prijatá</b>	<b>Prijatá</b>
Vypočítaný :	2,763933367	0,052222408	1,372407864	0,224409483
Teoretický :	5,991464547	5,991464547	5,991464547	5,991464547
Pravdepodobnosť :	0,251084264	0,974226746	0,503483704	0,893861228



Obr. 36



Obr. 37

Všetky diagnózy (stĺpce) obsahujú hodnoty, ktoré vykazujú normálne rozdelenie. Podľa krabicového grafu Obr. 37 stĺpce neobsahujú odľahlé hodnoty. Analyzované dáta sú nezávislé. Overením výberových predpokladov je potvrdená vhodnosť vstupných dát pre ďalšie spracovanie a vyhodnotenie ANOVA výsledkov.

Celkový priemer :	46,132
Celkový rozptyl :	702,504
Priemerný štvorec :	692,173
Reziduálny rozptyl :	95,613
Reziduálny súčet štvorcov :	6406,117
Celkový súčet štvorcov :	47067,80
Vysvetlený súčet štvorcov :	40661,69

Počet úrovní faktoru : 4

Stĺpec	Počet hodnôt	Efekty faktorov	Priemer úrovne
C911	17	42,338	88,4705
D500	17	-15,191	30,9411
D51	17	-13,367	32,7647
D693	17	-13,779	32,3529

Test významnosti celkového vplyvu faktorov :

Záver	Teoretický	Vypočítaný	Pravdepodobnosť
<b>Významný</b>	2,7481	135,409	1,16E-27

Párové porovnávanie úrovní dvojíc Scheffého metóda

Porovnávané dvojice	Rozdiel	Významnosť	Pravdepodobnosť
C911 - D500	57,529	Významný	2,25E-23
C911 - D51	55,705	Významný	1,19E-22
C911 - D693	56,117	Významný	8,14E-23
D500 - D51	-1,82	Nevýznamný	0,963
D500 - D693	-1,41	Nevýznamný	0,982
D51 - D693	0,41	Nevýznamný	0,999

## Záver

Test významnosti vplyvu diagnózy na lymfocyty BMPV vykazuje štatistickú závislosť. Vypočítaná hodnota F kritéria je 135,409, čo je viac ako kritická hodnota 2,7481. Hypotéza  $H_0$  o rovnosti stredných hodnôt lymfocytov pre jednotlivé diagnózy je zamietnutá. Viacnásobné porovnanie pomocou Scheffého procedúry sú testované lineárne kontrasty pre zadané kombinácie úrovní,  $H_0: \mu_i - \mu_j = 0$ . Tri dvojice faktora diagnóza majú hypotézu  $H_0$  zamietnutú. Jedná sa o diagnózu C911 v kombinácií so všetkými ďalšími diagnózami. Ostatné úrovne sa považujú za zhodné.

## Dvojfaktorová ANOVA bez opakovania (ADSTAT)

### Zadanie

Preverte vplyv diagnózy a pohlavia na hodnotu lymfocytov v krvi pacientov na hladine významnosti  $\alpha = 0,05$ . Faktor A reprezentuje pohlavie pacienta a faktor B diagnózu: C911, D500, D51, D693.

### Dáta

	Muži	Ženy
D500	50	30
C911	90	90
D51	29	51
D693	19	32

### Riešenie

#### (1) PRŮMĚRY A EFEKTY ÚROVNÍ:

Celkový průměr = 4.8875E+01  
 Reziduální rozptyl = 2.4919E+02

F A K T O R A:			F A K T O R B:		
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	4.0000E+01	-8.8750E+00	1	4.7000E+01	-1.8750E+00
2	9.0000E+01	4.1125E+01	2	5.0750E+01	1.8750E+00
3	4.0000E+01	-8.8750E+00			
4	2.5500E+01	-2.3375E+01			

Tukeyho C = -3.5808E-02

#### (3) TABULKA ANOVA PRO MODEL S TUKEYHO INTERAKCÍ:

H<sub>0</sub>: Efekty faktoru A jsou nulové, H<sub>A</sub>: ... nejsou nulové

Kvantil F(1-alfa,n-1,mn-m-n) = 19.164

H<sub>0</sub>: Efekty faktoru B jsou nulové, H<sub>A</sub>: ... nejsou nulové

Kvantil F(1-alfa,m-1,mn-m-n) = 18.513

H<sub>0</sub>: Interakce I je nulová, H<sub>A</sub>: ... není nulová

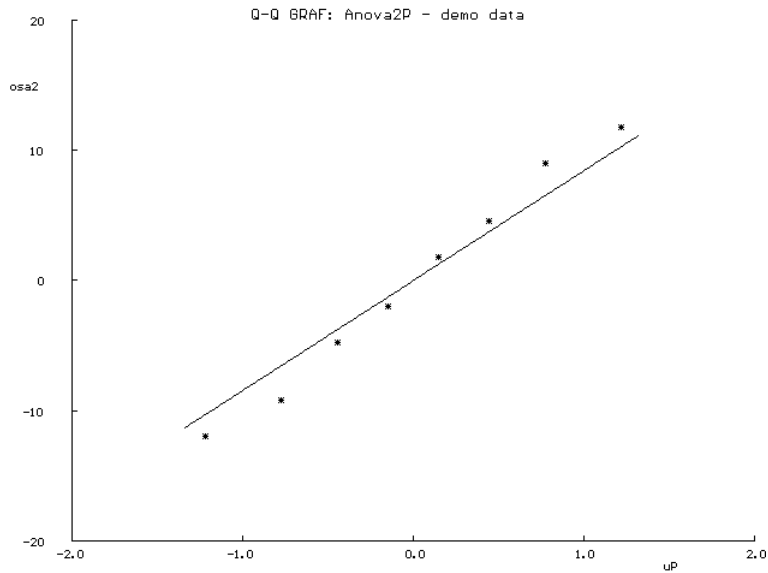
Kvantil F(1-alfa,1,mn-m-n) = 18.513

(Zde I znamená efekt Tukeyho interakce.)

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H <sub>0</sub> je	Spočtená hlad.úž.
Mezi úrovněmi A	n-1 = 3	4.7904E+03	1.5968E+03	6.408	Akceptována	0.138
Mezi levels B	m-1 = 1	2.8125E+01	2.8125E+01	0.113	Akceptována	0.769
Interakce	1	2.1594E+01	2.1594E+01	0.087	Akceptována	0.796
Rezidua	mn-m-n = 2	4.9838E+02	2.4919E+02			
Celková	mn-1 = 7	5.3169E+03	7.5955E+02			

**(4) TRANSFORMACE:**

Odhad mocninné transformace : 2.7501E+00  
Rozptyl odhadu transformace : 2.3563E+01  
Akceptovatelný interval : (-2.1041E+00, 7.6043E+00)



**Obr. 38 Q-Q graf**

**Záver**

Rozdiel obsahu lymfocytov v krvi pacientov nie je štatisticky významný pre analyzované faktory pohlavie a diagnóza. Testačné kritérium pre faktor A dosahuje hodnotu 6,408, čo je menej ako kritická hodnota 19,164. Testačné kritérium pre faktor B má hodnotu 0,113, čo je menej ako 18,513. V obidvoch prípadoch bola potvrdená hypotéza  $H_0$ .

## Dvojfaktorová ANOVA s opakováním (ADSTAT)

### Zadanie

U troch vekových skupín pacientov (Faktor A) bol sledovaný obsah hemoglobínu v krvi separátne pre skupinu mužov a skupinu žien (Faktor B). Na hladine významnosti  $\alpha = 0.05$  vyšetríte, či vek ovplyvňuje obsah hemoglobínu a či je odlišnosť medzi výsledkami u mužov a u žien.

### Dáta

#### V S T U P

#### (1) DATA A PODMÍNKY:

Hladina významnosti alfa : 0.050  
Transformace : Ne  
Počet úrovní parametru A, n : 2  
Počet úrovní parametrů B, m : 3  
Počet opakování v jedné buňce, o : 6  
Jméno výstupního souboru : RESULTS.TXT

#### (2) MĚŘENÍ NA RŮZNÝCH ÚROVNÍCH FAKTORŮ A ,B:

	Úrovně faktoru B		
	1	2	3
(pro každou úroveň faktorů A a B je měření 6 krát opakováno)			
Úrovně faktoru A			
1	1.4600E+02 1.5000E+02 1.5100E+02 1.6200E+02 1.6500E+02 1.4700E+02	1.2100E+02 1.3800E+02 1.3300E+02 1.3300E+02 1.2000E+02 1.4300E+02	1.4600E+02 1.5300E+02 1.2900E+02 9.3000E+01 1.0800E+02 1.4100E+02
2	1.2200E+02 1.1700E+02 1.1700E+02 1.4400E+02 1.3100E+02 1.2700E+02	1.3500E+02 1.4000E+02 1.5100E+02 1.2000E+02 1.2300E+02 1.1400E+02	1.2100E+02 1.2900E+02 1.0800E+02 1.4300E+02 1.3500E+02 1.2200E+02

### Riešenie

#### (1) PRŮMĚRY A ÚROVNĚ EFEKTŮ:

Celkový průměr = 1.3272E+02  
Reziduální rozptyl = 1.9088E+02

F A K T O R A:			F A K T O R B:		
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	1.3772E+02	5.0000E+00	1	1.3992E+02	7.1944E+00
2	1.2772E+02	-5.0000E+00	2	1.3092E+02	-1.8056E+00
			3	1.2733E+02	-5.3889E+00



(3) TABULKA ANOVA PRO MODEL S INTERAKCEMI FAKTORŮ A, B:

H<sub>0</sub>: Efekty faktoru A jsou nulové, H<sub>A</sub>: ... nejsou nulové

Kvantil  $F(1-\alpha, n-1, mn(o-1)) = 4.171$

H<sub>0</sub>: Efekty faktoru B jsou nulové, H<sub>A</sub>: ... nejsou nulové

Kvantil  $F(1-\alpha, m-1, mn(o-1)) = 3.316$

H<sub>0</sub>: Interakce I je nulová, H<sub>A</sub>: ... není nulová

Kvantil  $F(1-\alpha, (n-1)(m-1), nm(o-1)) = 3.316$

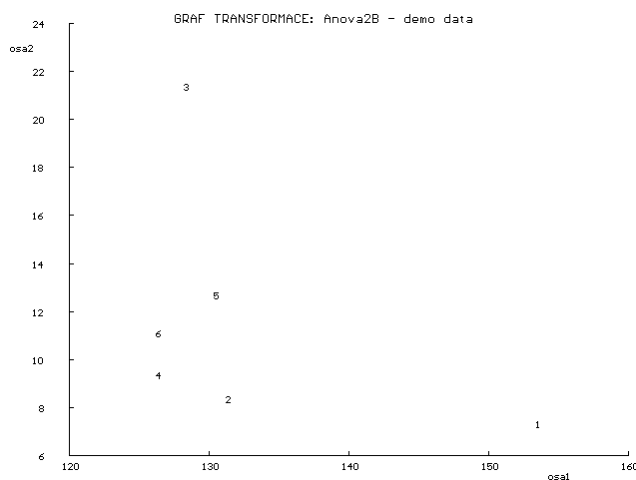
(Zde I znamená efekty interakcí A a B dohromady)

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H <sub>0</sub> je	Spočtená hlad.výz.
Mezi úrovněmi A	n-1 = 1	9.0000E+02	9.0000E+02	4.715	Zamítnuta	0.038
Mezi úrovněmi B	m-1 = 2	1.0087E+03	5.0436E+02	2.642	Akceptována	0.088
Interakce	(n-1)(m-1) = 2	1.3282E+03	6.6408E+02	3.479	Zamítnuta	0.044
Rezidua	mn(o-1) = 45	5.7263E+03	1.9088E+02			
Celkový	mno-1 = 35	8.9632E+03	2.5609E+02			

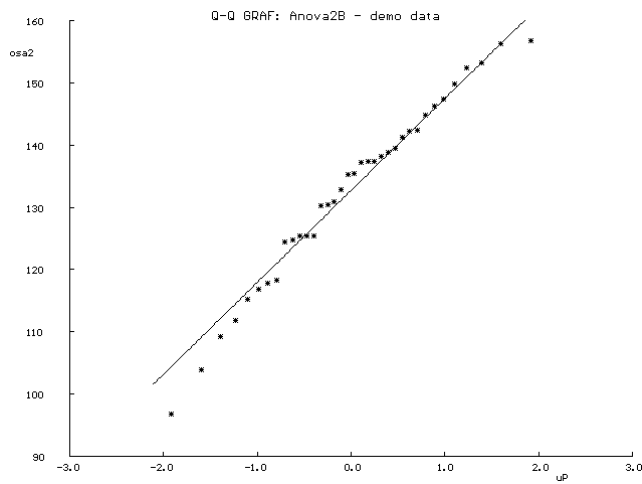
(5) ZKOUŠKA TRANSFORMACE:

Korelační koeficient, R : -0.424

(Pokud je R blízké nule není transformace nutná.)



Obr. 39 Graf transformácie



**Obr. 40 Q-Q graf**

## **Záver**

Vplyv faktoru pohlavie (A) je štatisticky významný a vplyv faktoru vek (B) je nevýznamný. Testačné kritérium pre faktor A dosahuje hodnotu 4,715, čo je viac ako kritická hodnota 4,171. Pre faktor B vypočítaná hodnota kritéria dosahuje hodnotu 2,642, čo je menej ako hraničná hodnota 3,316. Na základe výsledkov korelácie a grafov Obr. 39, Obr. 40 transformácia nie je potrebná. Korelačný koeficient dosahuje hodnotu -0,42. Interakcia faktorov je v teste uvedená ako významná. Vzhľadom na jej logickú neopodstatnenosť nie je potrebná.