

Univerzita Pardubice
Fakulta chemicko-technologická

Katedra analytické chemie

SEMESTRÁLNÍ PRÁCE

Licenční studium

Statistické zpracování dat při managementu jakosti

Předmět

Statistická analýza jednorozměrných dat

Obsah:

1.	Statistická analýza velkých výběrů	3
1.1.	Zadání	3
1.2.	Data	3
1.3.	Program	3
1.4.	Průzkumová (exploratorní) analýza dat EDA	4
1.5.	Ověření předpokladů o datech	7
1.6.	Transformace dat	9
1.7.	Základní analýza jednorozměrných dat	12
1.8.	Hodnocení	13
1.9.	Závěr	14
2.	Statistická analýza malých výběrů	15
2.1.	Zadání	15
2.2.	Data	15
2.3.	Hornův postup	15
2.4.	Program	16
2.4.1	Průzkumová (exploratorní) analýza dat EDA	17
2.4.2	Ověření základních předpokladů	20
2.5.	Srovnání výsledků	22
2.6.	Hodnocení	23
2.7	Závěr	23
3.	Statistické testování	24
3.1.	Test shody středních hodnot (testování shodnosti)	24
3.1.1	Zadání	24
3.1.2	Data	24
3.1.3	Program	24
3.1.4	EDA a ověření základních předpokladů	24
3.1.5	Hodnocení základních předpokladů o datech	26
3.1.6	Porovnání dvou výběrů	26
3.1.7	Hodnocení	29
3.1.8	Závěr	29
3.2.	Test střední hodnoty (testování správnosti)	30
3.2.1	Zadání	30
3.2.2	Data	30
3.2.3	Program	30
3.2.4	Průzkumová (exploratorní) analýza dat EDA	30
3.2.5	Hodnocení EDA	31
3.2.6	Základní analýza dat	31
3.2.7	Hodnocení základních předpokladů o datech	33
3.2.8	Celkové hodnocení	33
3.2.9	Závěr	33
3.3.	Párový test	34
3.3.1	Zadání	34
3.3.2	Data	34
3.3.3	Program	34
3.3.4	Ověření základních předpokladů o datech	34
3.3.5	Hodnocení	35
3.3.6	Závěr	35
Přílohy:	Output 1 Statistická analýza velkých výběrů	TOC1.vts
	Output 2 Statistická analýza malých výběrů	TOC2.vts
	Output 3 Test shody středních hodnot	TOC3.vts
	Output 4 Test střední hodnoty	TOC4.vts
	Output 5 Párový test	TOC5.vts
	CD disk	

1. Statistická analýza velkých výběrů

1.1. Zadání

Typ rozdělení, odlehle hodnoty, určení střední hodnoty obsahu TOC (celkového organického uhlíku) ve vodném roztoku.

V rámci evaluace analytické metody vznikl požadavek na statistickou analýzu dat získaných v uplynulém evaluačním období. V laboratoři Biochemické a chemické kontroly byl stanovován obsah celkového organického uhlíku u kontrolního vzorku (roztok sacharózy s koncentrací 0,50 mg TOC/l). Stanovení bylo prováděno na TOC analyzátoru metodou TOC, což znamená, že nejprve byl změřen obsah celkového uhlíku, pak obsah anorganického uhlíku a odečtením těchto hodnot byl stanoven obsah TOC ve vzorku.

Proveďte předpoklady o výběru, určete střední hodnotu a intervaly spolehlivosti. Identifikujte odlehle hodnoty a určete, zda je vhodné tyto hodnoty vypustit.

1.2. Data

Obsah TOC [mg/l]
n = 100

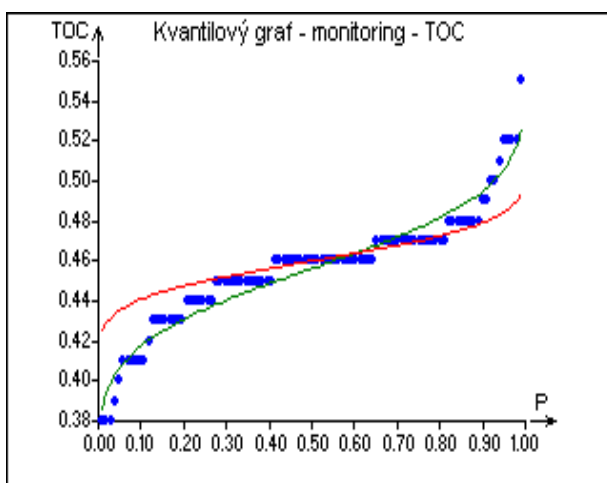
0,45	0,52	0,45	0,45
0,46	0,46	0,49	0,38
0,52	0,43	0,38	0,46
0,48	0,43	0,45	0,45
0,41	0,43	0,47	0,48
0,43	0,47	0,50	0,46
0,41	0,48	0,47	0,49
0,44	0,48	0,44	0,52
0,46	0,47	0,47	0,43
0,46	0,45	0,45	0,41
0,46	0,46	0,48	0,41
0,48	0,38	0,45	0,39
0,46	0,46	0,46	0,47
0,46	0,47	0,41	0,46
0,47	0,44	0,47	0,46
0,48	0,47	0,45	0,46
0,55	0,45	0,45	0,47
0,46	0,47	0,46	0,44
0,46	0,46	0,44	0,47
0,46	0,40	0,41	0,45
0,47	0,43	0,46	0,47
0,46	0,44	0,45	0,52
0,47	0,46	0,43	0,42
0,48	0,45	0,51	0,43
0,50	0,44	0,47	0,46

1.3. Program

Expert 2.27

Základní statistika

1.4. Průzkumová (exploratorní) analýza dat EDA



Kvantilový graf

Empirické hodnoty dat jsou proloženy kvantilovou funkcí normálního rozdělení.

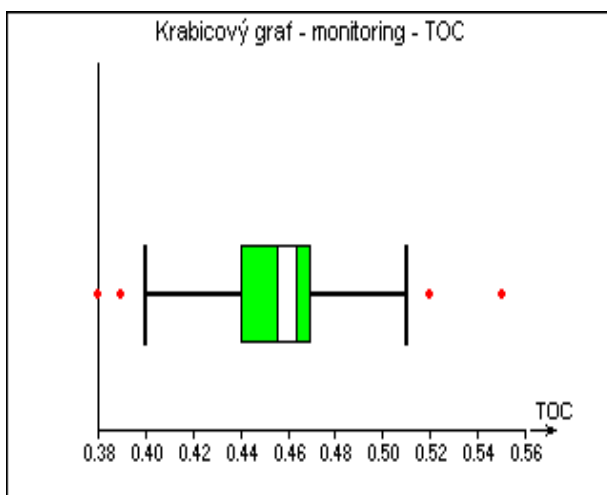
Zelená křivka – nerobustní (aritmetický průměr a rozptyl)

Červená křivka – robustní (medián a mediánová odchylka).

Hodnocení:

Data lépe prokládá zelená křivka, pravděpodobně normální rozdělení.

Nahoře i dole diagnostikovány podezřelé hodnoty.



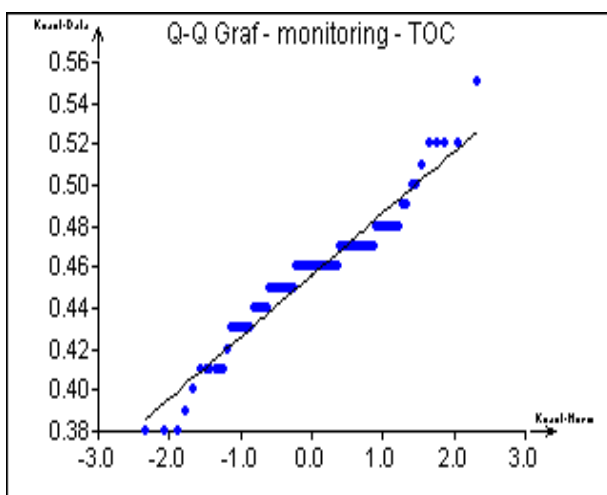
Krabicový graf

V okolí kvartilů a vnitřních hradeb je patrná symetrie. Medián je mírně posunutý směrem k vyšším hodnotám.

Hodnocení:

Jedná se o symetrické rozdělení, pravděpodobně normální.

Jsou diagnostikovány podezřelé hodnoty ležící mimo interval vnitřních hradeb.



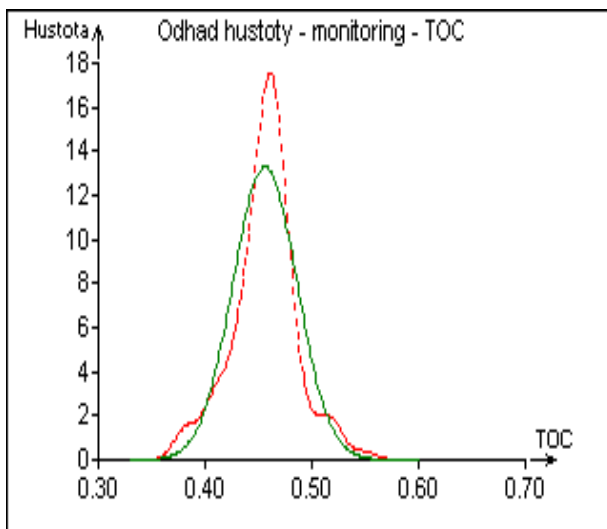
Kvantilově - kvantilový graf

Vysoká koncentrace dat kolem střední hodnoty a mírné konkávně-konvexní tvar naznačuje, že se jedná o data s vyšší špičatostí než je normální. Kolmice vztyčená v 0 se promítá na osu y jako aritmetický průměr.

Hodnocení:

Jedná se o symetrické pravděpodobně Laplaceovo rozdělení.

Jsou diagnostikovány podezřelé hodnoty: nahoře 4, dole 4.



Hustota pravděpodobnosti

Zelená čára – hustota pravděpodobnosti normálního rozdělení.
 Přerušovaná červená čára – jádrový odhad hustoty vypočtený na základě dat naznačuje, že se jedná o data s vyšší špičatostí než je normální. Kolmice spuštěná z vrcholu empirické křivky ukazuje na ose x hodnotu mediánu, kolmice spuštěná z vrcholu Gaussovy křivky ukazuje na ose x hodnotu aritmetického průměru. Není zde žádný zřejmý posun.

Hodnocení:

Ve srovnání s Gaussovým rozdělením je patrné, že se jedná o data s vyšší špičatostí (pravděpodobně Laplaceovo rozdělení), zešikmení směrem k vyšším hodnotám je zanedbatelné.

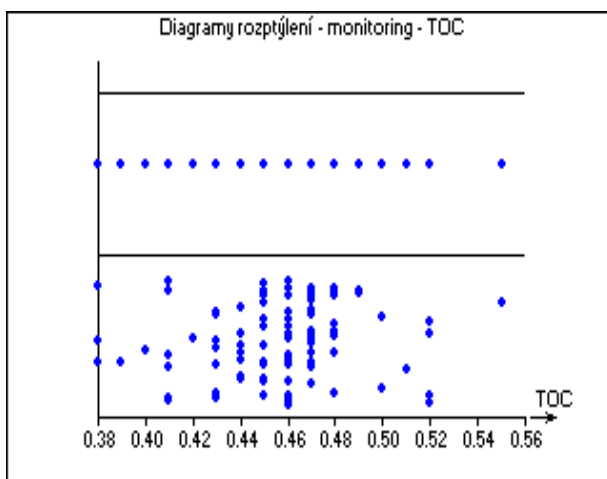
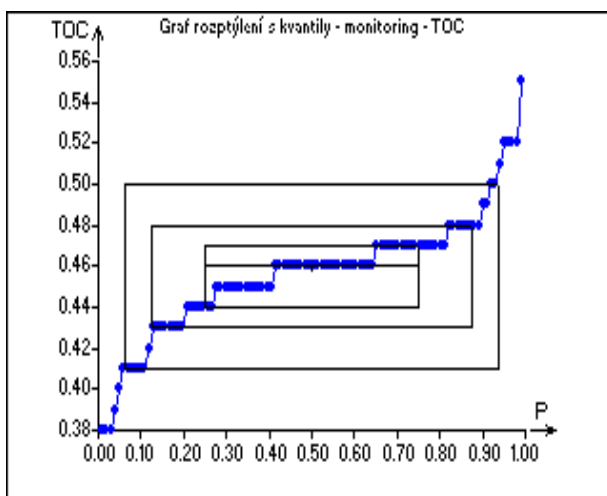


Diagram rozptýlení

Vysoká koncentrace dat kolem střední hodnoty naznačuje, že se jedná o data s vyšší špičatostí než je normální.

Hodnocení:

Jedná se o symetrické pravděpodobně Laplaceovo rozdělení. Nahoře i dole jsou diagnostikovány podezřelé hodnoty.

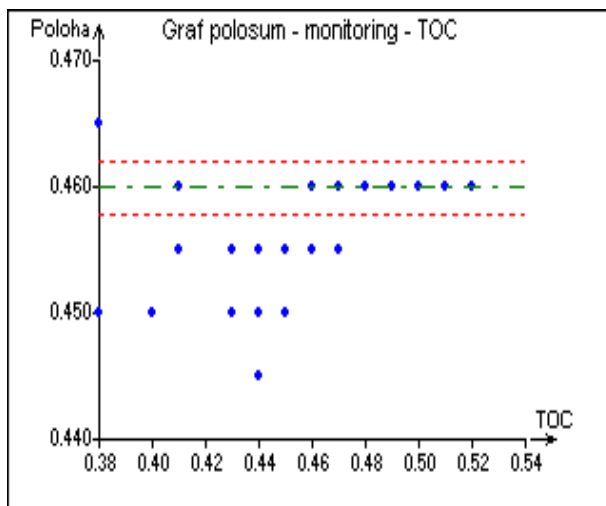


Graf rozptýlení s kvantily

Vzájemná poloha obdélníků dokazuje symetrické rozdělení. Jsou indikovány body vně sedecilových obdélníků.

Hodnocení:

Jedná se o symetrické rozdělení. Podezřelé hodnoty: nahoře 8, dole >10.

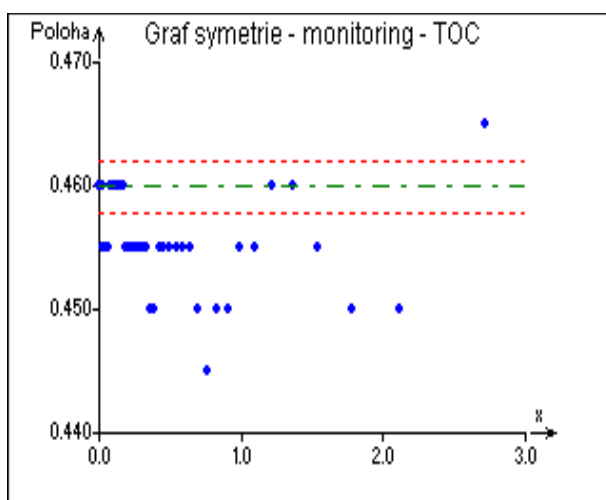


Graf polosum

Body neoscilují kolem horizontální přímky, leží přímo na ní nebo ve spodní části grafu mimo interval spolehlivosti mediánu.

Hodnocení:

Jedná se o symetrické rozdělení s delšími konci a s velmi úzkým intervalem spolehlivosti mediánu.

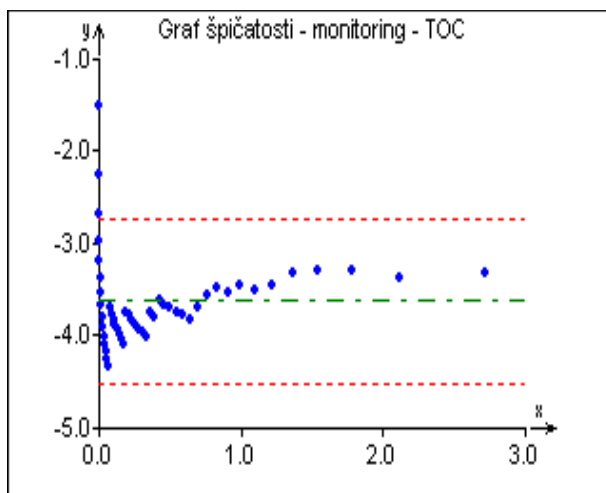


Graf symetrie

Směrnice trendu je odhadem parametru šikmosti. Velice úzký interval spolehlivosti mediánu způsobuje, že velké množství bodů leží mimo tento interval.

Hodnocení:

Data jsou uspořádána v mírně klesajícím trendu, který indikuje kladnou šikmost, tzn. mírné sešikmení směrem k vyšším hodnotám. Body mimo interval spolehlivosti mediánu jsou indikovány jako vybočující ze symetrického rozdělení.

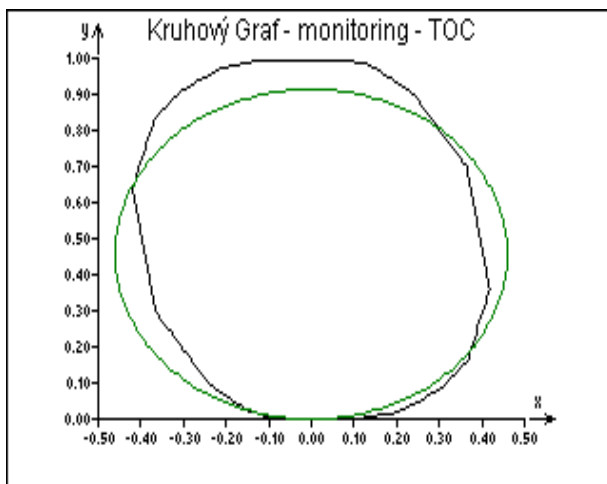


Graf špičatosti

Směrnice trendu je úměrná odchylce špičatosti (od 3).

Hodnocení:

Velké množství bodů neleží na rovnoběžce s osou x, což ukazuje na jiné rozdělení než normální.

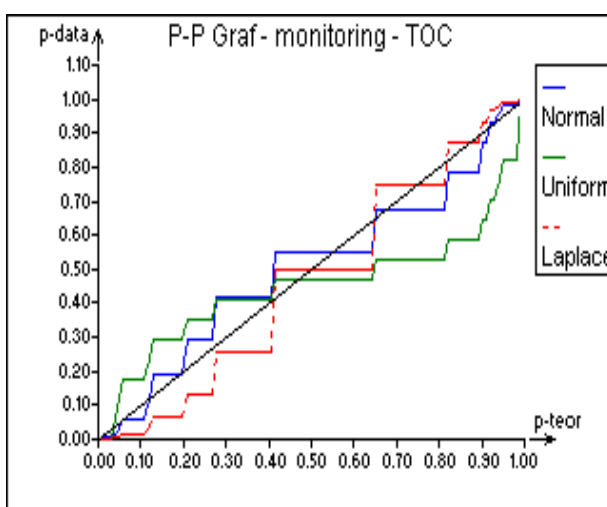


Kruhový graf

Zelený kruh je optimální tvar pro normální rozdělení.

Hodnocení:

Kruh vytvořený z posuzovaných dat se zcela nekryje s optimálním kruhem, ukazuje na velice mírné zešikmení.



Pravděpodobnostní graf

Empirická křivka souhlasí s křivkou symetrického rozdělení (normálního a Laplaceova).

Závěr:

Jedná se o symetrické rozdělení.

Závěr EDA: Rozdělení je symetrické, pravděpodobně normální nebo Laplaceovo s velkým množstvím vybočujících hodnot nahoře i dole. Ty, vzhledem k charakteru experimentu, nelze vyloučit, protože by to vedlo ke ztrátě velkého množství dat.

1.5. Ověření předpokladů o datech

Název úlohy :	Monitoring
Data:	Všechna
Řád trendu :	4
Testovaná hodnota :	0
Vyhlazení hustoty :	0,5
Hladina významnosti :	0,05
Název sloupce :	TOC
Počet platných dat :	100

(1) Ověření normality rozdělení

Klasické parametry :

Název sloupce :	TOC
Průměr :	0,4562
Spodní mez :	0,4502454788
Horní mez :	0,4621545212
Rozptyl :	0,0009005656566

Statistická analýza jednorozměrných dat

Směr. odchylka :	0,03000942613
Šikmost	-0,07217374904
Odchylka od 0 :	Nevýznamná
Špičatost :	4,13741127
Odchylka od 3 :	Významná
Polosuma	0,465
Modus :	0,467449505

<u>Test normality :</u>	
Název sloupce :	TOC
Průměr :	0,4562
Rozptyl :	0,0009005656566
Šikmost	-0,07217374904
Špičatost :	4,13741127
Normalita :	Přijata
Vypočtený :	0,1606655179
Teoretický :	5,991464547
Pravděpodobnost :	0,9228092223

(2) Ověření nezávislosti dat

Znaménkový test :

Závěr : **Data jsou nezávislá**

(3) Ověření homogenity

<u>Vybočující body :</u>	
Název sloupce :	TOC
Homogenita :	Zamítnuta
Počet vyb. bodů:	1
Spodní mez :	0,37358
Horní mez :	0,53642

<u>Robustní parametry :</u>	
Název sloupce :	TOC
Medián :	0,46
IS spodní :	0,4549381291
IS horní :	0,4650618709
Med. směr. odchylka :	0,002551067266
Mediánový rozptyl :	6,507944197E-006
10% Průměr :	0,4562222222
10% IS spodní :	0,4505950107
10% IS horní :	0,4618494337
10% Směr. odchylka :	0,0198805524
10% Rozptyl :	0,0003952363636
20% Průměr :	0,4565
20% IS spodní :	0,4511279312
20% IS horní :	0,4618720688
20% Směr. odchylka :	0,014157555
20% Rozptyl :	0,0002004363636
40% Průměr :	0,4578333333
40% IS spodní :	0,4537681099
40% IS horní :	0,4618985568
40% Směr. odchylka :	0,00745044233
40% Rozptyl :	5,550909091E-005

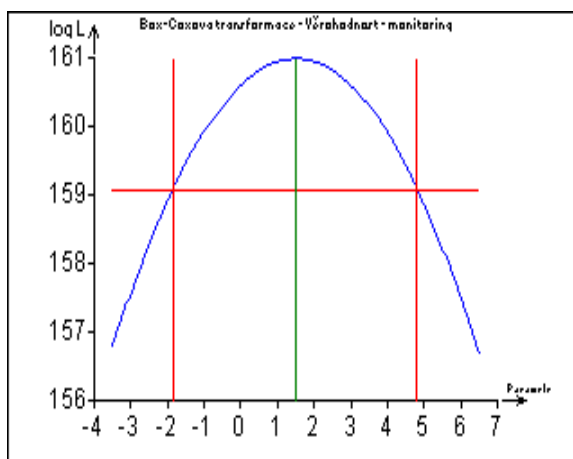
(4) Závěr – ověření předpokladů:

V datech se vyskytly mírné odchylky od normality.
 Třetí statistický moment - šikmost - odchylka od 0 byla hodnocena jako nevýznamná.
 Čtvrtý statistický moment - špičatost - odchylka od 3 byla hodnocena jako významná.
 Její hodnota 4,137 se nachází mezi normálním a Laplaceovým rozdělením, má však blíže k normálnímu.
 Pro přiblížení rozdělení výběru k normálnímu bude provedena transformace.

1.6. Transformace dat

Box-Coxova transformace dat

Název úlohy : Monitoring
 Data: Všechna
 Vybrané sloupce : TOC

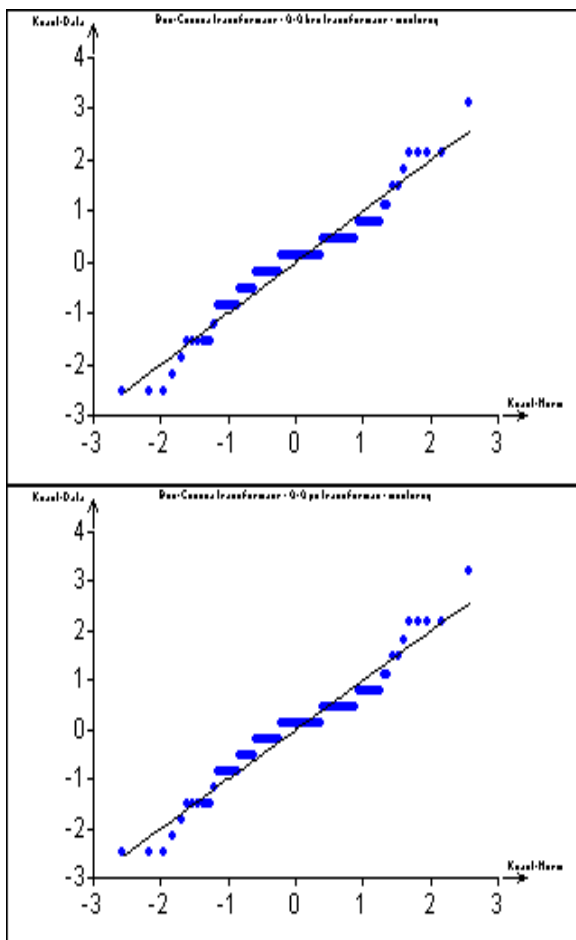


Box-Coxova transformace Věrohodnost

Z grafu je patrné, že 95% ní konfidenční interval parametru λ zahrnuje i hodnotu 1.

Hodnocení:

Ze statistického hlediska není transformace nutná.



Box-Coxova transformace Q-Q před transformací a Q-Q po transformaci

Body na grafu po transformaci nejsou blíže přímce než na grafu před transformací.

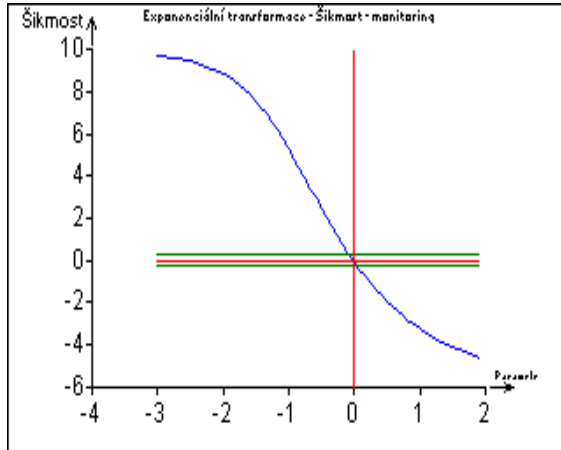
Hodnocení:

Transformace nebyla úspěšná.

Optimální parametr :	1,506063843
Dolní mez parametru :	-1,835644709
Horní mez parametru :	4,797521083
Věrohodnost bez transformace :	160,9437912
Věrohodnost s transformací :	160,9876579
Oprávněnost transformace :	Ne
Pravděpodobnost :	23,2641921919496 %
Zvolený parametr :	1,506063843
Věrohodnost :	160,9876579
Opravený průměr :	0,4565009292
LCL :	0,363594382
UCL :	0,5439069323
LWL :	0,4262114394
UWL :	0,486183891

Exponenciální transformace dat :

Název úlohy : Monitoring
 Data: Všechna
 Vybrané sloupce : TOC

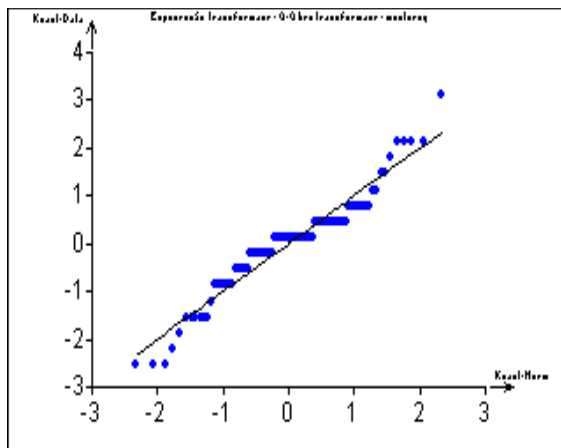


Exponenciální transformace Šikmost

Průsečík svislé přímky s křivkou leží v intervalu spolehlivosti pro šikmost.

Hodnocení:

Ze statistického hlediska není transformace nutná.

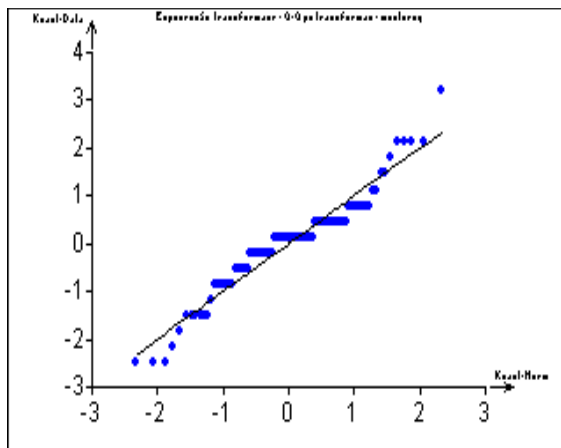


Exponenciální transformace Q-Q před transformací a Q-Q po transformaci

Body na grafu po transformaci nejsou blíže přímce než na grafu před transformací.

Hodnocení:

Transformace nebyla úspěšná.



Optimální parametr :	-0,01544189453
Zvolený parametr :	-0,01544189453
Oprávněnost transformace :	Ne
Opravený průměr :	0,4564281974
Interval spolehlivosti :	
Spodní :	0,4504665717
Horní :	0,4623715991
LCL :	0,3643137981
UCL :	0,5444041929
LWL :	0,3954938337
UWL :	0,5155191262

Závěr – transformace se ukázala jako neopodstatněná, nevedla ke zlepšení rozdělení.

1.7. Základní analýza jednorozměrných dat

Název úlohy : Monitoring
Data: Všechna

Řád trendu : 4
Testovaná hodnota : 0
Vyhlazení hustoty : 0,5
Hladina významnosti : 0,05
Název sloupce : TOC
Počet platných dat : 100

Klasické parametry :

Název sloupce : TOC
Průměr : 0,4562
Spodní mez : 0,4502454788
Horní mez : 0,4621545212
Rozptyl : 0,0009005656566
Směr. odchylka : 0,03000942613
Šikmost: -0,07217374904
Odchylka od 0 : Nevýznamná
Špičatost : 4,13741127
Odchylka od 3 : Významná
Polosuma: 0,465
Modus : 0,467449505

Robustní parametry :

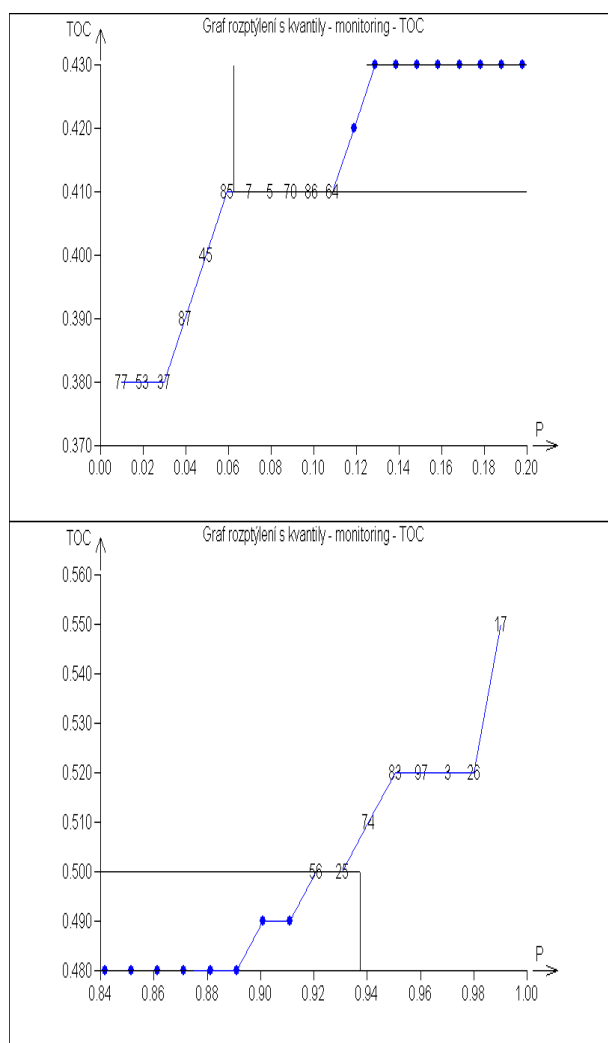
Název sloupce : TOC
Medián : 0,46
IS spodní : 0,4549381291
IS horní : 0,4650618709
Med. směr. odchylka : 0,002551067266
Medianový rozptyl: 6,507944197E-006
10% Průměr : 0,4562222222
10% IS spodní : 0,4505950107
10% IS horní : 0,4618494337
10% Směr. odchylka : 0,0198805524
10% Rozptyl : 0,0003952363636
20% Průměr : 0,4565
20% IS spodní : 0,4511279312
20% IS horní : 0,4618720688
20% Směr. odchylka : 0,014157555
20% Rozptyl : 0,0002004363636
40% Průměr : 0,4578333333

40% IS spodní : 0,4537681099
 40% IS horní : 0,4618985568
 40% Směr. odchylka : 0,00745044233
 40% Rozptyl : 5,550909091E-005

1.8. Hodnocení

Identifikace odlehlých bodů

Přestože v rámci ověření homogenity (CDA) byl odhalen jen 1 vybočující bod, EDA jich diagnostikovala více.



Graf rozptýlení s kvantily (výřezy s identifikací podezřelých dat)

Byly identifikovány podezřelé hodnoty na a vně sedecíů.

Č. měření	Obsah TOC [mg/l]
77, 53, 37	0,38
87	0,39
45	0,40
85, 7, 5, 70, 86, 64	0,41
56, 25	0,50
74	0,51
83, 97, 3, 26	0,52
17	0,55

Hodnocení:

Protože rozdělení bylo určeno jako normální, ovšem blíží se Laplaceovu, které se vyznačuje delšími konci a připouští výskyt výrazněji odchýlených hodnot, bude jako odlehlá hodnota označeno pouze měření č. 17.

Rozdělení je symetrické, pravděpodobně normální, nehomogenní. Byl odhalen jen 1 vybočující bod, který vzhledem k charakteru experimentu nelze vyloučit, Bude přihlédnuto k možnosti ovlivnění odhadů parametrů polohy a rozptýlení.

1.9. Hodnocení

V datech se vyskytly pouze nepatrné odchylky od normality, které výrazně neovlivní parametry polohy a rozptýlení. Není nutné použít robustních M odhadů, protože od klasických se výrazněji neodlišují. Průměrný obsah TOC u kontrolního vzorku roztoku sacharózy je 0,456 mg TOC/l se směrodatnou odchylkou 0,030; $IS_D = 0$, kontrolního vzorku 450 a $IS_H = 0,462$.

Správná hodnota obsahu TOC v kontrolním vzorku je 0,50 mg TOC/l, nespadá do intervalu spolehlivosti měření a způsob stanovení TOC je **NEVYHOVUJÍCÍ**.

2. Statistická analýza malých výběrů

2.1. Zadání

Hornova metoda pivotů k určení parametrů polohy a rozptýlení; porovnání výsledků s klasickými a robustními odhady polohy a rozptýlení určenými pomocí softwaru QC Expert.

V rámci validace nové analytické metody vznikl požadavek na statistickou analýzu dat získaných během validace. V laboratoři Biochemické a chemické kontroly byl stanoven obsah celkového organického uhlíku u kontrolního vzorku (roztok sacharózy s koncentrací 0,50 mg TOC/l). Stanovení bylo prováděno na TOC analyzátoru metodou NPOC. Vzorek je okyselen na pH 2 až 3 přidáním kyseliny fosforečné a probubláván inertním plynem, aby byly odstraněny všechny IC (anorganický uhlík) komponenty. Zbylý TC (celkový uhlík) ve vzorku po probublání je stanoven jako celkový organický uhlík a je označován jako TOC (okyselený a probubláný). Takto získaná hodnota TOC je označována jako NPOC, aby byl patrný rozdíl mezi hodnotou získanou tímto způsobem a hodnotou získanou metodou TOC pomocí odečtu koncentrace anorganického uhlíku od koncentrace celkového uhlíku.

2.2. Data

Obsah TOC [mg/l]
n = 12

0,510	0,482	0,490	0,509
0,501	0,509	0,486	0,488
0,508	0,505	0,506	0,493

2.3. Hornův postup

Statistické zpracování malých výběrů dat je často nepřesné a je zatíženo vysokou mírou statistické nejistoty. Z těchto důvodů je vhodné pro zpracování dat, kde $4 \leq n \leq 20$ použít Hornův postup, který vychází z pořádkových statistik. Hloubky pivotů přibližně odpovídají výběrovým kvartilům F.

(1) Pořádkové statistiky

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	0,482	0,486	0,488	0,490	0,493	0,501	0,505	0,506	0,508	0,509	0,509	0,510

(2) Hloubka pivotu

$$H = \frac{\left(\text{int} \frac{n+1}{2} \right)}{2} = 3$$

(3) Dolní a horní pivot

$$x_D = x_{(H)} = x_{(3)} = 0,488$$

$$x_H = x_{(n+1-H)} = x_{(10)} = 0,509$$

- (4) Pivotová polosuma (parametr polohy)

$$P_L = \frac{x_D + x_H}{2} = 0,4985$$

- (5) Pivotové rozpětí (parametr rozptýlení)

$$R_L = x_H - x_D = 0,021$$

- (6) Kvantil $t_{L,0.975(12)}$ je určen odečtem z tabulky

$$t_{L,0.975(12)} = 0,483$$

- (7) 95% interval spolehlivosti střední hodnoty

$$P_L - R_L t_{L,0.975(12)} \leq \mu \leq P_L + R_L t_{L,0.975(12)}$$

$$0,4884 \leq \mu \leq 0,5086$$

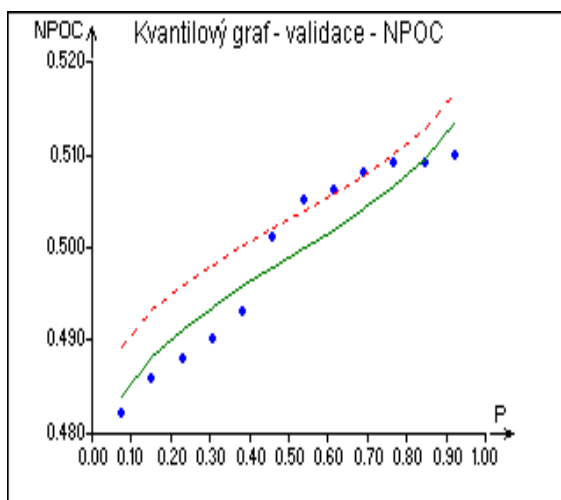
Závěr analýzy provedené dle Horna: S 95% statistickou jistotou se správná hodnota kontrolního vzorku (roztok sacharózy s koncentrací 0,50 mg TOC/l) nalézá v intervalu $L_D = 0,4884$ a $L_H = 0,5086$.

2.4. Program

Expert 2.27

Základní statistika

2.4.1 Průzkumová (exploratorní) analýza dat EDA



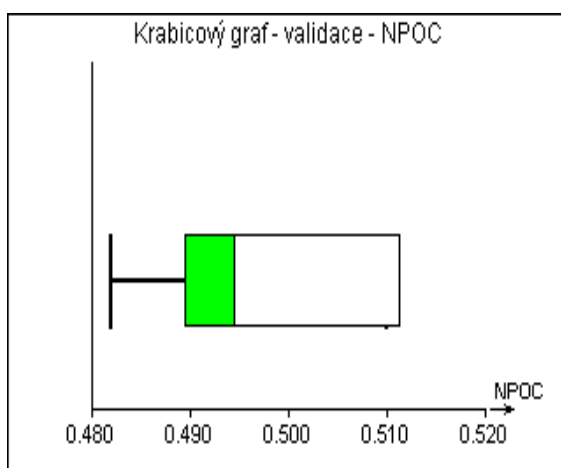
Kvantilový graf

Empirické hodnoty dat jsou proloženy kvantilovou funkcí normálního rozdělení.

Zelená křivka – nerobustní (aritmetický průměr a rozptyl)

Červená křivka – robustní (medián a mediánová odchylka).

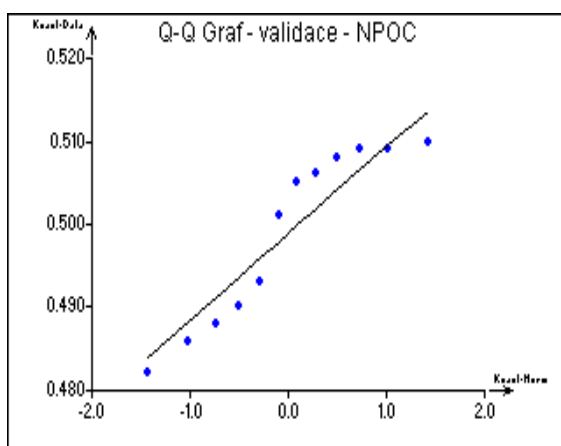
Data lépe prokládá zelená křivka, pravděpodobně normální rozdělení.



Krabicový graf

V okolí kvartilů a vnitřních hradeb není patrná symetrie.

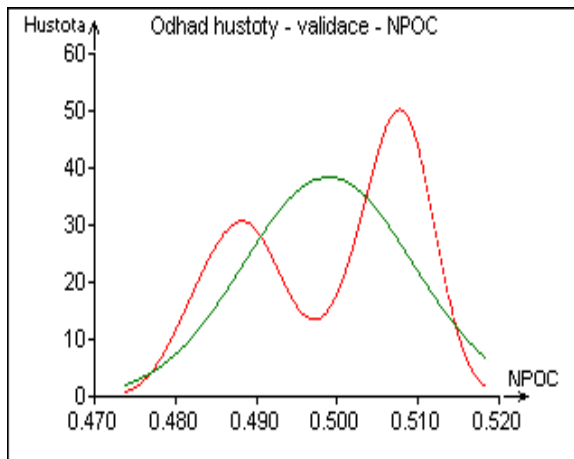
Nejsou diagnostikovány podezřelé hodnoty ležící mimo interval vnitřních hradeb.



Kvantilově - kvantilový graf

Kolmice vztyčená v 0 se promítá na osu y jako aritmetický průměr.

Nejsou diagnostikovány podezřelé hodnoty.



Hustota pravděpodobnosti

Zelená čára – hustota pravděpodobnosti normálního rozdělení.

Přerušovaná červená čára – jádrový odhad hustoty vypočtený na základě dat naznačuje, že se jedná o výrazně bimodální data.

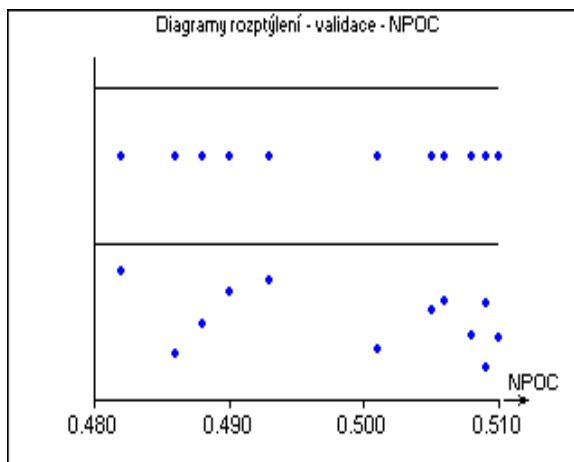
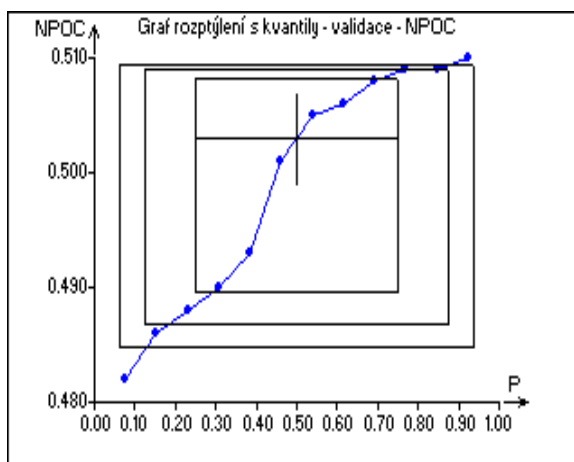


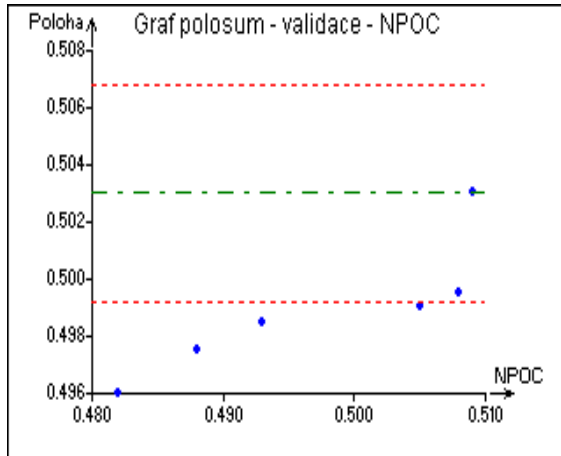
Diagram rozptýlení

Absence dat kolem střední hodnoty naznačuje, že se jedná o bimodální data.



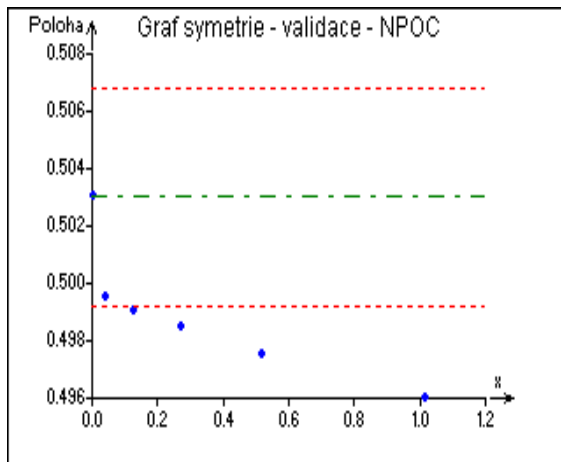
Graf rozptýlení s kvantily

Vzájemná poloha obdélníků nedokazuje symetrické rozdělení (zešikmení směrem k vyšším hodnotám). Jsou indikovány body vně sedcílových obdélníků.



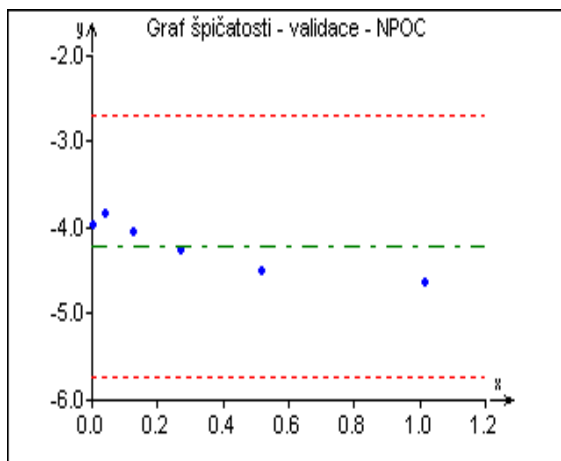
Graf polosum

Body neoscilují kolem horizontální přímky.



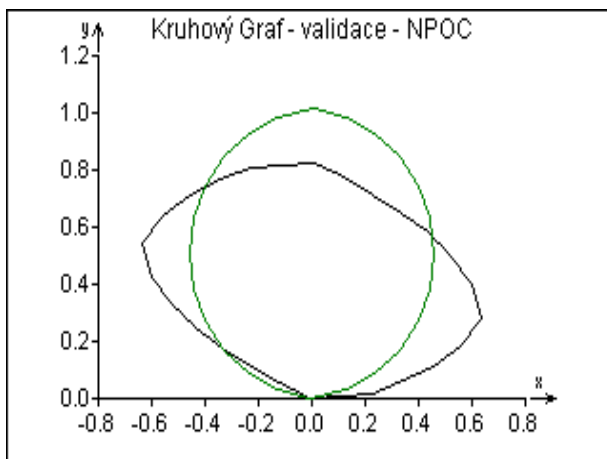
Graf symetrie

Data jsou uspořádána v mírně klesajícím trendu, který indikuje kladnou šikmost, tzn. mírné zešikmení směrem k vyšším hodnotám.



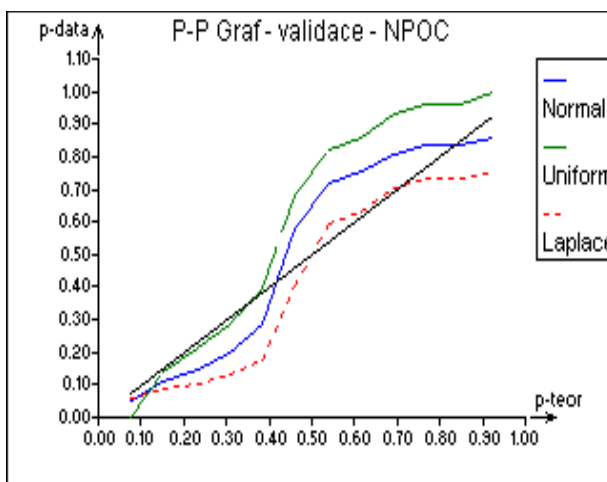
Graf špičatosti

Body leží na rovnoběžce s osou x, což ukazuje na normální rozdělení.



Kruhový graf

Zelený kruh je optimální tvar pro normální rozdělení. Kruh vytvořený z posuzovaných dat se nekryje s optimálním kruhem, ukazuje na možnou asymetrii dat.



Pravděpodobnostní graf

Empirická křivka nejvíce souhlasí s křivkou symetrického rozdělení

Závěr EDA: Nelze s jistotou prohlásit, že rozdělení je symetrické a normální, přestože je mu nejbližší. Rozložení dat v jednotlivých grafech ukazuje možnost rovnoměrného rozdělení. Podezřelé hodnoty nelze vzhledem k charakteru experimentu vyloučit, protože by to vedlo ke ztrátě dat.

2.4.2 Ověření předpokladů o datech

Základní analýza dat

Název úlohy :	Validace
Data:	Všechna
Řád trendu :	4
Testovaná hodnota :	0
Vyhlazení hustoty :	0,5
Hladina významnosti :	0,05
Název sloupce :	NPOC
Počet platných dat :	12

(1) Klasické parametry :

Název sloupce :	NPOC
Průměr :	0,4989166667
Spodní mez :	0,4923167225

Statistická analýza jednorozměrných dat

Horní mez :	0,5055166109
Rozptyl :	0,0001079015152
Směr. odchylka :	0,01038756541
Šikmost	-0,3780206553
Odchylka od 0 :	Nevýznamná
Špičatost :	1,515858302
Odchylka od 3 :	Nevýznamná
Polosuma	0,496
Modus :	0,5099102564

(2) Robustní parametry :

Název sloupce :	NPOC
Medián :	0,503
IS spodní :	0,4912087915
IS horní :	0,5147912085
Mediánová směr. odchylka :	0,005357241259
Mediánový rozptyl :	2,870003391E-005
10% Průměr :	0,4995
10% IS spodní :	0,4917892229
10% IS horní :	0,5072107771
10% Směr. odchylka :	0,008397029297
10% Rozptyl :	7,051010101E-005
20% Průměr :	0,5
20% IS spodní :	0,4902760411
20% IS horní :	0,5097239589
20% Směr. odchylka :	0,006755469165
20% Rozptyl :	4,563636364E-005
40% Průměr :	0,5005
40% IS spodní :	0,4875085971
40% IS horní :	0,5134914029
40% Směr. odchylka :	0,005017016498
40% Rozptyl :	2,517045455E-005

(3) Znaménkový test :

Závěr : Data jsou nezávislá

(4) Analýza malých výběrů

N :	12
Střední hodnota :	0,4985
Spodní mez (5%) :	0,489617
Horní mez (95%) :	0,507383
Spodní mez (2.5%) :	0,488357
Horní mez (97.5%) :	0,508643
Pivotové rozpětí :	0,021

(5) Test normality :

Název sloupce :	NPOC
Průměr :	0,4989166667
Rozptyl :	0,0001079015152
Šikmost	-0,3780206553
Špičatost :	1,515858302
Normalita :	Přijata
Vypočtený :	0,6702748089
Teoretický :	5,991464547
Pravděpodobnost :	0,7152398025

(6) Vybočující body :

Název sloupce : NPOC
 Homogenita : Přijata
 Počet vybočujících bodů : 0
 Spodní mez : 0,449
 Horní mez : 0,547

(7) Autokorelace :

Řád autokorelace : 4
 Název sloupce : NPOC
 Počet : 0,2059725245

Řád autokorelace 1
 Korelační koeficient : -0,112482155
 Pravděpodobnost : 0,3709735462
 Závěr : Nevýznamný

Řád autokorelace 2
 Korelační koeficient : -0,5010999555
 Pravděpodobnost : 0,07005020907
 Závěr : Nevýznamný

Řád autokorelace 3
 Korelační koeficient : 0,109229095
 Pravděpodobnost : 0,3898404273
 Závěr : Nevýznamný

Řád autokorelace 4
 Korelační koeficient : 0,2059725245
 Pravděpodobnost : 0,3122926973
 Závěr : Nevýznamný

2.5. Srovnání výsledků

K určení parametrů polohy a rozptýlení byla použita Hornova metoda pivotů spolu s klasickými a robustními odhady polohy a rozptýlení určenými pomocí softwaru QC Expert.

Tabulka 01

Hornova metoda pivotů		QC Expert – analýza malých výběrů		QC Expert - klasické odhady		QC Expert robustní odhady	
P_L	0,4985	P_L	0,4985	\bar{x}	0,4989	$\tilde{x}_{0,5}$	0,503
R_L	0,021	R_L	0,021	s	0,0104	$s(\tilde{x}_{0,5})$	0,0054
L_D	0,4884	L_D	0,4884	L_D	0,4923	L_D	0,4912
L_H	0,5086	L_H	0,5086	L_H	0,5055	L_H	0,5148

2.6. Hodnocení

Data byla zpracována dle Hornova postupu pro malé výběry, a to jak pomocí pivotů, tak pomocí programu QC Expert. Z Tabulky 01 vyplývá, že oběma použitými způsoby bylo dosaženo stejných výsledků.

Použití EDA pro takto malý soubor dat se ukázalo jako nevhodné. Z grafických výstupů nebylo možné jednoznačně určit druh rozdělení a rozhodnout, které parametry polohy a rozptýlení nejlépe charakterizují analyzovaná data. Z Tabulky 01 je patrné, že při použití klasických odhadů získáme velice úzký interval spolehlivosti a robustní odhady posunují střední hodnotu i intervaly spolehlivosti směrem k vyšším hodnotám, jak již bylo patrné z grafických výstupů v rámci EDA.

2.7. Závěr

Průměrný obsah TOC u kontrolního vzorku roztoku sacharózy je 0,4895 mg TOC/l s R_L 0,021; $IS_D = 0,4884$ a $IS_H = 0,5086$.

Správná hodnota obsahu TOC v kontrolním vzorku je 0,50 mg TOC/l, spadá do intervalu spolehlivosti měření; způsob stanovení TOC je **VYHOVUJÍCÍ**.

3. Statistické testování

3.1. Test shody středních hodnot (testování shodnosti)

3.1.1 Zadání

Porovnání dvou výběrů pomocí testu shodnosti s použitím softwaru QC Expert na hladině významnosti $\alpha = 0,05$.

V laboratoři Biochemické a chemické kontroly v rámci validace nové analytické metody vznikl požadavek na statistickou analýzu dat získaných během validace. Byl stanovován obsah celkového organického uhlíku u kontrolního vzorku (roztok sacharózy s koncentrací 0,50 mg TOC/l). Stanovení bylo prováděno na dvou TOC analyzátořech, které pracují každý na jiném principu. V obou případech byla použita metoda NPOC.

3.1.2 Data

Obsah TOC [mg/l]
 $n_1 = 6$; $n_2 = 6$

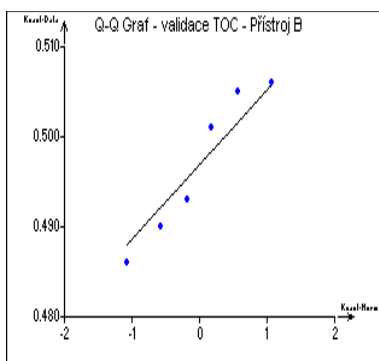
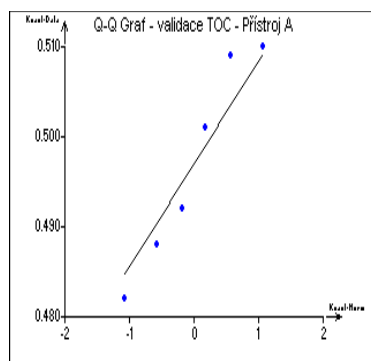
Přístroj A	Přístroj B
0,510	0,490
0,501	0,486
0,508	0,506
0,482	0,509
0,509	0,488
0,505	0,493

3.1.3 Program

QC Expert 2.27

Základní statistika, Porovnání dvou výběrů

3.1.4 EDA a ověření základních předpokladů o datech



Kvantilově - kvantilový graf

Průměr (kolmice vztyčená v 0) se jeví jako shodný. Nejsou diagnostikovány podezřelé hodnoty.

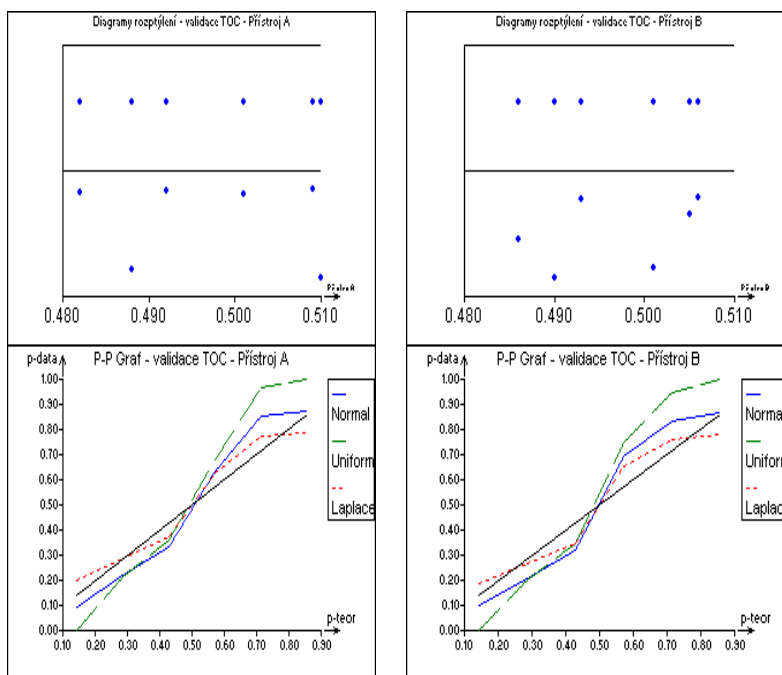


Diagram rozptýlení

Absence dat kolem střední hodnoty naznačuje, že se v obou případech jedná o bimodální data.

Pravděpodobnostní graf

Empirická křivka u obou rozdělení nejvíce souhlasí s křivkou symetrického rozdělení

Základní analýza dat

Název úlohy :	Validace TOC	
Data:	Všechna	
Řád trendu :	4	
Testovaná hodnota :	0,5	
Vyhlazení hustoty :	0,5	
Hladina významnosti :	0,05	
Název sloupce :	Přístroj A	Přístroj B
Počet platných dat :	6	6

(1) Klasické parametry :		
Název sloupce :	Přístroj A	Přístroj B
Průměr :	0,497	0,4968333333
Spodní mez :	0,4849429024	0,4880929351
Horní mez :	0,5090570976	0,5055737316
Rozptyl :	0,000132	6,936666667E-005
Směr. odchylka :	0,01148912529	0,008328665359
Šikmost	-0,03467136688	-0,09842937092
Odchylka od 0 :	Nevýznamná	Nevýznamná
Špičatost :	1,478842975	1,395595859
Odchylka od 3 :	Nevýznamná	Nevýznamná
Polosuma	0,496	0,496
Modus :	0,4957857143	0,4972380952

(2) Robustní parametry :		
Název sloupce :	Přístroj A	Přístroj B
Medián :	0,4965	0,497
IS spodní :	0,4781383639	0,4838845456
IS horní :	0,5148616361	0,5101154544
Med. směr. odchylka :	0,007142988346	0,005102134533
Mediánový rozptyl :	5,102228251E-005	2,603177679E-005
10% Průměr :	0,4975	0,49725

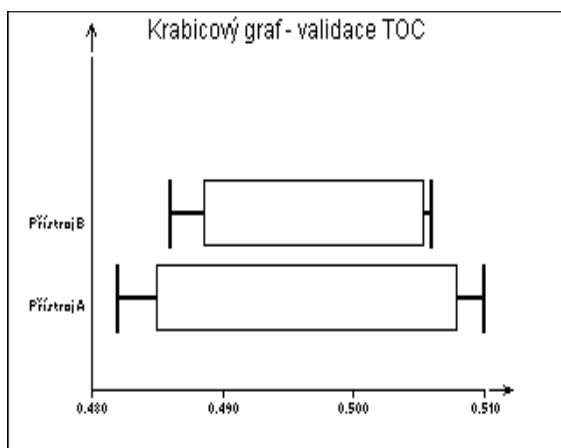
Statistická analýza jednorozměrných dat

10% IS spodní :	0,4772296795	0,482512681
10% IS horní :	0,5177703205	0,511987319
10% Směr. odchylka :	0,007286212246	0,005381036662
10% Rozptyl :	5,308888889E-005	2,895555556E-005
(3) Znaménkový test :		
Závěr :	Data jsou nezávislá	Data jsou nezávislá
(4) Analýza malých výběrů:		
N :	6	6
Střední hodnota :	0,4985	0,4975
Spodní mez (5%) :	0,482561	0,486115
Horní mez (95%) :	0,514439	0,508885
Spodní mez (2.5%) :	0,476765	0,481975
Horní mez (97.5%) :	0,520235	0,513025
Pivotové rozpětí :	0,021	0,015
(5) Test normality :		
Název sloupce :	Přístroj A	Přístroj B
Průměr :	0,497	0,4968333333
Rozptyl :	0,000132	6,936666667E-005
Šikmost :	-0,03467136688	-0,09842937092
Špičatost :	1,478842975	1,395595859
Normalita :	Přijata	Přijata
Vypočtený :	0,01390306673	0,03936480515
Teoretický :	5,991464547	5,991464547
Pravděpodobnost :	0,9930725727	0,9805100313
(6) Vybočující body :		
Název sloupce :	Přístroj A	Přístroj B
Homogenita :	Přijata	Přijata
Počet vyboč. bodů :	0	0
Spodní mez :	0,45065	0,46125
Horní mez :	0,53235	0,52575

3.1.5 Hodnocení základních předpokladů o datech

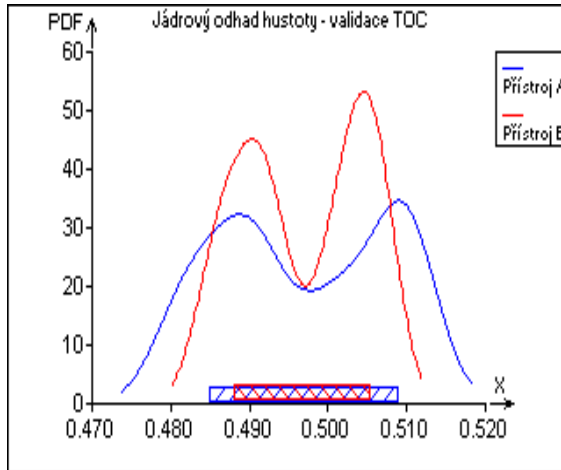
Z ověření základních předpokladů pro jednotlivé výběry vyplývá, že data v obou výběrech jsou nezávislá, homogenní, bez odlehklých bodů. Test normality u obou výběrů prokázal Gaussovo rozdělení.

3.1.6 Porovnání dvou výběrů



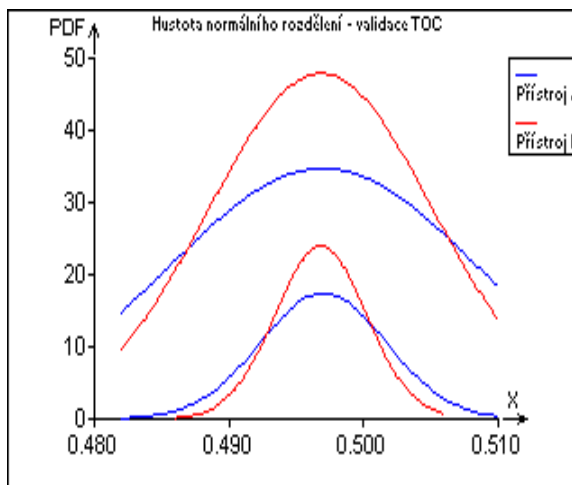
Krabicový graf

Nejsou diagnostikovány podezřelé hodnoty ležící mimo interval vnitřních hradeb.



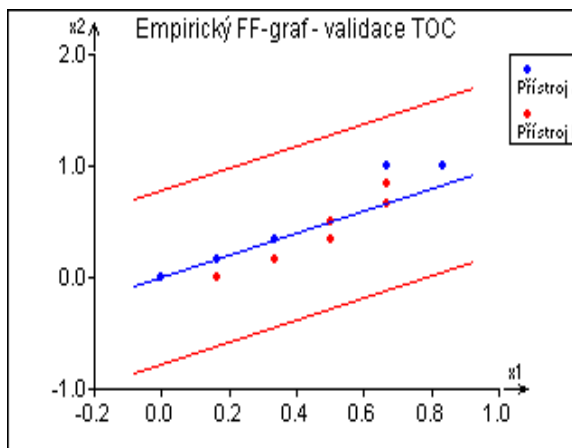
Jádrový odhad hustoty

Oba soubory jsou bimodální, intervaly spolehlivosti se překrývají, průměry jsou statisticky shodné na hladině významnosti $\alpha = 0,05$.



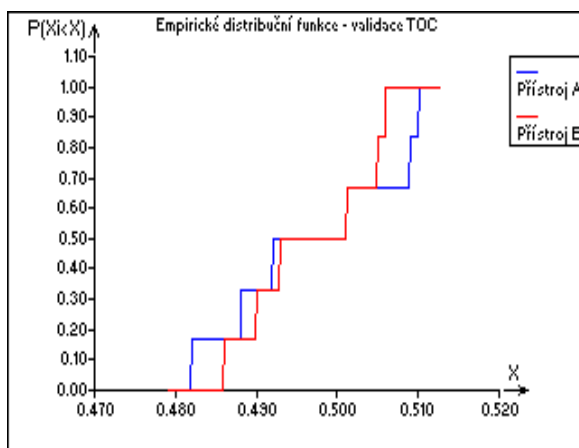
Hustota normálního rozdělení

Gausova křivka odpovídá průměru a rozptylu obou přístrojů.



Empirický F-F graf

Data obou výběrů leží na centrální přímce, takže oba výběry jsou shodné.



Empirické distribuční funkce
Distribuční funkce obou rozdělení jsou shodné

Porovnání dvou výběrů

Název úlohy :	Validace TOC	
Data:	Všechna	
Hladina významnosti :	0,05	
Porovnávané sloupce :	Přístroj A	Přístroj B
Počet dat :	6	6
Průměr :	0,497	0,4968333333
Směr. odchylka :	0,01148912529	0,008328665359
Rozptyl :	0,000132	6,936666667E-005
Korel. koef. R(x,y) :	-0,2257310869	

Test shody rozptylů

Poměr rozptylů :	1,902931283	
Počet stupňů volnosti :	5	5
Kritická hodnota :	5,050329058	
Závěr :	Rozptyly jsou SHODNÉ, H₀ přijata	
Pravděpodobnost :	0,2485639994	

Test shody průměrů pro SHODNÉ rozptyly

t-statistika :	0,02876938541
Počet stupňů volnosti :	10
Kritická hodnota :	2,228138852
Závěr :	Průměry jsou SHODNÉ
Pravděpodobnost :	0,9776145786

Test dobré shody rozdělení dvouvýběrový K-S test

Diference DF :	0,3333333333
Kritická hodnota :	0,7841002757
Závěr :	Rozdělení jsou SHODNÁ

3.1.7 Hodnocení

Oboustranný klasický test potvrdil shodu středních hodnot na hladině významnosti $\alpha = 0,05$.

3.1.8 Závěr

Na obou testovaných přístrojích bylo dosaženo stejných výsledků se stejnou variabilitou.

3.2. Test střední hodnoty (testování správnosti)

3.2.1 Zadání

Testování střední hodnoty pomocí testu správnosti s použitím softwaru QC Expert na hladině významnosti $\alpha = 0,05$.

V laboratoři Biochemické a chemické kontroly v rámci validace nové analytické metody vznikl požadavek na statistickou analýzu dat získaných během validace. Byl stanovován obsah celkového organického uhlíku u tří roztoků sacharózy s koncentracemi 0,25 mg TOC/l, 0,50 mg TOC/l a 1,0 mg TOC/l. Stanovení bylo prováděno na TOC analyzátoru metodou NPOC.

3.2.2 Data

Obsah TOC [mg/l]

$n_1 = 9$; $n_2 = 9$; $n_3 = 9$

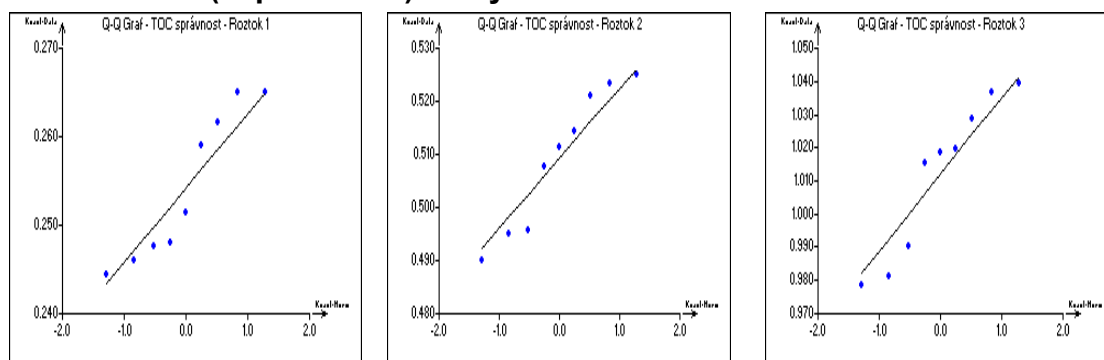
Roztok 1	Roztok 2	Roztok 3
0,2650	0,5211	1,0285
0,2445	0,4951	1,0365
0,2649	0,5115	0,9905
0,2590	0,4899	1,0155
0,2476	0,5235	1,0185
0,2514	0,5143	1,0195
0,2459	0,5076	0,9785
0,2479	0,4956	0,9814
0,2615	0,5249	1,0395

3.2.3 Program

QC Expert 2.27

Základní statistika

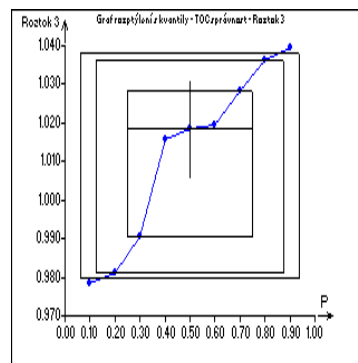
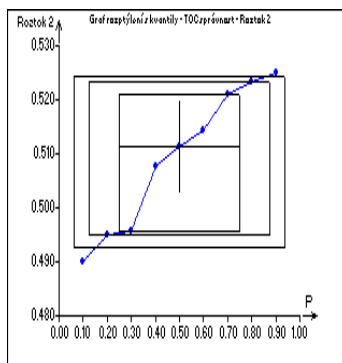
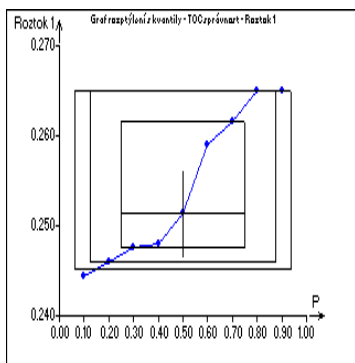
3.2.4 Průzkumová (exploratorní) analýza dat EDA



Kvantilově - kvantilový graf

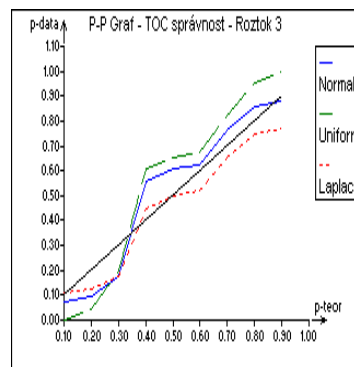
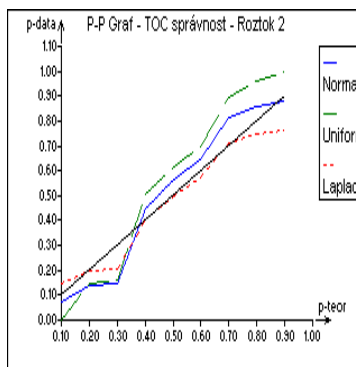
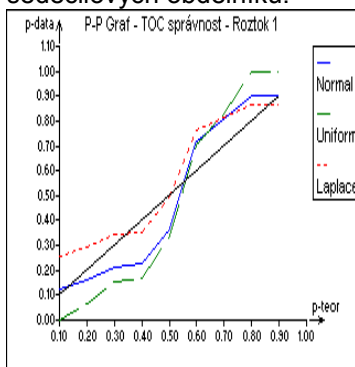
Rozložení dat naznačuje normální rozdělení s absencí dat kolem střední hodnoty u roztoků 1 a 3.

Nejsou diagnostikovány podezřelé hodnoty.



Graf rozptýlení s kvantily

Vzájemná poloha obdélníků dokazuje symetrické rozdělení. Jsou indikovány body vně sedciových obdélníků.



Pravděpodobnostní graf

Empirická křivka u obou rozdělení nejvíce souhlasí s křivkou symetrického rozdělení

3.2.5 Hodnocení EDA

Z ověření základních předpokladů pro jednotlivé výběry vyplývá, že data ve všech výběrech jsou nezávislá, homogenní, bez odlehlých bodů.

3.2.6 Základní analýza dat

Název úlohy : TOC správnost
Data: Všechna

Řád trendu : 4
Testovaná hodnota : 1
Vyhlazení hustoty : 0,5
Hladina významnosti : 0,05
Název sloupce : Roztok 1
Počet platných dat : 9

Roztok 2
9

Roztok 3
9

(1) Klasické parametry :	Roztok 1	Roztok 2	Roztok 3
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Průměr :	0,2541888889	0,5092777778	1,012044444
Spodní mez :	0,2477489776	0,49917244	0,9943059778
Horní mez :	0,2606288002	0,5193831156	1,029782911
Rozptyl :	7,019111111E-005	0,0001728319444	0,0005325428
Směr. odchylka :	0,008378013554	0,01314655637	0,0230769
Šikmost	0,2162415039	-0,2456808018	-0,38859498
Odchylka od 0 :	Nevýznamná	Nevýznamná	Nevýznamná

Statistická analýza jednorozměrných dat

Špičatost :	1,348901636	1,586902116	1,669874819
Odchylna od 3 :	Nevýznamná	Nevýznamná	Nevýznamná
Polosuma	0,25475	0,5074	1,009
Modus :	0,2469377778	0,5150555556	1,028828889
(2) t-test			
Testovaná hodnota :	0,25	0,50	1,00
Rozdíl :	Nevýznamný	Nevýznamný	Nevýznamný
Vypočtený :	1,499957787	2,117157721	1,565780131
Teoretický :	2,306004135	2,306004135	2,306004135
Pravděpodobnost :	0,08600699757	0,03356073597	0,07801637
Konfidenční interval levý:	0,2489957827	0,5011288934	0,9977402503
Konfidenční interval pravý:	0,2593819951	0,5174266621	1,026348639
(3) Robustní parametry :			
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Medián :	0,2514	0,5115	1,0185
IS spodní :	0,2402227338	0,4947929285	0,9860859281
IS horní :	0,2625772662	0,5282070715	1,050914072
Med. směr. odchylka :	0,004847028	0,00724503	0,01405638
Medianový rozptyl :	2,349367855E-005	5,249047472E-005	0,0001975818
10% Průměr :	0,2540285714	0,5098142857	1,012914286
10% IS spodní :	0,2453034307	0,4969283598	0,9892795153
10% IS horní :	0,2627537122	0,5227002117	1,036549056
10% Směr. odchylka :	0,006626126739	0,009754358	0,017253409
10% Rozptyl :	4,390555556E-005	9,51475E-005	0,0002976801
20% Průměr :	0,2540285714	0,5098142857	1,012914286
20% IS spodní :	0,2453034307	0,4969283598	0,9892795153
20% IS horní :	0,2627537122	0,5227002117	1,036549056
20% Směr. odchylka :	0,006626126739	0,009754358	0,017253409
20% Rozptyl :	4,390555556E-005	9,51475E-005	0,0002976801
(4) Znaménkový test :			
Závěr :	Data jsou nezávislá	Data jsou nezávislá	Data jsou nezávislá
(5) Analýza malých výběrů			
N :	9	9	9
Střední hodnota :	0,25455	0,50835	1,0095
Spodní mez (5%) :	0,2449868	0,490806	0,983356
Horní mez (95%) :	0,2641132	0,525894	1,035644
Spodní mez (2.5%) :	0,2418315	0,4850175	0,97473
Horní mez (97.5%) :	0,2672685	0,5316825	1,04427
Pivotové rozpětí :	0,0139	0,0255	0,03800
(6) Test normality :			
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Průměr :	0,2541888889	0,5092777778	1,012044444
Rozptyl :	7,019111111E-005	0,0001728319444	0,0005325428
Šikmost	0,2162415039	-0,2456808018	-0,38859498
Špičatost :	1,348901636	1,586902116	1,669874819
Normalita :	Přijata	Přijata	Přijata
Vypočtený :	0,2108177519	0,2623514901	0,5891873473
Teoretický :	5,991464547	5,991464547	5,991464547
Pravděpodobnost :	0,8999564768	0,8770636213	0,7448341717
(7) Vybočující body :			
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Homogenita :	Přijata	Přijata	Přijata
Počet vybočujících bodů :	0	0	0

Statistická analýza jednorozměrných dat

Spodní mez :	0,221665	0,45958	0,910915
Horní mez :	0,283235	0,54982	1,089985
(8) Autokorelace :			
Řád autokorelace :	4		
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Počet :	-0,3765146039	0,4197962232	0,2541117951
(9) Řád autokorelace 1			
Korelační koeficient :	-0,2511933909	-0,4668580143	-0,045058399
Pravděpodobnost :	0,2742248225	0,1217589044	0,4578148917
Závěr :	Nevýznamný	Nevýznamný	Nevýznamný
(10) Test významnosti trendu :			
Název sloupce :	Roztok 1	Roztok 2	Roztok 3
Směrnice :	-0,0008233333333	0,000555	-0,002355
Významnost :	Nevýznamný	Nevýznamný	Nevýznamný
Pravděpodobnost :	0,7581233867	0,616459893	0,7667879337

3.2.7 Hodnocení základních předpokladů o datech

Z ověření základních předpokladů pro jednotlivé výběry vyplývá, že data ve všech výběrech jsou nezávislá, homogenní, bez odlehlých bodů. Test normality prokázal Gaussovo rozdělení.

3.2.8 Celkové hodnocení

Pro 95 % statistickou jistotu byly nalezeny následující intervalové odhady:

Roztok 1 – testovaná koncentrace 0,25 mg TOC/l

Interval pro \bar{x} $0,248 < \mu < 0,261$

Interval pro $\tilde{x}_{0,5}$ $0,240 < \mu < 0,263$

Interval pro P_L $0,242 < \mu < 0,267$

Roztok 2 – testovaná koncentrace 0,50 mg TOC/l

Interval pro \bar{x} $0,499 < \mu < 0,519$

Interval pro $\tilde{x}_{0,5}$ $0,495 < \mu < 0,528$

Interval pro P_L $0,485 < \mu < 0,532$

Roztok 3 – testovaná koncentrace 1,00 mg TOC/l

Interval pro \bar{x} $0,994 < \mu < 1,030$

Interval pro $\tilde{x}_{0,5}$ $0,986 < \mu < 1,051$

Interval pro P_L $0,975 < \mu < 1,044$

3.2.9 Závěr

Z uvedených intervalových odhadů vyplývá, že správné hodnoty všech tří testovaných koncentrací 0,25 mg TOC/l, 0,50 mg TOC/l a 1,0 mg TOC/l leží uvnitř intervalových odhadů. Testovaná metoda poskytuje v rozmezí 0,25 až 1,00 mg TOC/l správné výsledky.

3.3. Párový test

3.3.1 Zadání

Porovnání dvou výběrů pomocí párového testu s použitím softwaru QC Expert na hladině významnosti $\alpha = 0,05$.

V laboratoři Biochemické a chemické kontroly v rámci validace nové analytické metody vznikl požadavek na statistickou analýzu dat získaných během validace. Byl stanovován obsah celkového organického uhlíku u šesti různých vzorků. Stanovení bylo prováděno na dvou TOC analyzátořech, které pracují každý na jiném principu. V obou případech byla použita metoda NPOC.

3.3.2 Data

Obsah TOC [mg/l]

$n_1 = 6$; $n_2 = 6$

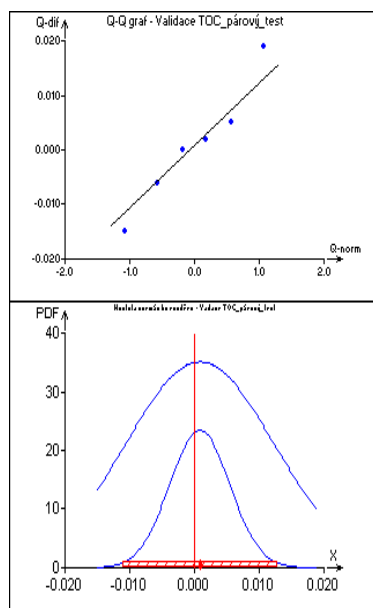
Přístroj A	Přístroj B
0,508	0,490
0,503	0,501
0,024	0,019
0,222	0,222
0,253	0,259
1,473	1,488

3.3.3 Program

QC Expert 2.27

Základní statistika, Porovnání dvou výběrů, Párové porovnání

3.3.4 Ověření základních předpokladů o datech

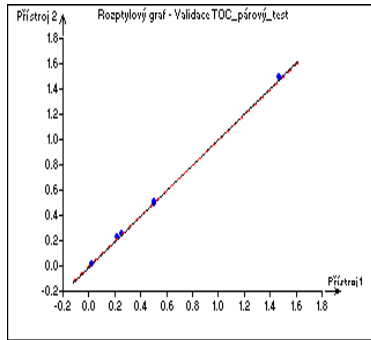


Q-Q graf

Body leží přibližně na přímce a lze tedy předpokládat normální rozdělení. Spolehlivost použitého testu by měla být dobrá.

Hustota normálního rozdělení

Nula leží v intervalu spolehlivosti průměru rozdílu, rozdíl je tedy statisticky nevýznamně odlišný od nuly.



Rozptylový graf

Přímky, které reprezentují skutečnou závislost dat z obou přístrojů a závislost odpovídající nevýznamnému rozdílu jsou shodné.

Závěr EDA: Pro malý počet dat nelze s jistotou prohlásit, že rozdělení je symetrické a normální, přestože je mu nejbližší.

Párové porovnání dvou výběrů

Hypotézy : $H_0: d = \mu_1 - \mu_2 = 0$
 $H_A: d \neq 0$

Název úlohy : Validace TOC_párový_test

Hladina významnosti : 0,05
 Porovnávané sloupce : Přístroj 1 Přístroj 2

(1) Analýza difference X - Y

Počet dat : 6
 Průměrná diference : 0,000833333333333
 Interval spolehlivosti: -0,01110289184 0,01276955851
 Směr. odchylka : 0,01137394684
 Rozptyl : 0,0001293666667

Korel. koef. R(x,y) : 0,9998256228

(2) Test významnosti rozdílu

t-statistika : 0,002041241452
 Počet stupňů volnosti : 5
 Kritická hodnota : 2,570581836
 Závěr : **Rozdíly jsou NEVÝZNAMNÉ, H_0 přijata**
 Pravděpodobnost : 0,4323088295

3.3.5 Hodnocení

Párový test potvrdil shodu středních hodnot na hladině významnosti $\alpha = 0,05$.

3.3.6 Závěr

Na obou testovaných přístrojích bylo dosaženo stejných výsledků.