

UNIVERZITA PARDUBICE
Fakulta chemicko-technologická

Katedra analytické chemie

Nám. Čs. Legií 565, 532 10 Pardubice

15. licenční studium
INTERAKTIVNÍ STATISTICKÁ ANALÝZA DAT

Semestrální práce

ANOVA

2015

Vedoucí studia a odborný garant: Prof. RNDr. Milan Meloun, DrSc.

Vyučující: Prof. RNDr. Milan Meloun, DrSc.

Duben 2015

Autor práce: Mgr. Veronika Pilařová

OBSAH

ÚLOHA 1: JEDNOFAKTOROVÁ ANOVA	3
Zadání	3
Data	3
Program	3
Řešení	3
Závěr	6
ÚLOHA 2: DVOUFAKTOROVÁ ANOVA BEZ OPAKOVÁNÍ	7
Zadání	7
Data	7
Program	7
Řešení	7
Hodnocení	11
Závěr	12
ÚLOHA 3: DVOUFAKTOROVÁ ANOVA S OPAKOVÁNÍM	13
Zadání	13
Data	13
Program	13
Řešení	13
Hodnocení	17
Závěr	17

ÚLOHA 1: JEDNOFAKTOROVÁ ANOVA

Zadání

Na 3 kolonách stejného typu, stejného sorbentu (BEH 2-EP) a různé šarže (FAKTOR A) byl proměřen standard látky *ibuprofenum*, pro ověření stejné účinnosti kolon. Analýza byla provedena za stejných podmínek se stejným roztokem standardu, na každou kolonu byl vzorek nastříknut dvanáctkrát. Na základě ploch píků určete, zda je rozptyl a odezva píků pro testované kolony stejná.

Data

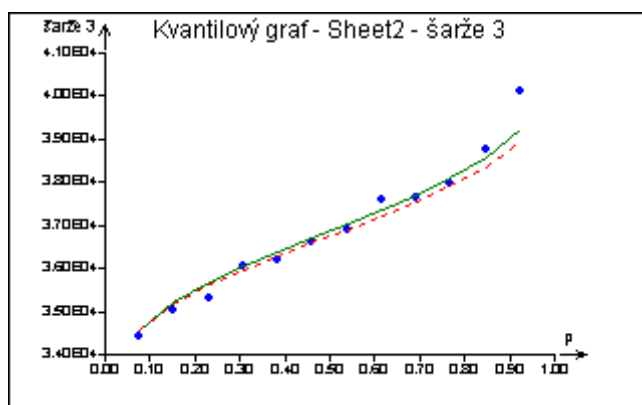
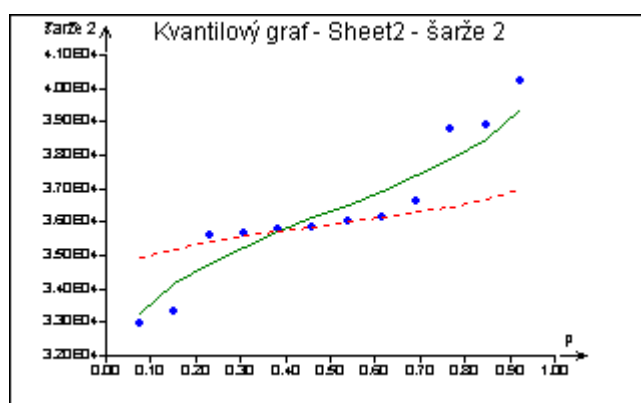
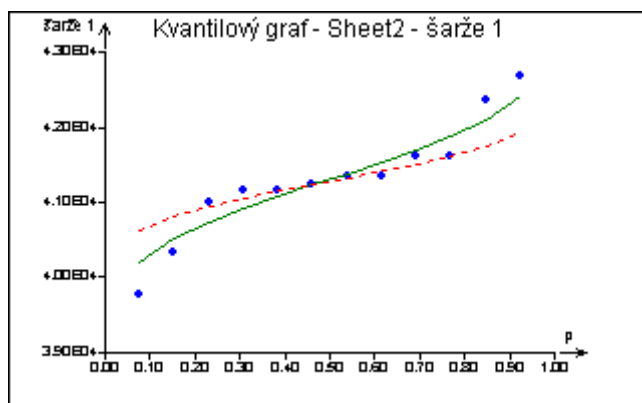
Šarže kolony	Inj. 1	Inj. 2	Inj. 3	Inj. 4	Inj. 5	Inj. 6	Inj. 7	Inj. 8	Inj. 9	Inj. 10	Inj. 11	Inj. 12
0108332731	42368	42685	41165	40337	41231	41621	41355	41340	41000	41625	39767	41162
0108341561	40236	35627	33342	35783	35682	35842	36021	36143	36591	38901	38765	32946
0108341341	38759	36033	35292	36211	35024	34440	36621	37607	40118	36888	37954	37649

Program

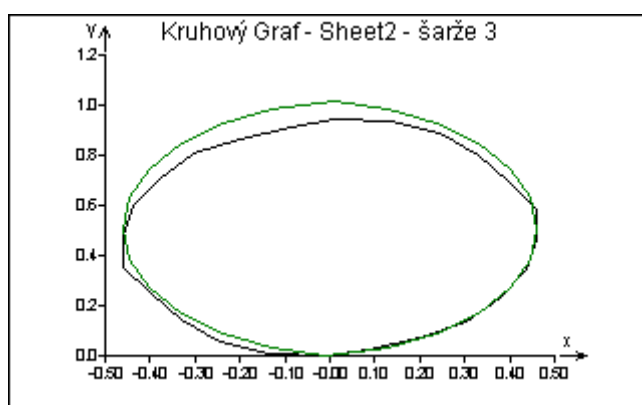
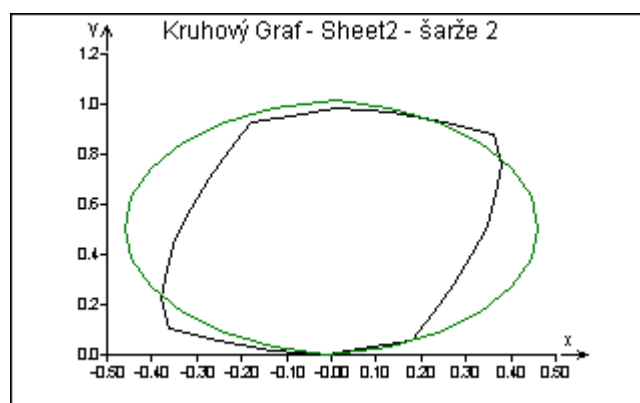
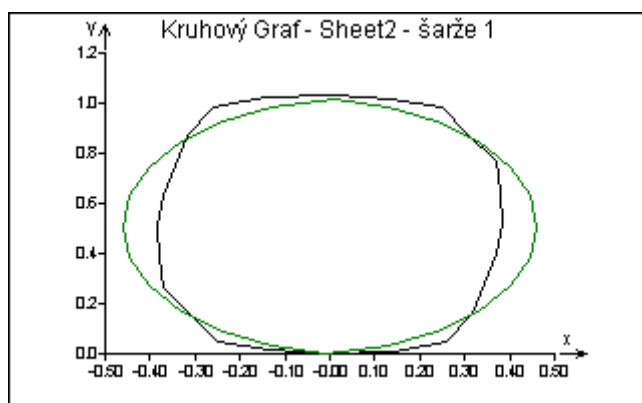
QC.Expert 3.3/Jednofaktorová ANOVA...

Řešení

Kvantilový graf pro odezvy při 3 měřeních na 3 kolonách jiné šarže



Kruhový graf pro odezvy při 3 měřeních na 3 kolonách jiné šarže



Klasické parametry:

Sloupce:	šarže 1	šarže 2	šarže 3
Průměr:	41304,66667	36323,25	36883
Spodní mez:	40809,487	34971,00732	35843,54787
Horní mez:	41799,84633	37675,49268	37922,45213
Rozptyl:	607396,9697	4529562,932	2676430,727
Směr. odchylka:	779,3567666	2128,276987	1635,980051
Šikmost:	-0,1362229546	0,2530112563	0,3588291829
Odchylka od 0:	Nevýznamná	Nevýznamná	Nevýznamná
Špičatost:	3,066070039	2,531568419	2,471873656
Odchylka od 3:	Nevýznamná	Nevýznamná	Nevýznamná

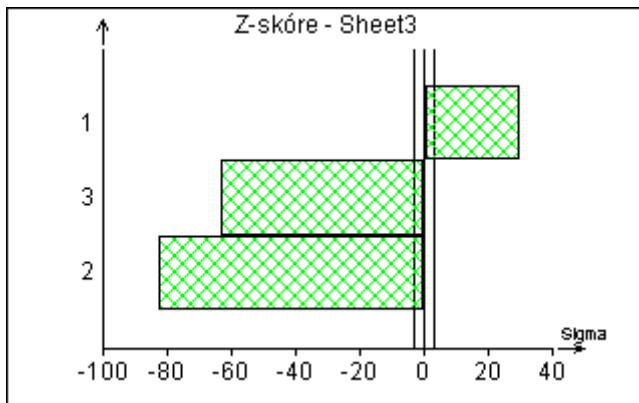
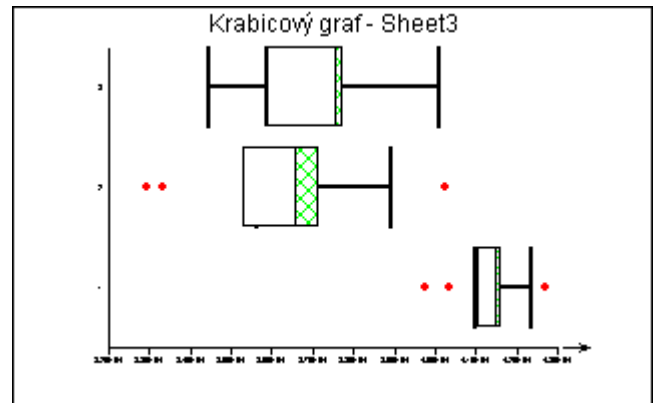
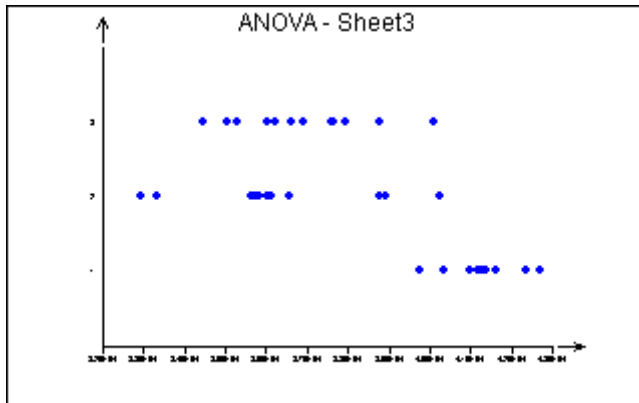
Znaménkový test:

Závěr:	Data jsou nezávislá	Data jsou závislá	Data jsou závislá
--------	---------------------	-------------------	-------------------

Test normality:

Sloupce:	šarže 1	šarže 2	šarže 3
Normalita:	Přijata	Přijata	Přijata
Testové kritérium:	0,1339126435	0,3378336656	0,6142807346
Kritický kvantil $\chi^2(22)$:	5,991464547	5,991464547	5,991464547
p-hodnota:	0,9352360499	0,8445791418	0,7355473469

Grafické výstupy - ANOVA



Všechny použité šarže vykazují normální rozdělení dat. Podle krabicového grafu obsahují 2 sloupce odlehle hodnoty.

Celkový průměr: 38170,30556

Celkový rozptyl: 7561804,847

Průměrný čtverec: 7351754,712

Reziduální rozptyl: 2604463,543

Reziduální součet čtverců: 85947296,92

Celkový součet čtverců: 264663169,6

Vysvětlený součet čtverců: 178715872,7

Počet úrovní faktoru: 3

Sloupec	Počet hodnot	Efekty faktorů	Průměr úrovně
1	12	3134,361111111111	41304,6666666667
2	12	-1847,055555555555	36323,25
3	12	-1287,305555555555	36883

Test významnosti celkového vlivu faktoru:

Závěr	Teoretický	Vypočítaný	Pravděpodobnost
Významný	3,284917651	34,30953626	8,717435974E-009

Párové porovnávání dvojic úrovní

Scheffého metoda

<i>Srovnávaná dvojice</i>	<i>Rozdíl</i>	<i>Významnost</i>	<i>Pravděpodobnost</i>
1 - 2	4981,416667	Významný	6,269953415E-008
1 - 3	421,666667	Významný	6,794891037E-007
2 - 3	-559,75	Nevýznamný	0,6997632718

Závěr

Test významnosti vlivu šarže kolony na odezvu látky vykazuje statistickou závislost. Vypočítaná hodnota F kritéria je 34,309, což je hodnota vyšší než kritická hodnota 3,285. Hypotéza H_0 o shodě středních hodnot odezvy pro jednotlivé šarže je zamítnuta. Vícenásobné porovnání Scheffého metodou nám ukazuje, že 2 dvojice mají hypotézu H_0 zamítnutu ($H_0: \mu_i - \mu_j = 0$). Jedná se o dvojice, ve kterých je porovnávána kolona č. 1.

ÚLOHA 2: DVOUFAKTOROVÁ ANOVA BEZ OPAKOVÁNÍ

Zadání

Na 4 stejných kolonách (rozměry, sorbent BEH, typ) uchovávaných ve 4 různých rozpouštědlech (oxid uhličitý, isopropanol, acetonitril a hexan) byl po dobu půl roku každý měsíc měřen standard paracetamolu. Byl sledován vliv času (FAKTOR B) a činidla, ve kterém byla kolona uchovávána (FAKTOR A), na retenční čas zvoleného standardu. Který z faktorů ovlivňuje více změnu retenčního času?

Data

	Měsíc 1	Měsíc 2	Měsíc 3	Měsíc 4	Měsíc 5	Měsíc 6
<i>oxid uhličitý</i>	0,71	0,63	0,65	0,63	0,65	0,73
<i>isopropanol</i>	0,66	0,62	0,64	0,65	0,72	0,76
<i>acetonitril</i>	0,66	0,64	0,68	0,66	0,70	0,79
<i>hexan</i>	0,65	0,61	0,63	0,60	0,65	0,79

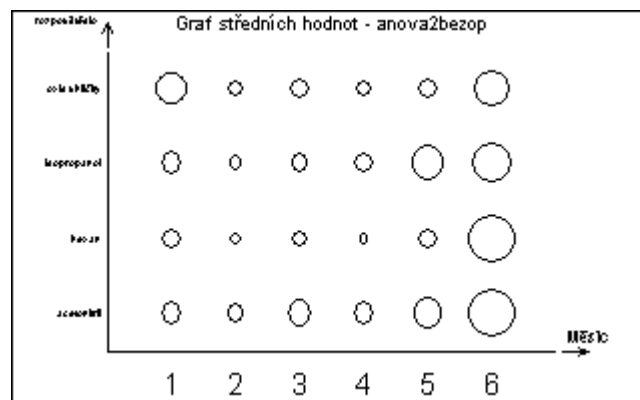
Program

QC.Expert 3.3, ANOVA - Dvofaktorová...

Řešení

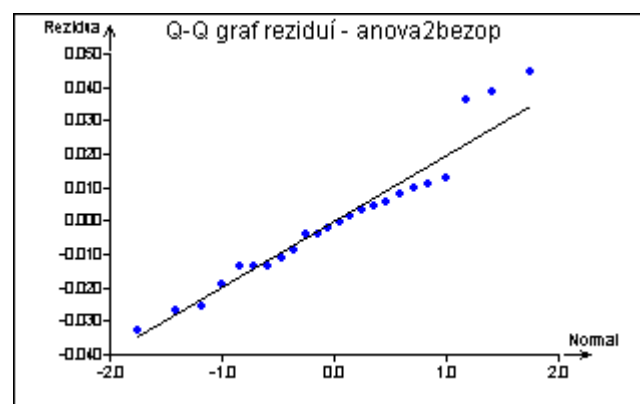
Graf středních hodnot

Z grafu jsou patrné rozdíly mezi hodnotami v jednotlivých celách. Statistická významnost však musí být potvrzena v matematických výstupech.



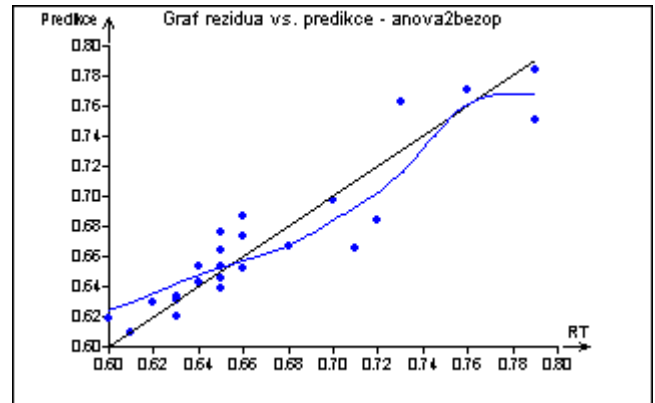
Q-Q graf reziduí

Z grafu jsou patrné 3 vybočující hodnoty. Ostatní body vykazují normální rozdělení, které kopíruje černou přímkou.



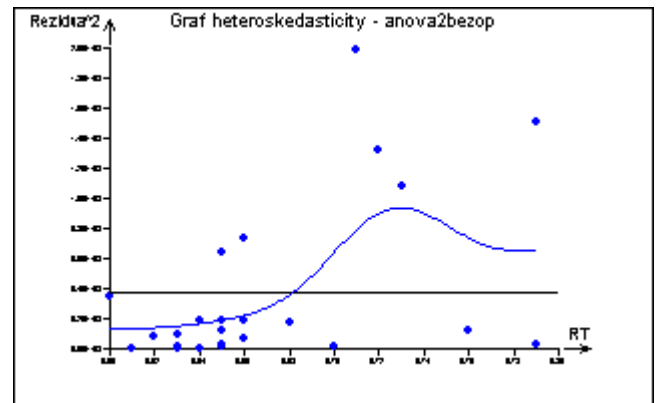
Graf rezidua vs. Predikce

Modrá křivka, znázorňující přibližný průběh střední hodnoty predikce, přimyká k černé přímce ($x = y$). Anova model by měl pracovat efektivně a dobře.



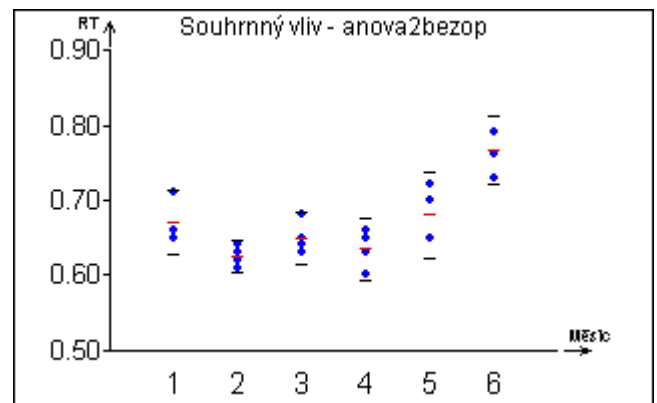
Graf heteroskedasticity

Modrá křivka nekopíruje křivku černou. Data vykazují nekonzstantnost rozptylu.



Souhrnný vliv faktoru 1 (čas)

Graf vyjadřuje souhrnný vliv faktoru na odezvu. Červeně vyznačené hodnoty značí významný vliv faktoru.



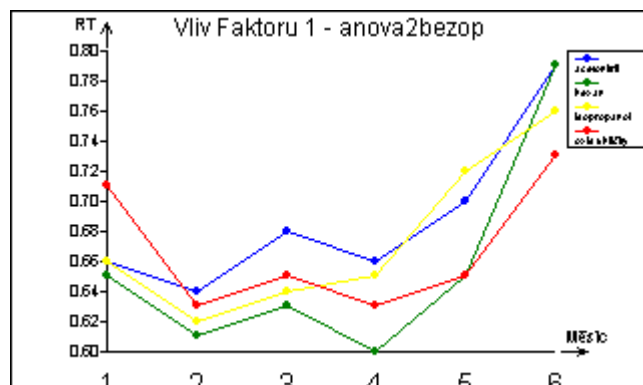
Souhrnný vliv faktoru 2 (medium)

Graf vyjadřuje souhrnný vliv faktoru na odezvu. Červeně vyznačené hodnoty značí významný vliv faktoru. Faktor nemá významný vliv.



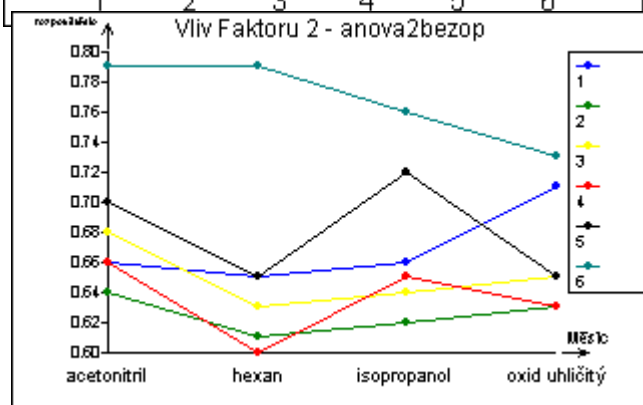
Vliv faktoru 1 (uchovací médium)

Křivky mají obdobný průběh, což značí vliv faktoru na odezvu.



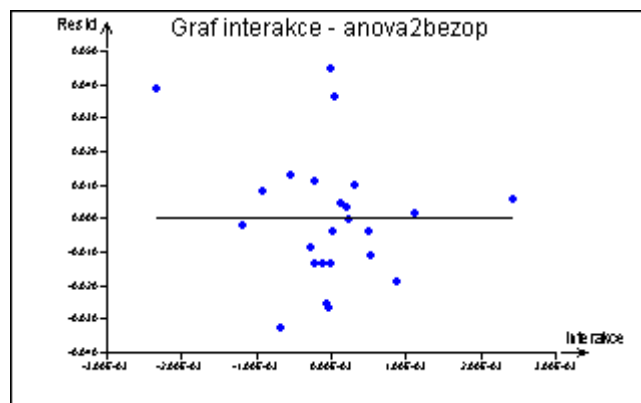
Vliv faktoru 2 (čas)

Podle náhodného uspořádání křivek nemá faktor výrazný vliv na odezvu.



Graf vlivu interakce

Body nevykazují rostoucí tendenci, interakce ve statisticky posuzovaném modelu je nevýznamná. Toto potvrzuje i barva přímk (černá), která by v případě statistické významnosti interakce byla červená.



Analýza rozptylu

První faktor: Měsíc

Druhý faktor: rozpouštědlo

Odezva: RT

Typ modelu: vyvážený bez opakování

Faktory úrovně:

Měsíc: 1 2 3 4 5 6

Uchovací médium: acetonitril hexan isopropanol oxid uhličitý

Počet opakování 1

Tabulka středních hodnot RT

Úrovně	acetonitril	hexan	isopropanol	oxid uhličitý
1	0,66	0,65	0,66	0,71
2	0,64	0,61	0,62	0,63
3	0,68	0,63	0,64	0,65
4	0,66	0,6	0,65	0,63
5	0,7	0,65	0,72	0,65
6	0,79	0,79	0,76	0,73

Průměry pro faktor: Měsíc

Úroveň	1	2	3	4	5	6
Průměr	0,67	0,625	0,65	0,635	0,68	0,7675

Průměry pro faktor: rozpouštědlo

Úroveň	acetonitril	hexan	isopropanol	oxid uhličitý
Průměr	0,6883333333	0,655	0,675	0,6666666667

Celkový průměr : 0,67125

Parametry modelu

Faktor	Úroveň faktoru	Příspěvek úrovně
Měsíc	1	-0,00125
Měsíc	2	-0,04625
Měsíc	3	-0,02125
Měsíc	4	-0,03625
Měsíc	5	0,00875
Měsíc	6	0,09625

rozpouštědlo	acetonitril	0,01708333333
rozpouštědlo	hexan	-0,01625
rozpouštědlo	isopropanol	0,00375
rozpouštědlo	oxid uhličitý	-0,004583333333

Tabulka ANOVA

Zdroj variability	Součet čtverců	průměrný čtverec	Stupně volnosti	Směr. odch.	F-kritérium	Kritický kvantil	Závěr	p-hodnota
Měsíc	0,0529875	0,0105975	5	0,102944 1596	17,80261316	2,901294536	Významný	7,813467246E-006
rozpouštědlo	0,0035458 33333	0,0011819444 44	3	0,034379 41891	1,985534298	3,287382104	Nevýznamný	0,1594821631

Interakce	0,0003410 577602	0,0003410577 602	1	0,094494 2679	0,5559790514	4,600109937	Nevýznamný	0,4682200873
Rezidua	0,0085881 08906	0,0006134363 505	14	0,024767 64725				
Celkem	0,0654625	0,0028461956 52	23	0,053349 74838				

Hodnocení

Faktor B (čas)

Hypotézy: H_0 : efekty faktoru A jsou nulové.
 H_A : efekty faktoru A nejsou nulové.

Kritický kvantil: $F = 2,901$
F-kritérium: $F_e = 17,802$
p-hodnota: $7,813 \text{ E-}006$

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 17,802$ má hodnotu vyšší než kritický kvantil, je nulová hypotéza zamítnuta. Efekty faktoru B nejsou nulové a faktor A je statisticky významný na hladině $\alpha = 0,05$.

Faktor A (médium)

Hypotézy: H_0 : efekty faktoru B jsou nulové.
 H_A : efekty faktoru B nejsou nulové.

Kritický kvantil: $F = 3,287$
F-kritérium: $F_e = 1,985$
p-hodnota: $0,159$

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 1,985$ má hodnotu nižší než kritický kvantil, je nulová hypotéza přijata. Efekty faktoru A jsou nulové a faktor A je statisticky nevýznamný na hladině $\alpha = 0,05$.

Interakce

Hypotézy: H_0 : efekty interakce faktorů jsou nulové.
 H_A : efekty interakce faktorů nejsou nulové.

Kritický kvantil: $F = 4,600$
F-kritérium: $F_e = 0,556$
p-hodnota: $0,468$

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 0,556$ má hodnotu nižší než kritický kvantil, je nulová hypotéza přijata. Efekty interakce faktoru A a B jsou nulové a interakce je statisticky nevýznamná na hladině $\alpha = 0,05$.

Závěr

Posun retenčního času je ovlivněn především faktorem B (časem, po který byla kolona uchováвана). Vliv média, ve kterém byly kolony uchováваны (faktor A), na posun retenčního času nemá statistický vliv. Interakce obou faktorů nemá významný statistický vliv na sledovanou odezvu.

ÚLOHA 3: DVOUFAKTOROVÁ ANOVA S OPAKOVÁNÍM

Zadání

Na kolonách typu BEH a BEH 2-EP (faktor A) uchovávaných v různých rozpouštědlech – oxid uhličitý, isopropanol, acetonitril, hexan (faktor B), byl změřen standard ketoprofenu. Měření bylo provedeno po 6 měsících od zahájení testování a nástřik na každou kolonu byl opakován třikrát. Vyšetřete, zda retenční čas ovlivňuje sorbent, nebo rozpouštědlo, ve kterém je kolona uchována.

Data

ketoprofen	nástřik	oxid uhličitý	isopropanol	acetonitril	hexan
BEH	1	1,379	1,419	1,482	1,454
	2	1,381	1,42	1,481	1,454
	3	1,375	1,42	1,484	1,454
BEH 2-EP	1	1,377	1,26	1,255	1,165
	2	1,372	1,26	1,253	1,165
	3	1,373	1,261	1,252	1,165

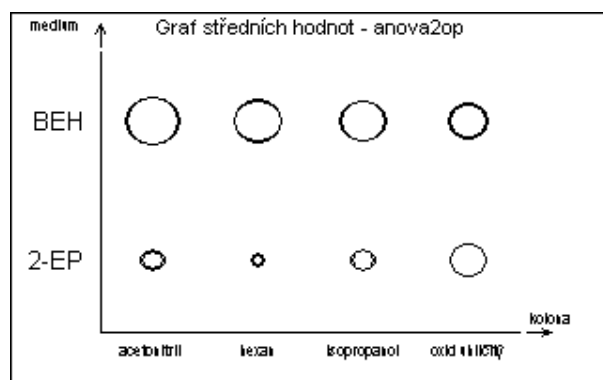
Program

QC.Expert 3.3, ANOVA - Dvoufaktorová...

Řešení

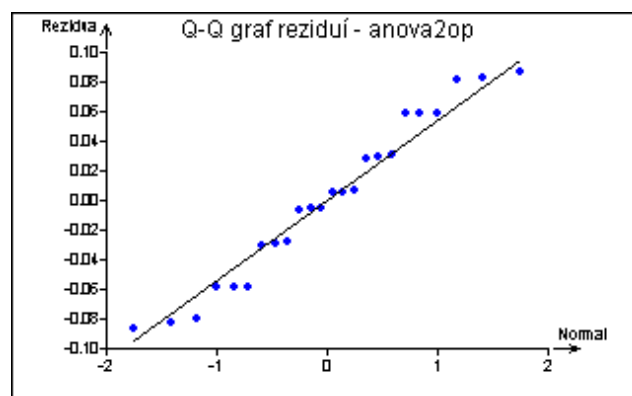
Graf středních hodnot

Jsou patrné rozdíly v celách. Statistickou významnost musí potvrdit matematické postupy.



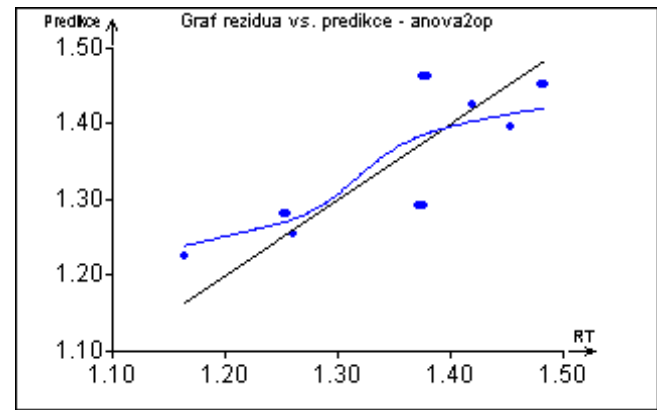
Q-Q graf reziduí

Body leží přibližně okolo černé přímky, nejsou přítomny vybočující hodnoty, rezidua jsou normálně rozložena.



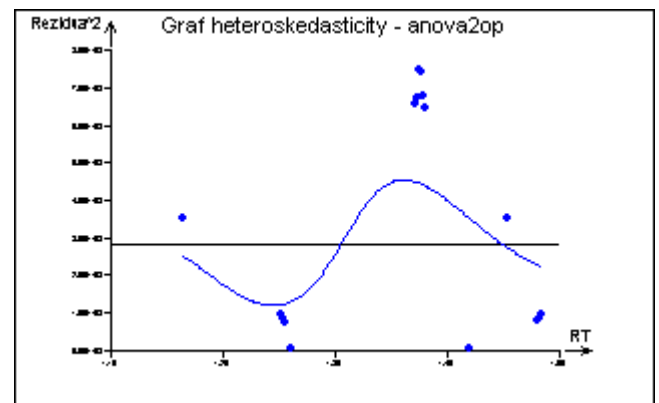
Graf rezidua vs. predikce

Modrá křivka, která znázorňuje přibližný průběh střední hodnoty predikce, se částečně přibližuje černé přímce. Vliv faktoru by mohl mít význam z hlediska statistiky.



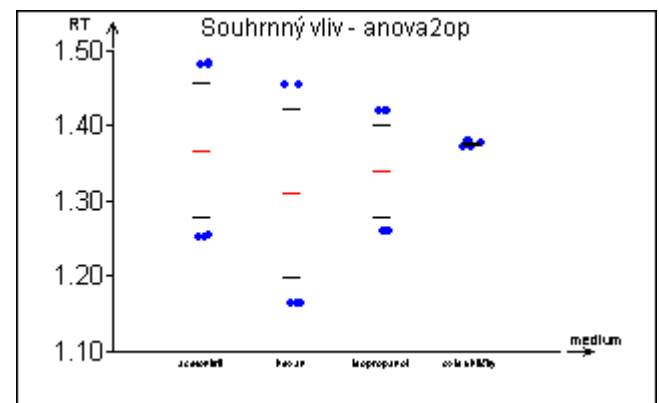
Graf heteroskedasticity

Data nevykazují konstantnost rozptylu, což může ovlivnit spolehlivost výsledku.



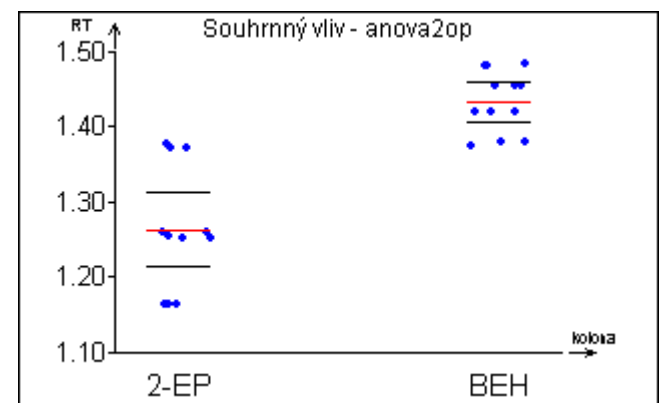
Souhrnný vliv faktoru A/1 (kolona)

Je zde vyjádřen souhrnný vliv faktoru na odezvu. 3 hodnoty jsou vyznačeny červeně – mají významný faktor.



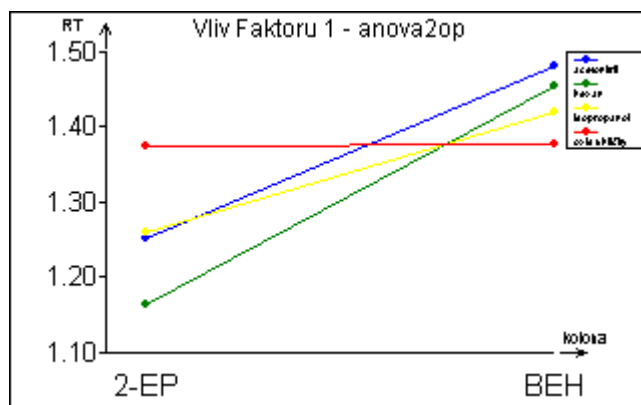
Souhrnný vliv faktoru B/2 (medium)

Je zde vyjádřen souhrnný vliv faktoru na odezvu. 2 hodnoty jsou vyznačeny červeně – mají významný faktor.



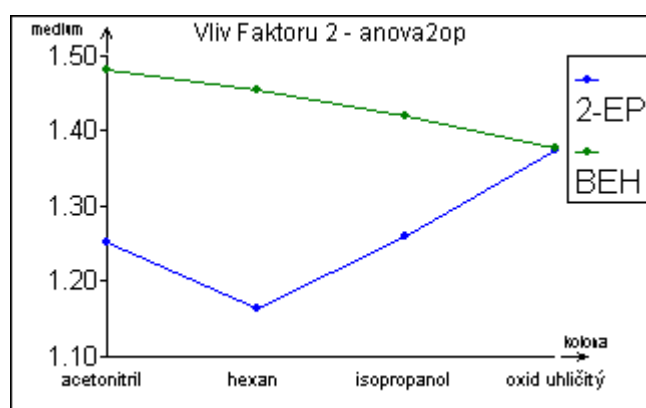
Vliv faktoru A/1

Křivky vykazují podobný, lehce chaotický průběh. Faktor může být ze statistického hlediska významný.



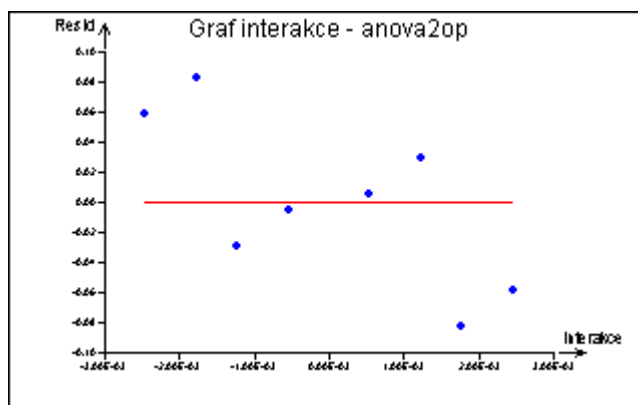
Vliv faktoru B/2

Křivky vykazují chaotický průběh. Faktor je ze statistického hlediska nevýznamný.



Graf interakce

Přímka nemá vzestupnou tendenci, ale má červenou barvu, což znamená, že interakce faktorů by mohla být statisticky významná.



Analýza rozptylu

První faktor : kolona

Druhý faktor : medium

Odezva : RT

Typ modelu: vyvážený s opakováním

Faktory úrovně:

Kolona: 2-EP BEH

Medium: acetonitril hexan isopropanol oxid uhlíčitý

Počet opakování 3

Tabulka středních hodnot RT

Úroveň	acetonitril	hexan	isopropanol	oxid uhličitý
2-EP	1,253333333	1,165	1,260333333	1,374
BEH	1,482333333	1,454	1,419666667	1,378333333

Průměry pro faktor: kolona

Úroveň:	2-EP	BEH
Průměr:	1,263166667	1,433583333

Průměry pro faktor: medium

Úroveň:	acetonitril	hexan	isopropanol	oxid uhličitý
Průměr:	1,367833333	1,3095	1,34	1,376166667

Celkový průměr: 1,348375

Parametry modelu

Faktor	Úroveň faktoru	Příspěvek úrovně
kolona	2-EP	-0,08520833333
kolona	BEH	0,08520833333
medium	acetonitril	0,01945833333
medium	hexan	-0,038875
medium	isopropanol	-0,008375
medium	oxid uhličitý	0,02779166667

Tabulka ANOVA

Zdroj variability	Součet čtverců	Průměrný čtverec	Stupně volnosti	Směr. odch.	F-kritérium	Kritický kvantil	Závěr	p-hodnota
kolona	0,1742510417	0,1742510417	1	0,417433877	64338,84615	4,493998478	Významný	2,867189651E-030
medium	0,01639445833	0,005464819444	3	0,07392441711	2017,779487	3,238871517	Významný	7,777985039E-021
Interakce	0,06780079167	0,02260026389	3	0,1503338415	8344,712821	3,238871517	Významný	9,244917342E-026
Rezidua	4,333333333E-005	2,708333333E-006	16	0,001645701472				
Celkem	0,258489625	0,01123867935	23	0,1060126377				

Hodnocení

Faktor A (kolona)

Hypotézy:	H ₀ : efekty faktoru A jsou nulové. H _A : efekty faktoru A nejsou nulové.
Kritický kvantil:	F = 4,494
F-kritérium:	F _e = 64338,846
p-hodnota:	2,867E-030

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 64338,846$ má hodnotu vyšší než kritický kvantil, je nulová hypotéza zamítnuta. Efekty faktoru A nejsou nulové a faktor A je statisticky významný na hladině $\alpha = 0,05$.

Faktor B (médiu)

Hypotézy:	H ₀ : efekty faktoru B jsou nulové. H _A : efekty faktoru B nejsou nulové.
Kritický kvantil:	F = 3,239
F-kritérium:	F _e = 2017,779
p-hodnota:	7,778E-021

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 2017,779$ má hodnotu vyšší než kritický kvantil, je nulová hypotéza zamítnuta. Efekty faktoru B jsou nenulové a faktor B je statisticky významný na hladině $\alpha = 0,05$.

Interakce

Hypotézy:	H ₀ : efekty interakce faktorů jsou nulové. H _B : efekty interakce faktorů nejsou nulové.
Kritický kvantil:	F = 3,239
F-kritérium:	F _e = 8344,712
p-hodnota:	9,245E-026

Protože Fischerovo-Snedecerovo testační kritérium $F_e = 8344,712$ má hodnotu vyšší než kritický kvantil, je nulová hypotéza zamítnuta. Efekty interakce faktorů A a B nejsou nulové a interakce je statisticky významná na hladině $\alpha = 0,05$.

Závěr

Na změnu retenčního času má významný vliv sorbent i použité uchovávací médium.