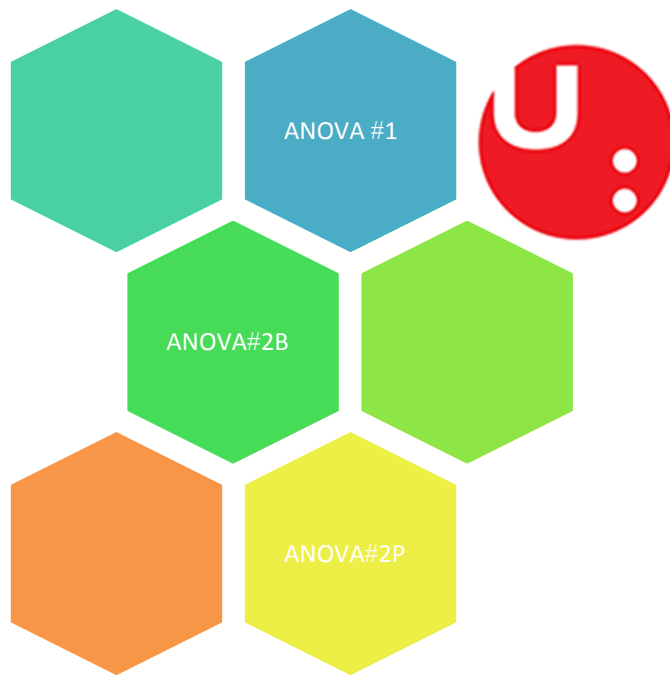


Licenční studium Galileo: Statistické zpracování dat



Analýza rozptylu ANOVA

Semestrální práce

Lenka Husáková

Pardubice 2015

Obsah

1	Jednofaktorová ANOVA.....	3
1.1	Zadání	3
1.2	Data	3
1.3	Použitý software.....	3
1.4	Ověření předpokladů o datech a posouzení významu transformace	3
1.5	Průměry a efekty úrovní faktoru	4
1.6	Detekce vlivných bodů	4
1.7	Tabulka ANOVA	5
1.8	Závěr	5
1.9	Literatura	6
2	Dvoufaktorová ANOVA bez opakování.....	7
2.1	Zadání	7
2.2	Data	7
2.3	Použitý software.....	7
2.4	Ověření předpokladů o datech a posouzení významu transformace	8
2.5	Průměry a efekty úrovní.....	9
2.6	Tabulka ANOVA	9
2.7	Vícenásobné porovnání úrovní jednotlivých faktorů	12
2.8	Volba optimálních úrovní jednotlivých faktorů.....	15
2.9	Závěr	16
2.10	Literatura	16
3	Dvoufaktorová ANOVA s opakováním.....	17
3.1	Zadání	17
3.2	Data	17
3.3	Použitý software.....	17
3.4	Ověření základních předpokladů o datech.....	17
3.5	Transformace dat	18
3.6	Průměry a efekty úrovní.....	19
3.7	Tabulka ANOVA	19
3.8	Vícenásobné porovnání úrovní jednotlivých faktorů	20
3.9	Závěr	21
3.10	Literatura	21

1 Jednofaktorová ANOVA

1.1 Zadání

Vliv doby lyofilizace na zajištění sušiny ve vzorku jater

S ohledem na nutnost uchování unikátního biologického vzorku a pro potřeby budoucího stanovení vybraných elementů v sušině, byla pro konzervaci a přípravu vzorku vydřích jater k analýze zvolena lyofilizace. Optimalizace doby lyofilizace byla provedena na vzorcích vepřových jater. Pro tento účel bylo lyofilizováno 25 vzorků s navážkou 200 mg (± 1 mg) po dobu 12, 24 a 48 hodin. Po uplynutí každého z intervalů lyofilizace byla stanovena sušina vzorku. Výsledky experimentu byly analyzovány pomocí jednofaktorové analýzy rozptylu.

Cílem bylo posoudit vliv doby lyofilizace na zajištění sušiny, konkrétně pak nalézt minimální dobu zajišťující celkové odstranění vlhkosti ze vzorku.

1.2 Data

Tabulka 1.1: Obsah sušiny (%) stanovený ve 25 vzorcích vepřových jater po lyofilizaci vzorku po dobu 12, 24 a 48 h (n = 75)

	Vzorek							
Faktor	1	2	3	23	24	25
A1 (12 h)	28.24	27.88	35.50	30.40	29.70	31.12
A2 (24 h)	24.84	27.71	32.56	30.28	29.28	31.00
A4 (48 h)	25.14	28.61	32.53	30.31	29.29	30.91

1.3 Použitý software

Naměřená data byla vyhodnocena pomocí statistického programu Adstat 2.0 (TriloByte Statistical Software, s.r.o., ČR) a NCSS 2007 (NCSS, Statistical Software, USA).

1.4 Ověření předpokladů o datech a posouzení významu transformace

Jelikož je analýza rozptylu založena na předpokladu, že každý z výběrů pochází z populace s normálním rozdělením se stejnou směrodatnou odchylkou a že máme náhodný výběr z každé populace (skupiny) a všechna pozorování jsou nezávislá, bylo před vlastní ANOVA analýzou provedeno ověření základních předpokladů o datech. Na základě výsledků vyhodnocených programem NCSS 2007 a uvedených v tabulce 1.2 je zřejmé, že data vykazují normální rozdělení a mají shodné rozptyly.

Kromě výše uvedeného postupu bylo pro posouzení kvality dat využito programu Adstat 2.0, kde byl v modulu ANOVA#1 posuzován význam transformace vstupních dat. Jelikož hodnota korelačního koeficientu dosáhla hodnoty 0.768, byla provedena logaritmická transformace dat. Po provedené logaritmické transformaci se hodnota korelačního koeficientu výrazně nezměnila (0.780), transformace tedy nebyla nutná a další analýza byla provedena pro původní data.

Tab. 1.2: Testy výběrových předpokladů

Předpoklad	Testační kritérium	Spočtená hladina významnosti	Závěr testu (0.05) H_0
Test šikmosti reziduí	-0.6597	0.509	Přijata
Test špičatosti reziduí	0.6161	0.538	Přijata
Omnibus test reziduí	0.8148	0.665	Přijata
Modif. Levenův test homogenity rozptylů	0.0587	0.943	Přijata

1.5 Průměry a efekty úrovní faktoru

Hodnoty průměru a efekty úrovní faktoru doby lyofilizace na obsah sušiny ve vzorku jater shrnuje tabulka 1.3. Informace o hodnotách jednotlivých skupinových průměrů poskytují i krabicový graf a diagram průměrných hodnot na obrázku 1.1.

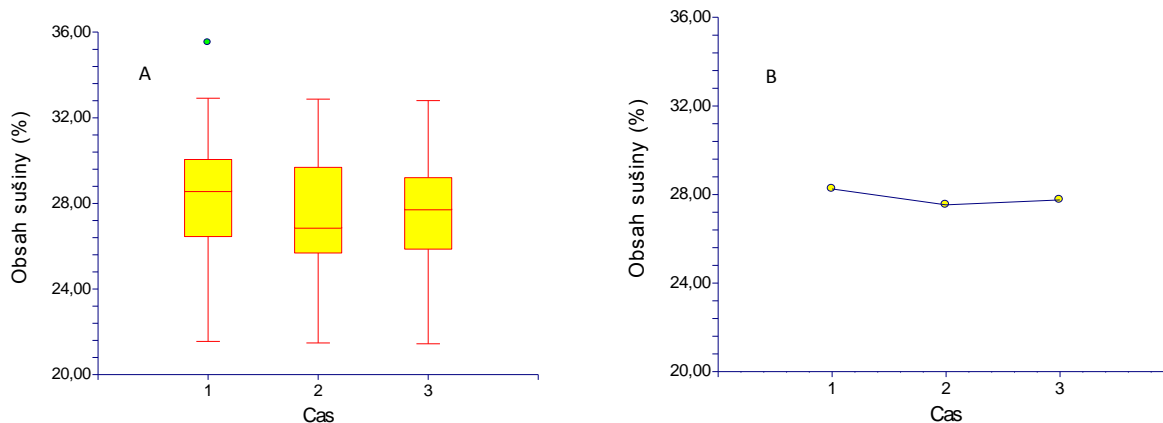
Tab. 1.3: Průměry obsahu sušiny (%) pro jednotlivé úrovně faktoru a efekty těchto úrovní

Úroveň	Počet hodnot	Průměr	Efekt
1	25	28.25	0.406
2	25	27.53	-0.314
3	25	27.75	-0.092

Celkový průměr 27.85, celkový rozptyl 9.27, $H_{ii} = 0.04$

1.6 Detekce vlivných bodů

V modulu ANOVA#1 v programu Adstat 2.0 byla indikována přítomnost jednoho odlehlého bodu. Přítomnost odlehlého bodu ve skupině dat první úrovně faktoru, tj. výsledků stanovení sušiny po 12 h lyofilizace, je patrná rovněž z krabicového grafu vyhodnoceného programem NCSS 2007.



Obr. 1.1 (A) Krabicový graf a (B) Diagram průměrů obsahu sušiny v závislosti na době lyofilizace

1.7 Tabulka ANOVA

Byl proveden F-test významnosti sledovaného faktoru. Jelikož Fischerovo-Snedecorovo testační kritérium $F_e = 0.365$ nabývá menší hodnoty než kvantil $F_{0.95}(2, 72) = 3.124$, je nulová hypotéza H_0 : Efekty faktoru A jsou nulové přijata a faktor A je statisticky nevýznamný (viz tab. 1.4).

Tab. 1.4: ANOVA tabulka testování vlivu doby lyofilizace na stanovení obsahu sušiny v játrech

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H_0	Spočtená α
Mezi úrovněmi	$k-1 = 2$	6.789	3.395	0.365	Přijata	0.695
Rezidua	$n-k = 72$	667.7	9.273			
Celkový	$n-1 = 74$	674.5	9.115			

$$F_{1-\alpha}(k-1, n-k) = 3.124$$

1.8 Závěr

Z výsledků jednofaktorové analýzy rozptylu vyplývá, že nulová hypotéza o shodě středních hodnot výběrů reprezentujících jednotlivé úrovně sledovaného faktoru času byla na hladině významnosti 0.05 přijata ($F_e < F_{0.95}(2, 72)$). Mezi hodnotami stanovení obsahu sušiny v závislosti na době lyofilizace vzorku jater nebyl shledán statisticky významný rozdíl. Pro lyofilizaci testovaného množství jater (200 mg) lze tedy použít nejkratší sledovanou dobu 12 h. Použití kratší doby lyofilizace zvýší výrazným způsobem výkon analytické metody (průchodnost vzorků) a sníží ekonomické náklady.

1.9 Literatura

1. Meloun M., Militký J.: Kompendium statistického zpracování dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2196-8.
2. Meloun M., Militký J.: Interaktivní statistická analýza dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2173-9.
3. Nováková L., Husáková L., Urbanová I., Šrámková J., Beran V., Sborník příspěvků Studentská vědecká odborná činnost 2013/2014, str. 143–148. Univerzita Pardubice, Pardubice 2014. ISBN: 978-80-7395-819-0.

2 Dvoufaktorová ANOVA bez opakování

2.1 Zadání

Vliv doby mletí a velikosti navážky na stanovení vybraných oxidů ve vzorcích identifikačních tělísek

Byla optimalizována metoda stanovení vybraných oxidů ve vzorcích identifikačních tělísek, sloužících k předvýbuchové i povýbuchové identifikaci průmyslových výbušnin, metodou atomové absorpční spektrometrie za využití přímého dávkování suspenzí.

Cílem bylo posoudit vliv doby mletí (faktor A) ovlivňující velikost částic a navážky vzorku (faktor B) na výsledky stanovení.

2.2 Data

Tab. 2.1: Výsledky stanovení vybraných oxidů (%) ve vzorcích identifikačních tělísek v závislosti na době mletí (min) a navážce (mg)

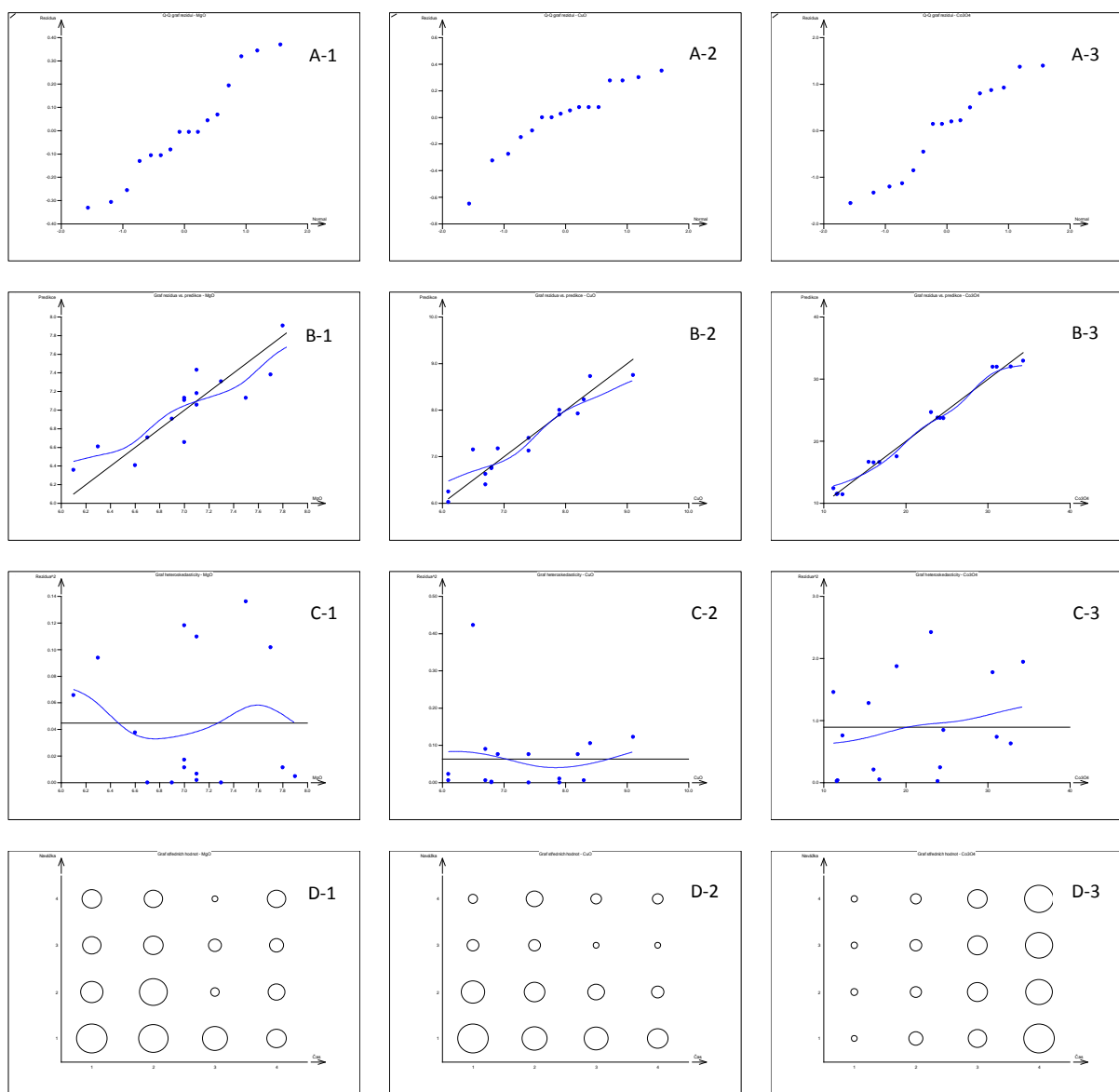
Faktor A	Faktor B	MgO	PbO	CuO	ZnO	Co ₃ O ₄
3 min	2 mg	7.9	7.9	9.1	8.6	11.2
	5 mg	7.3	8.5	8.2	7.9	12.3
	10 mg	7.0	8.1	6.8	7.3	11.6
	25 mg	7.1	8.4	6.5	7.3	11.7
5 min	2 mg	7.8	9.1	8.4	9.5	18.9
	5 mg	7.7	8.7	7.9	9.2	16.1
	10 mg	7.1	8.2	6.8	8.4	16.8
	25 mg	7.0	8.8	7.4	8.7	15.5
7 min	2 mg	7.5	9.2	8.3	9.6	23.1
	5 mg	6.3	8.4	7.4	8.9	24.6
	10 mg	6.6	7.7	6.1	8.2	24.2
	25 mg	6.1	8.3	6.7	8.6	23.9
10 min	2 mg	7.1	8.0	7.9	9.7	34.3
	5 mg	6.9	8.6	6.9	8.4	30.6
	10 mg	6.7	8.5	6.1	8.1	31.1
	25 mg	7.0	8.6	6.7	8.2	32.8

2.3 Použitý software

Data byla analyzována za využití programů Adstat 1.25 a QC-Expert 2.9 (TriloByte Statistical Software, s.r.o., ČR), dále pak programu NCSS 2007 (NCSS Statistical Software, USA) a Statistica 12 (StatSoft, Inc., USA).

2.4 Ověření předpokladů o datech a posouzení významu transformace

Pro posouzení základních předpokladů o datech bylo použito vybraných grafických diagnostik programu QC-Expert 2.9 v modulu dvoufaktorová ANOVA. Analýza byla provedena pro případ stanovení MgO, CuO a Co₃O₄. Data pro ostatní analyty jsou dostupná v příloze na CD. Na základě Q-Q grafu reziduí není patrný výrazný odklon experimentálních dat od normality. Z grafu rezidua vs. predikce, který vyjadřuje efektivitu modelu ANOVA, je zřejmé, že data leží přibližně na přímce $y = x$, lze tedy usuzovat na významnost modelu. Z grafu heteroskedasticity není patrná výrazná závislost velikosti chyby na pozorované hodnotě, data tedy vykazují homoskedasticitu.



Obr. 2.1 Q-Q graf reziduí (A), graf rezidua vs predikce (B), graf heteroskedasticity (C) a graf středních hodnot (D) pro původní data (1), data po Box-Coxově (2) a logaritmické (3) transformaci.

V programu Adstat 1.25 byl dále sledován vliv matematické transformace na kvalitu dat. Jelikož odhad exponentu mocninné transformace leží pro všechny studované případy v akceptovatelném intervalu, není třeba data transformovat (viz Tab. 2.2).

Na základě výše uvedených údajů byla analýza rozptylu provedena pro původní data uvedená v tabulce 2.1.

Tab. 2.2 Odhady exponentů mocninné transformace, jejich rozptylu a akceptovatelné intervaly pro data odpovídající stanovení vybraných oxidů kovů ve vzorcích IT v závislosti na době mletí a navážce

Analyt	Odhad mocninné transformace	Rozptyl odhadu transformace	Akceptovatelný interval	
			Dolní mez	Horní mez
MgO	3.40	26.8	-1.78	8.58
PbO	-12.8	629.0	-37.9	12.3
CuO	-0.70	6.35	-3.22	1.82
ZnO	2.01	2.88	0.31	3.70
Co ₃ O ₄	-1.23	4.48	-3.35	0.88

2.5 Průměry a efekty úrovní

Hodnoty průměrů a efekty úrovní faktoru doby mletí a faktoru velikosti navážky na stanovení vybraných oxidů kovů ve vzorcích identifikačních tělísek shrnuje tabulka 2.3. Uvedeny jsou i hodnoty celkového průměru, reziduálního rozptylu a Tukeyho konstanta. Informace o hodnotách jednotlivých skupinových průměrů pro stanovení MgO, CuO a Co₃O₄ poskytuje graf středních hodnot na obrázku 2.1-D. Hodnoty skupinových průměrů jednotlivých úrovní sledovaných faktorů pro všechny sledované analyty shrnuje obr. 2.3.

2.6 Tabulka ANOVA

Při posuzování významnosti jednotlivých faktorů a jejich vzájemné interakce na stanovení jednotlivých oxidů byly testovány následující hypotézy:

H₀: Efekty faktoru A jsou nulové, H_A: nejsou nulové; Kvantil $F_{1-\alpha}(n-1, mn-m-n) = 4.066$

H₀: Efekty faktoru B jsou nulové, H_A: nejsou nulové; Kvantil $F_{1-\alpha}(m-1, mn-m-n) = 4.066$

H₀: Interakce I je nulová, H_A: není nulová; Kvantil $F_{1-\alpha}(1, mn-m-n) = 5.318$

Pro případ kdy Fischerovo-Snedecorovo testační kritérium F_e nabývá větší hodnoty než kvantil $F_{0.95}(3, 8) = 4.066$, je nulová hypotéza H₀: Efekty faktoru A, resp. B jsou nulové zamítnuta a faktory A a B jsou statisticky významné. Jak je patrné z tabulky 2.4 oba faktory byly na hladině významnosti 0.05 statisticky významné pro případ stanovení MgO, CuO a ZnO. Pro stanovení Co₃O₄ byl statisticky významným faktorem shledán pouze faktor A, tj. doba mletí. Pro případ stanovení PbO nebyl ani jeden z testovaných faktorů shledán jako významný.

Jelikož Fischerovo-Snedecorovo testační kritérium F_e nabývá ve všech sledovaných případech nižší hodnoty než kvantil $F_{0.95}(1, 8) = 5.318$, je nulová hypotéza H₀: Interakce I je nulová přijata a interakce mezi faktory A a B, tj. dobou mletí a navážkou vzorku je statisticky nevýznamná.

Tab. 2.3: Průměry koncentrací vybraných oxidů kovů ve vzorcích identifikačních tělísek (%) pro jednotlivé úrovně faktoru doby mletí vzorku (A) a navážky (B) a efekty jednotlivých úrovní

Faktor A			Faktor B		
MgO					
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	7.33	0.256	1	7.58	0.506
2	7.40	0.331	2	7.05	-0.019
3	6.63	-0.444	3	6.85	-0.219
4	6.93	-0.144	4	6.80	-0.269
Celkový průměr: 7.07; Reziduální rozptyl: 0.089; Tukeyho C: -0.340					
PbO					
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	8.20	-0.231	1	8.55	0.112
2	8.70	0.269	2	8.55	0.119
3	8.40	-0.031	3	8.10	-0.331
4	8.43	-0.006	4	8.53	0.094
Celkový průměr: 8.43; Reziduální rozptyl: 0.19; Tukeyho C: 1.638					
CuO					
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	7.65	0.325	1	8.43	1.100
2	7.63	0.300	2	7.60	0.275
3	7.13	-0.200	3	6.45	-0.875
4	6.90	-0.425	4	6.83	-0.500
Celkový průměr: 7.33; Reziduální rozptyl: 0.13; Tukeyho C: 0.232					
ZnO					
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	7.78	-0.763	1	9.35	0.813
2	8.95	0.413	2	8.60	0.063
3	8.83	0.288	3	8.00	-0.054
4	8.60	0.063	4	8.20	-0.338
Celkový průměr: 8.54; Reziduální rozptyl: 0.04; Tukeyho C: -0.118					
Co₃O₄					
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	11.70	-9.469	1	21.88	0.706
2	16.83	-4.344	2	20.90	-0.269
3	23.95	2.781	3	20.93	-0.244
4	32.20	11.03	4	20.98	-0.194
Celkový průměr: 21.17; Reziduální rozptyl: 1.79; Tukeyho C: 0.106					

Tab. 2.4: Tabulka ANOVA pro model s Tukeyho interakcí

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H_0	Spočtená α
MgO						
Mezi úrovněmi A	n-1 = 3	1.572	0.524	5.857	Zamítnuta	0.020
Mezi úrovněmi B	m-1 = 3	1.507	0.502	5.615	Zamítnuta	0.023
Interakce	1	0.017	0.017	0.191	Přijata	0.674
Rezidua	mn-m-n = 8	0.716	0.089			
Celkový	mn-1 = 15	3.794	0.253			
PbO						
Mezi úrovněmi A	n-1 = 3	0.507	0.169	0.913	Přijata	0.477
Mezi úrovněmi B	m-1 = 3	0.587	0.196	1.057	Přijata	0.419
Interakce	1	0.050	0.050	0.270	Přijata	0.618
Rezidua	mn-m-n = 8	1.481	0.185			
Celkový	mn-1 = 15	2.574	0.172			
CuO						
Mezi úrovněmi A	n-1 = 3	1.665	0.555	4.353	Zamítnuta	0.043
Mezi úrovněmi B	m-1 = 3	9.205	3.068	24.065	Zamítnuta	0.000
Interakce	1	0.051	0.051	0.403	Přijata	0.543
Rezidua	mn-m-n = 8	1.481	0.185			
Celkový	mn-1 = 15	2.574	0.172			
ZnO						
Mezi úrovněmi A	n-1 = 3	3.352	1.118	28.157	Zamítnuta	0.000
Mezi úrovněmi B	m-1 = 3	4.268	1.423	35.843	Zamítnuta	0.000
Interakce	1	0.012	0.012	0.313	Přijata	0.591
Rezidua	mn-m-n = 8	0.318	0.040			
Celkový	mn-1 = 15	7.938	0.529			
Co₃O₄						
Mezi úrovněmi A	n-1 = 3	951.8	317.3	177.67	Zamítnuta	0.000
Mezi úrovněmi B	m-1 = 3	2.672	0.891	0.499	Přijata	0.693
Interakce	1	1.769	1.769	0.991	Přijata	0.349
Rezidua	mn-m-n = 8	14.29	1.786			
Celkový	mn-1 = 15	968.8	64.58			

Počet úrovní faktoru A, n = 4; Počet úrovní faktoru B, m = 4.

2.7 Vícenásobné porovnání úrovní jednotlivých faktorů

Z výše uvedených výsledků vyplývá, že faktor A, tj. doba mletí má na hladině významnosti 0.05 statisticky významný vliv na stanovení všech oxidů s výjimkou PbO. Vliv druhého studovaného faktoru navážky byl shledán významným pro stanovení MgO, CuO a ZnO. Ačkoli lze z diagramů středních hodnot obsahu jednotlivých oxidů získat představu o hladinách jednotlivých úrovní obou studovaných faktorů, není z uvedených grafů (obr. 2.2) možné usoudit na statistickou významnost jednotlivých hladin. Pro tento účel bylo v programu NCSS 2007 provedeno vícenásobné porovnání úrovní pomocí Bonferroniho porovnání všech párů a Bonferroniho porovnání sloupců vůči kontrolnímu. Příslušná data jsou uvedena v tabulce 2.5 a 2.6. Data byla vyhodnocena pro oxidy kovů, u nichž byl sledovaný faktor shledán jako významný. Pro případ kdy byly pro stanovení oxidu významné oba faktory, bylo pro porovnání daného faktoru provedeno znáhodnění faktoru druhého.

Tab. 2.5: Bonferroniho porovnání všech párů

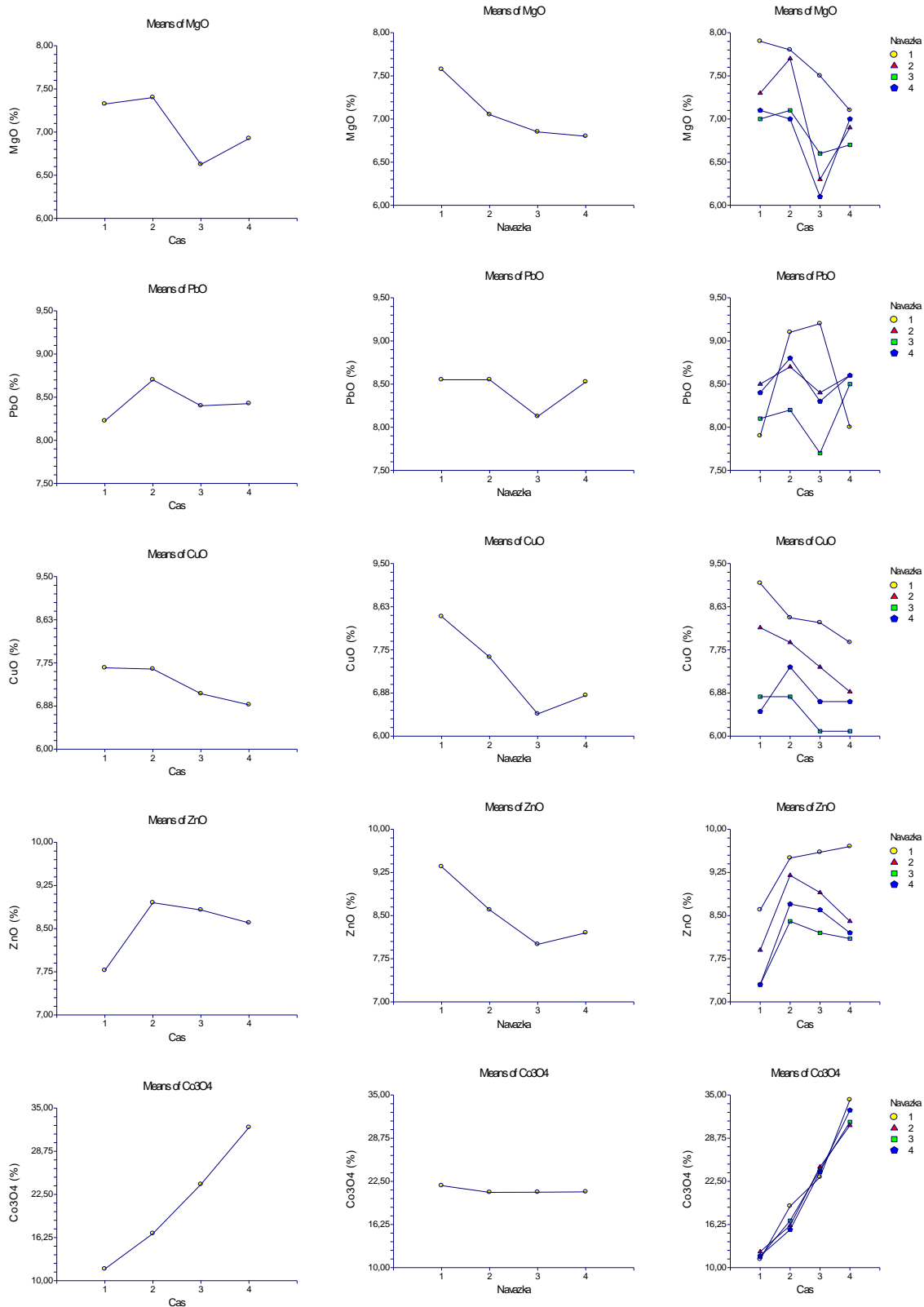
	Faktor A				Faktor B			
	Sloupec	Počet	Průměr	Liší se od	Úroveň	Počet	Průměr	Liší se od
MgO	3	4	6.625	1, 2	4	4	6.80	1
	4	4	6.925		3	4	6.85	1
	1	4	7.325	3	2	4	7.05	
	2	4	7.4	3	1	4	7.575	4, 3
SV=9, MSE=7.95E-02, Kritická hodnota 3.364								
CuO	4	4	6.900		3	4	6.45	2,1
	3	4	7.125		4	4	6.825	1
	2	4	7.625		2	4	7.6	3, 1
	1	4	7.650		1	4	8.425	3, 4, 2
SV=9, MSE=0.113, Kritická hodnota 3.364								
ZnO	1	4	7.775	4,3,2	3	4	8.0	2,1
	4	4	8.6	1	4	4	8.2	1
	3	4	8.825	1	2	4	8.6	3,1
	2	4	8.95	1	1	4	9.35	3,4,2
SV=9, MSE=3.53E-02, Kritická hodnota 3.364								
Co₃O₄	1	4	11.7	2,3,4	2	4	20.9	
	2	4	16.825	1,3,4	3	4	20.925	
	3	4	23.95	1,2,4	4	4	20.975	
	4	4	32.2	1,2,3	1	4	21.875	
SV=9, MSE=1.587, Kritická hodnota 3.364								

Faktor A: doba mletí (čas); Faktor B: navážka; MSE: hodnota průměrného čtverce chyb; SV: stupně volnosti; Kritická hodnota: tabulkový kvantil pro dané stupně volnosti a hladinu významnosti 0.05

Tab. 2.6: Bonferroniho porovnání sloupců vůči kontrolnímu

	Faktor A				Faktor B			
	Sloupec	Počet	Průměr	Liší se od	Úroveň	Počet	Průměr	Liší se od
MgO	3	4	6.625	1, 2	4	4	6.80	1
	4	4	6.925		3	4	6.85	1
	1	4	7.325	3	2	4	7.05	
	2	4	7.4	3	1	4	7.575	4, 3
SV=9, MSE=7.95E-02, Kritická hodnota 2.933								
CuO	4	4	6.900	2, 1	3	4	6.45	2, 1
	3	4	7.125		4	4	6.825	2, 1
	2	4	7.625	4	2	4	7.6	3, 4, 1
	1	4	7.650	4	1	4	8.425	3, 4, 2
SV=9, MSE=0.113, Kritická hodnota 3.364								
ZnO	1	4	7.775	4, 3, 2	3	4	8.0	2, 1
	4	4	8.6	1	4	4	8.2	2, 1
	3	4	8.825	1	2	4	8.6	3, 4, 1
	2	4	8.95	1	1	4	9.35	3, 4, 2
SV=9, MSE=3.53E-02, Kritická hodnota 3.364								
Co₃O₄	1	4	11.7	2, 3, 4	2	4	20.9	
	2	4	16.825	1, 3, 4	3	4	20.925	
	3	4	23.95	1, 2, 4	4	4	20.975	
	4	4	32.2	1, 2, 3	1	4	21.875	
SV=9, MSE=1.587, Kritická hodnota 3.364								

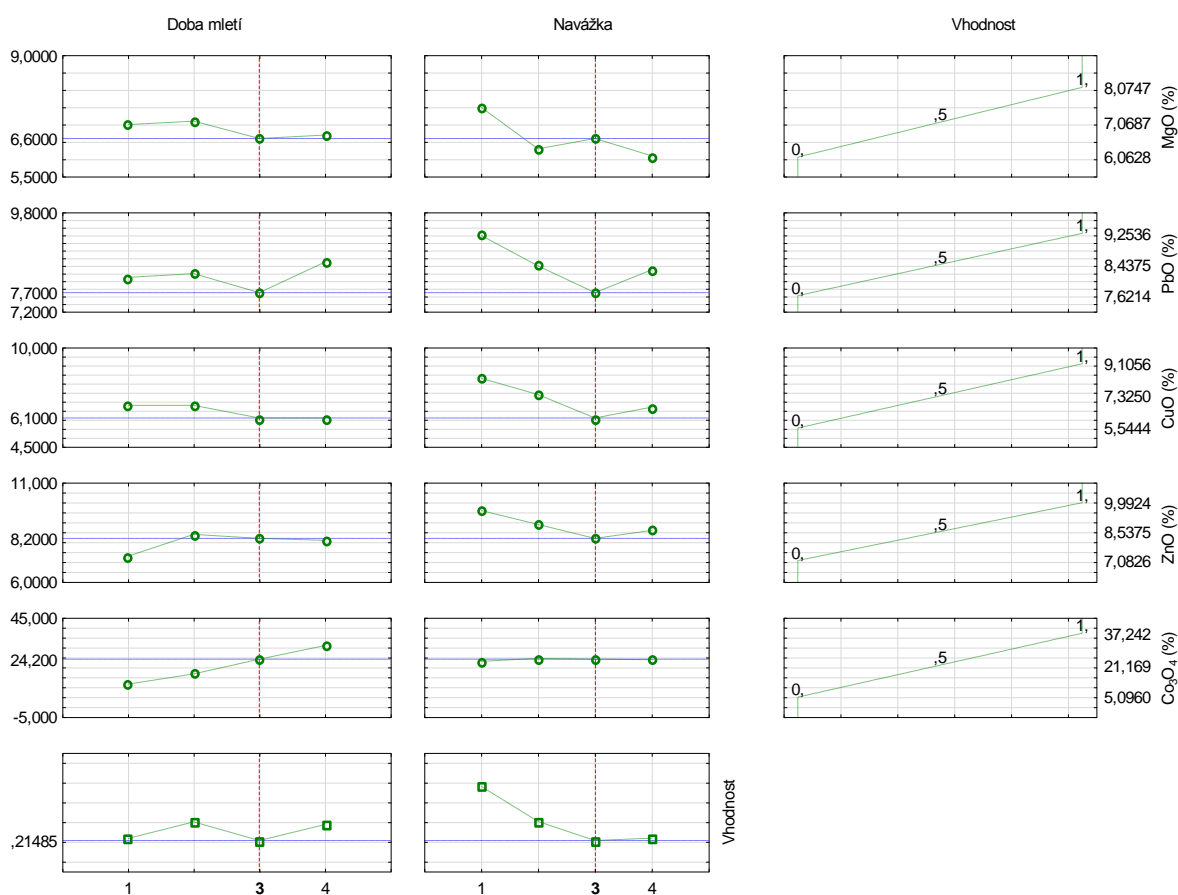
Faktor A: doba mletí (čas); Faktor B: navážka; MSE: hodnota průměrného čtverce chyb; SV: stupně volnosti; Kritická hodnota: tabulkový kvantil pro dané stupně volnosti a hladinu významnosti 0.05



Obr. 2.2 Diagram průměrů stanovení koncentrace vybraných oxidů (%) ve vzorcích identifikačních tělísek v závislosti na době mletí (čas) a navážce vzorku

2.8 Volba optimálních úrovní jednotlivých faktorů

V reálné praxi dochází poměrně často k situaci, kdy je nutné volit optimální hodnoty úrovní jednotlivých faktorů pro relativně komplexní systém, tj. jako v daném případě s ohledem na jednotlivé typy analytu. V tomto případě může být výše provedené víceúrovňové srovnání obtížné z hlediska nalezení optimálních úrovní jednotlivých faktorů. Z tohoto důvodu byly v programu Statistika 12 v modulu dvoufaktorové analýzy rozptýly sestrojeny profily předpovězených hodnot a vhodností (obr. 2.3), které umožňují na základě přiřazení významnosti sledované odezvy určit optimální parametry měření. Pro tento účel byla přiřazena nejvyšší významnost 1 odpovídající hodnotě nejvyšší odezvy (nejlepší citlivosti) a naopak nejnižší hodnotě odezvy nejnižší významnost 0. Optimální hodnoty úrovně faktorů jsou pro jednotlivé oxidy v diagramech na obrázku 2.3 označeny červenou čerchovanou čarou. Pro stanovení všech studovaných oxidů byla jako optimální pro přípravu suspenze indikována navážka 10 mg a doba mletí 7 min.



Obr. 2.3 Profily předpovězených hodnot a vhodnosti

2.9 Závěr

Při posuzování vlivu doby mletí a velikosti navážky na výsledky stanovení vybraných oxidů ve vzorcích detekčních tělísek metodou atomové absorpční spektrometrie s atomizací v plameni, za využití přímého dávkování suspenzí, byly na hladině významnosti 0.05 získány následující závěry. Pro případ stanovení MgO, CuO a ZnO nabývá Fischerovo-Snedecorovo testační kritérium F_e větší hodnoty než kvantily $F_{0.95}(3, 8)$ a nulová hypotéza H_0 : Efekty faktoru A, resp. B jsou nulové je zamítnuta a oba faktory jsou statisticky významné. Pro stanovení Co_3O_4 byl statisticky významným faktorem shledán pouze faktor A, tj. doba mletí. Pro případ stanovení PbO nebyl ani jeden z testovaných faktorů shledán statisticky významným. Jelikož Fischerovo-Snedecorovo testační kritérium F_e nabývá ve všech sledovaných případech nižší hodnoty než kvantil $F_{0.95}(1, 8)$, je nulová hypotéza H_0 : Interakce I je nulová přijata a interakce mezi faktory A a B, tj. dobou mletí a navážkou vzorku je statisticky nevýznamná. S cílem určit statisticky významné rozdíly mezi jednotlivými úrovněmi studovaných faktorů, provedeno bylo vícenásobné porovnání úrovní pomocí Bonferroniho porovnání všech párů a Bonferroniho porovnání sloupců vůči kontrolnímu. Výsledky Bonferroniho porovnání všech párů poskytlo shodné závěry jako Scheffého metoda (výsledky jsou uvedeny v příloze na CD). Ani jeden z postupů neindikoval rozdíly mezi jednotlivými úrovněmi faktoru A (navážka) pro stanovení CuO. Tyto rozdíly jsou však patrné z výsledků Bonferroniho porovnání sloupců vůči kontrolnímu. Optimální úrovně faktorů přípravy vzorku k analýze byly zvoleny pomocí diagramů profilů předpovězených hodnot a vhodností.

2.10 Literatura

1. Meloun M., Militký J.: Kompendium statistického zpracování dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2196-8.
2. Meloun M., Militký J.: Interaktivní statistická analýza dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2173-9.

3 Dvoufaktorová ANOVA s opakováním

3.1 Zadání

Vliv homogenizace vzorku a velikosti navážky na výsledky stanovení rtuti ve vzorcích jater termooxidační metodou

Sledován byl vliv homogenizace vzorku (faktor A) a velikosti navážky (faktor B) na výsledky stanovení rtuti ve vydřích játrech termooxidační metodou. Vzorky byly analyzovány pro navážku 5, 10 a 15 mg bez předchozí přípravy (nemleté) a po homogenizaci vibračním kulovým mlýnkem (mleté). Každý experiment byl opakován osmkrát. Výsledky byly analyzovány pomocí dvou faktorové analýzy rozptylu (ANOVA).

S ohledem na malé množství vzorku unikátního biologického materiálu bylo cílem nalézt optimální velikost navážky lyofilizovaného vzorku a posoudit vliv další úpravy (homogenizace) vzorku na výsledky stanovení s ohledem na dosažení reprezentativního analytického výsledku, zkrácení doby postupu přípravy vzorku k analýze a omezení rizika kontaminace, které zvyšuje v oblasti stopové analýzy každý krok analytického postupu.

3.2 Data

Tab. 3.1: Koncentrace rtuti (mg kg^{-1}) stanovená ve vzorcích vepřových jater termooxidační metodou v závislosti na homogenizaci vzorku (faktor A) a navážce (faktor B)

	B1 (5 mg)	B2 (10 mg)	B3 (15 mg)
A1 (nemleto)	0.021, 0.014, 0.007, 0.025 0.019, 0.006, 0.026, 0.027	0.012, 0.013, 0.008, 0.017 0.009, 0.020, 0.008, 0.015	0.007, 0.009, 0.016, 0.008 0.007, 0.011, 0.009, 0.012
A2 (pomleto)	0.019, 0.014, 0.025, 0.023 0.014, 0.029, 0.025, 0.012	0.012, 0.014, 0.012, 0.013 0.014, 0.020, 0.021, 0.012	0.011, 0.016, 0.014, 0.011 0.014, 0.008, 0.014, 0.007

3.3 Použitý software

Data byla analyzována za využití programů Adstat 1.25 a QC-Expert 2.9 (TriloByte Statistical Software, s.r.o., ČR) a NCSS 2007 (NCSS, USA).

3.4 Ověření základních předpokladů o datech

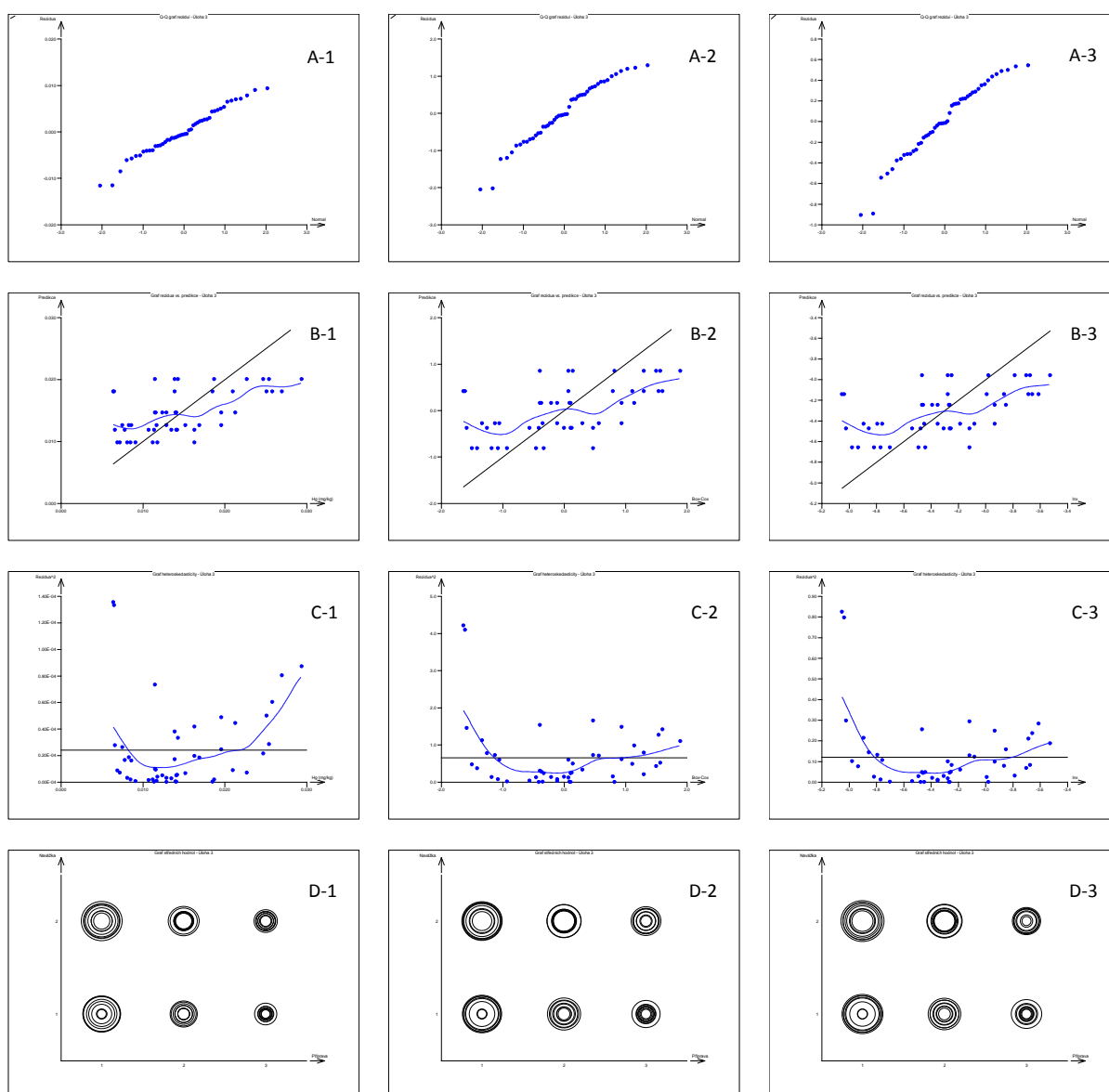
Pro posouzení základních předpokladů o datech bylo použito vybraných grafických diagnostik programu QC-Expert v modulu ANOVA (dvoufaktorová). Na základě Q-Q grafu reziduí není patrný výrazný odklon experimentálních dat od normality. Z grafu rezidua vs. predikce, který vyjadřuje efektivitu modelu ANOVA, je zřejmé, že data neleží na přímce $y = x$, a že body leží spíše na vodorovné linii. Lze tedy usuzovat na nevýznamnost modelu. Z grafu heteroskedasticity je patrná výrazná závislost velikosti chyby na pozorované hodnotě. Data vykazují heteroskedasticitu a výsledky analýzy tak mohou být méně spolehlivé.

3.5 Transformace dat

V programu QC-Expert byla provedena Box-Coxova a exponencionální transformace dat a sledován její vliv na kvalitu dat prostřednictvím výše zmíněných grafických diagnostik (obr. 3.1). Zatímco vliv transformace na Q-Q graf reziduí a graf rezidua vs. predikce je minimální, patrný je přínos transformace dat v grafu heteroskedasticity. Modrá křivka neparametrického vyhlazení čtverce reziduí lépe koresponduje s černou křivkou, která představuje odhad σ^2 .

V programu Adstat 1.25 byl dále sledován vliv logaritmické transformace na hodnotu korelačního koeficientu. Hodnota korelačního koeficientu R pro původní data činila 0.860, hodnota po logaritmické transformaci se pak výrazně snížila, a to na hodnotu -0.379, což dokazuje význam transformace.

Z výše uvedených důvodů byla analýza rozptylu provedena po transformaci vstupních dat. Pro transformaci byla zvolena logaritmická transformace provedená v programu Adstat 1.25.



Obr. 3.1 Q-Q graf reziduí (A), graf rezidua vs predikce (B), graf heteroskedasticity (C) a graf středních hodnot (D) pro původní data (1), data po Box-Coxově (2) a logaritmické (3) transformaci.

3.6 Průměry a efekty úrovní

Hodnoty průměrů a efekty úrovní faktoru homogenizace vzorku a faktoru velikosti navážky na obsah Hg ve vzorku jater po logaritmické transformaci shrnuje tabulka 3.2. Informace o hodnotách jednotlivých skupinových průměrů poskytuje graf středních hodnot na obrázku 3.1.

Tab. 3.2: Průměry obsahu Hg v játrech (mg kg^{-1}) pro jednotlivé úrovně faktoru homogenizace vzorku a navážky a jejich efekty

Faktor A			Faktor B		
Úroveň	Průměr	Efekt	Úroveň	Průměr	Efekt
1	-4.41	-0.092	1	-4.05	0.267
2	-4.23	0.092	2	-4.34	-0.019
			3	-4.57	-0.248

3.7 Tabulka ANOVA

Testovány byly následující hypotézy:

H_0 : Efekty faktoru A jsou nulové, H_A : nejsou nulové; Kvantil $F_{1-\alpha}(n-1, mn(o-1)) = 4.073$

H_0 : Efekty faktoru B jsou nulové, H_A : nejsou nulové; Kvantil $F_{1-\alpha}(m-1, mn(o-1)) = 3.220$

H_0 : Interakce (efekty interakci A a B dohromady) je nulová, H_A : není nulová;

Kvantil $F_{1-\alpha}((n-1)(m-1), nm(o-1)) = 3.220$

Z výsledků analýzy rozptylu vyplývají následující závěry:

Jelikož Fischerovo-Snedecorovo testační kritérium $F_e = 2.906$ nabývá větší hodnoty než kvantil $F_{0.95}(1, 42) = 4.073$, je nulová hypotéza H_0 : Efekty faktoru A jsou nulové přijata a faktor A je statisticky nevýznamný.

Jelikož Fischerovo-Snedecorovo testační kritérium $F_e = 7.683$ nabývá větší hodnoty než kvantil $F_{0.95}(2, 42) = 3.220$, je nulová hypotéza H_0 : Efekty faktoru B jsou nulové zamítnuta a faktor B je statisticky významný.

Jelikož Fischerovo-Snedecorovo testační kritérium $F_e = 0.001$ nabývá nižší hodnoty než kvantil $F_{0.95}(2, 42) = 3.220$, je nulová hypotéza H_0 : Interakce I je nulová přijata a interakce faktor A a B je statisticky nevýznamná.

Tab. 3.3: Tabulka ANOVA pro model s interakcí faktoru A a B po logaritmicke transformaci dat

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H_0	Spočtená α
Mezi úrovněmi A	$n-1 = 1$	0.403	0.403	2.906	Přijata	0.096
Mezi úrovněmi B	$m-1 = 2$	2.13	1.065	7.683	Zamítnuta	0.001
Interakce	$(n-1)(m-1) = 2$	3.08×10^{-4}	1.54×10^{-4}	0.001	Přijata	0.999
Rezidua	$mn(o-1) = 42$	5.82	0.139			
Celkový	$mno-1 = 47$	8.359	0.178			

Počet úrovní faktoru A, $n = 2$; Počet úrovní faktoru B, $m = 3$; Počet opakování, $o = 8$.

3.8 Vícenásobné porovnání úrovní jednotlivých faktorů

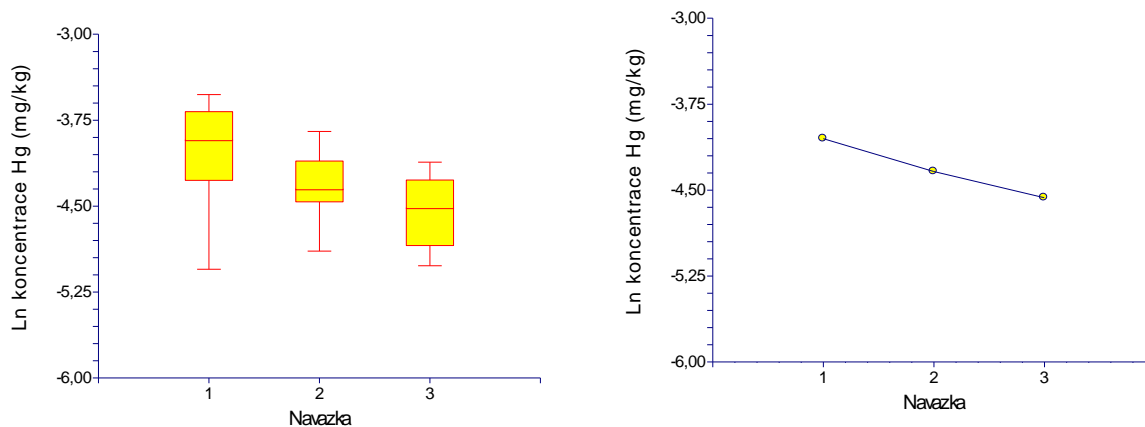
Z výše uvedených výsledků vyplývá, že na hladině významnosti 0.05 má statisticky významný vliv na výsledky stanovení Hg pouze navážka vzorku. Ačkoli je z grafu středních hodnot na obrázku 3.1 patrné, že pro nejnižší navážku vzorku bylo dosaženo často vyšší koncentrace Hg, bylo pro srovnání jednotlivých úrovní faktoru navážky provedeno párové porovnání dvojic Scheffého metodou. Toto srovnání bylo provedeno pro data po logaritmicke transformaci v programu Adstat 2.0, a to v modulu jedno faktorové analýzy rozptylu ANOVA#1. Toto srovnání bylo možné provést, jelikož byl vliv prvního sledovaného faktoru, tj. faktoru homogenizace vzorku shledán jako nevýznamný a mohl být tedy z další analýzy vypuštěn. Kromě toho výběrové sloupce po transformaci dat vykazují normalitu a homoskedasticitu.

Jak je patrné z tabulky 3.4, nulová hypotéza $H_0 (\mu_i - \mu_j = 0)$ byla na hladině významnosti 0.05 přijata pro všechny sledované dvojice s výjimkou dvojice 1–3. O statisticky nevýznamném rozdílu mezi úrovněmi 1 a 3 svědčí konfidenční intervaly, které neobsahují nulu (tab. 3.4). Hodnoty průměrů pro jednotlivé úrovně faktoru navážky jsou pro transformovaná data patrná z obrázku 3.2. Zde je z diagramu průměrů vyhodnoceného programem NCS 2007 patrný vyšší průměr hodnot koncentrací pro první (nejnižší) sledovanou navážku, z krabicového grafu je dále zřejmý výrazně vyšší rozptyl výsledků pro první úroveň sledovaného faktoru.

Tab. 3.4: Vícenásobné porovnání úrovní faktoru navážky Schéffeho metodou

Hypotéza H_0	Průměrný párový rozdíl	Dolní mez CI ^a	Horní mez CI ^a	Závěr
$P1 = P2$	0.287	-0.046	0.620	Přijata
$P1 = P3$	0.515	0.182	0.848	Zamítnuta
$P2 = P3$	0.228	-0.105	0.561	Přijata

^a 95% CI konfidenční interval



Obr. 3.2 (A) Krabicový graf a (B) diagram průměrů koncentrace rtuti po logaritmické transformaci zdrojových dat v lyofilizovaných vepřových játrech v závislosti na navážce vzorku

3.9 Závěr

Byl studován vliv faktoru velikosti navážky a vlivu homogenizace na výsledky stanovení rtuti v lyofilizovaných vzorcích vepřových jater metodou termooxidačního stanovení. Na základě ověření základních předpokladů o datech bylo zjištěno, že data vykazují normální rozdělení, jsou nezávislá a v datech není patrný trend. Nicméně data vykazovala heteroskedasticitu. Za účelem stabilizace rozptylu byla provedena transformace dat. Z grafu heteroskedasticity bylo pozorováno, že jak Box-Coxova, exponenciální i logaritmická transformace dat vedly ke zlepšení výše uvedeného problému. Dvou faktorová analýza rozptylu byla provedena pro data po logaritmické transformaci, z výsledků bylo zjištěno, že vliv další homogenizace vzorku nemá na hodnotu koncentrace rtuti statisticky významný vliv. Vzorky tedy není nutné před vlastní analýzou dále homogenizovat, což je výhodné nejen z hlediska výrazného zkrácení celkové doby analytického postupu, ale i z hlediska rizika kontaminace vzorku při procesu mletí. Faktor velikosti navážky byl shledán jako statisticky významný. Metodou vícenásobného porovnání jednotlivých úrovní faktoru navážky bylo zjištěno, že významný rozdíl mezi koncentracemi rtuti je pozorován pro nejnižší a nejvyšší sledovanou navážku. Tento fakt lze vysvětlit tím, že při analýze menšího podílu navážky vzorku může dojít ve srovnání s nejvyšším sledovaným množstvím k snazšímu uvolnění analytu z matrice vzorku. Nicméně jak je patrné z krabicového diagramu uvedeného na obr. 3.2 je při analýze malého množství vzorku indikován výrazně vyšší rozptyl výsledků pravděpodobně v důsledku rozdílné distribuce analytu ve vzorku. V daném případě může být analýza příliš malého množství za předpokladu nízkého opakování odpovědná za dosažení nereprezentativních analytických výsledků.

3.10 Literatura

1. Meloun M., Militký J.: Kompendium statistického zpracování dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2196-8.
2. Meloun M., Militký J.: Interaktivní statistická analýza dat, Vyd. 3., nakl. Karolinum, Praha, 2012. ISBN: 978-80-246-2173-9.