

Univerzita Pardubice  
Fakulta chemicko-technologická  
Katedra analytické chemie

# **ANOVA**

## **Semestrální práce**

Licenční studium GALILEO – Interaktivní statistická analýza dat

Brno, 2015

Doc. Mgr. Jan Muselík, Ph.D.  
VFU Brno, Ústav technologie léků

## **Obsah**

<b>Úloha 1. Jednofaktorová ANOVA .....</b>	<b>3</b>
<b>Úloha 2. Dvoufaktorová ANOVA bez opakování.....</b>	<b>6</b>
<b>Úloha 3. Dvoufaktorová ANOVA s opakováním (nevyvážená) .....</b>	<b>9</b>

# Úloha 1. Jednofaktorová ANOVA

**Zadání:** Cílem úlohy je ověřit zda různé druhy *Staphylea* L. (faktor A) významně ovlivňují obsah celkových flavonů a flavonolů v extraktech připravených z těchto rostlin. Celkový obsah flavonů a flavonolů je vyjádřen jako ekvivalent kvercetinu (QE g/100 g suchého extraktu) a byl změřen u 4 extraktů z různých druhů *Staphylea* L.

**Data:**

QE (g/100 g); n = 15, k = 4

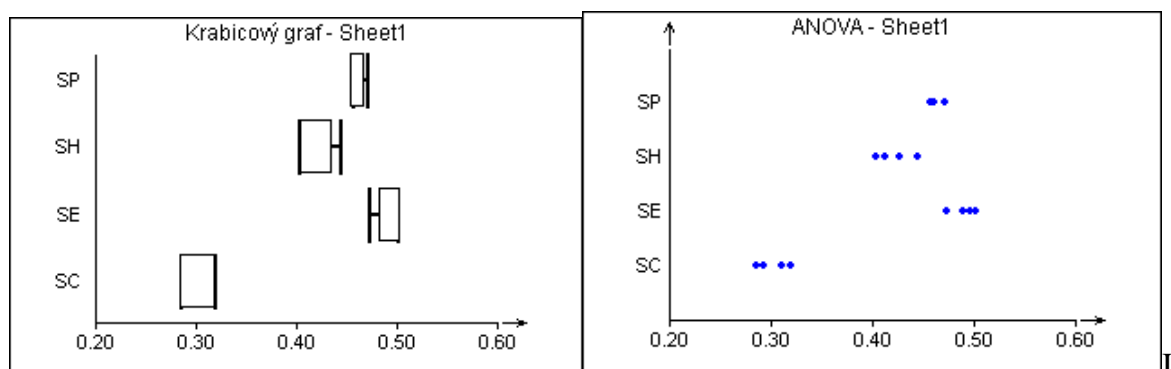
Opakování	<i>S. colchica</i> SC	<i>S. elegans</i> SE	<i>S. holocarpa</i> SH	<i>S. pinnata</i> SP
1	0,311	0,501	0,427	0,472
2	0,292	0,497	0,412	0,457
3	0,285	0,489	0,404	0,461
4	0,320	0,473	0,444	-

**Užitý program:** QC Expert

**Řešení:**

H<sub>0</sub>: Faktor A nemá vliv na měření v jednotlivých skupinách dat; druh *Staphylea* L. nemá vliv na obsah celkových flavonů a flavonolů.

H<sub>A</sub>: Faktor A má vliv na měření v jednotlivých skupinách dat; druh *Staphylea* L. má vliv na obsah celkových flavonů a flavonolů.



EDA

Graf 1.1 – Krabicový graf

Graf 1.2 – Diagram rozptýlení

Z krabicových grafů pro všechny skupiny na první pohled vyplývá, že jeden datový soubor vybočuje. Hodnoty pro vzorek SC se zdají být významně odlišné od všech ostatních. Vnitřní hradby vzorku SH se nepřekrývají s vnitřními hradbami ostatních vzorků, proto je možné soudit, že i tato skupina dat bude významně odlišná od ostatních vzorků.

## II. Ověření předpokladů výběru

Hladina významnosti:  $\alpha = 0,05$

skupina	SC	SE	SH	SP
Střední hodnota	0,303	0,487	0,424	0,463*
Pivotové rozpětí	0,035	0,028	0,040	0,037*
Homogenita	<i>Přijata</i>	<i>Přijata</i>	<i>Přijata</i>	-
vybočující b.	0	0	0	-

\* zvláště malé výběry (n = 3)

Střední hodnoty jednotlivých skupin dat se pohybují kolem hodnoty 0,45, až na vzorek SC, kde je střední hodnota 0,30. Již z tohoto porovnání je patrné, že se tato skupina dat liší od ostatních a je tedy pravděpodobné, že díky tomu bude nulová hypotéza testu ANOVA zamítnuta.

## III. ANOVA

Tabulka ANOVA

Teoretický	Vypočítaný	závěr	p-hodnota
3,6	128,5	H <sub>0</sub> je zamítnuta	7,63E-009

Párové porovnávání dvojic úrovní – Scheffého metoda

Srovnávaná dvojice	Rozdíl	Významnost	Pravděpodobnost
SC - SE	-0,188	Významný	1,56E-008
SC - SH	-0,120	Významný	1,71E-006
SC - SP	-0,161	Významný	1,76E-007
SE - SH	0,068	Významný	0,000354
SE - SP	0,027	Nevýznamný	0,182
SH - SP	-0,042	Významný	0,0238

***Závěr:***

V této úloze bylo snahou statisticky vyhodnotit vliv druhu *Staphylea L.* na obsah celkových flavonů a flavonolů. Nulová hypotéza o nevýznamnosti vlivu faktoru na naměřená data byla zamítnuta. Pomocí Scheffého metody vzájemného porovnání skupin byly potvrzeny významné rozdíly mezi všemi dvojicemi mimo dvojice SE – SP.

## Úloha 2. Dvoufaktorová ANOVA bez opakování

**Zadání:** Byla provedena zkouška obsahové stejnoměrnosti tablet klatrátu sodné soli warfarinu. Tablety byly lisovány z tabletových směsí mísených různě dlouhý čas (faktor A) a lišily se použitým plnivem (faktor B). U každé připravené šarže bylo analyzováno 10 tablet a ze zjištěných výsledků byla spočítána relativní směrodatná odchylka (RSD). Záměrem je zjistit, zda má délka mísení směsí a jejich složení významný vliv na obsahovou stejnoměrnost vyjádřenou jako RSD.

### Data:

relativní směrodatná odchylka (%)

	plnivo 1	plnivo 2	plnivo 3
5 min	5,3	5,9	4,1
10 min	4,1	6,2	3,5
15 min	4,3	4,3	2,6

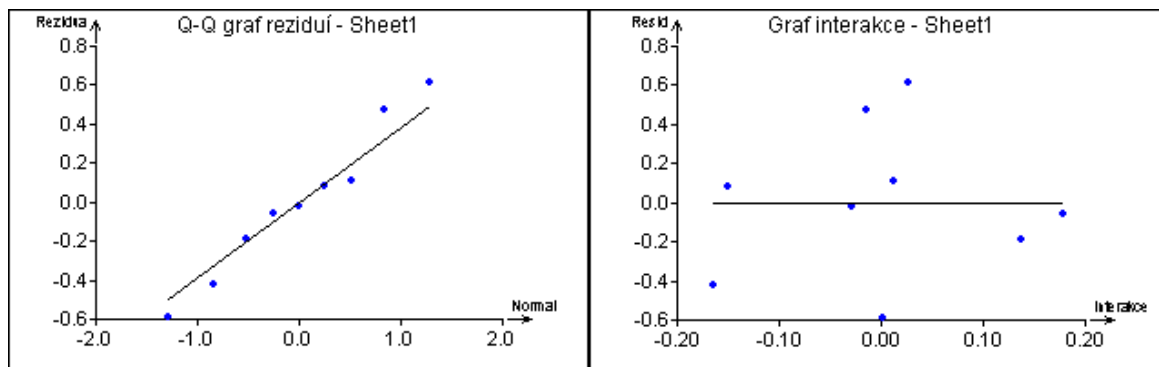
**Užitý program:** QC Expert

### Řešení:

$H_0$ : Faktory A, B a interakce mezi nimi nemají významný vliv na výsledky v jednotlivých skupinách dat; délka mísení (faktor A) a složení tablet (faktor B) nemají významný vliv na hodnotu RSD.

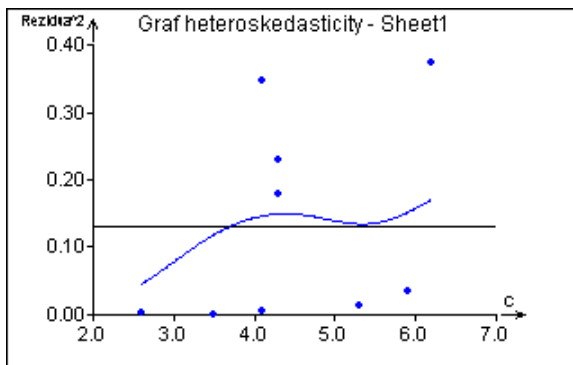
$H_A$ : Faktory A, B, nebo interakce mezi nimi mají významný vliv na výsledky v jednotlivých skupinách dat; délka mísení (faktor A) nebo složení tablet (faktor B) mají významný vliv na hodnotu RSD.

### I. Grafické diagnostiky

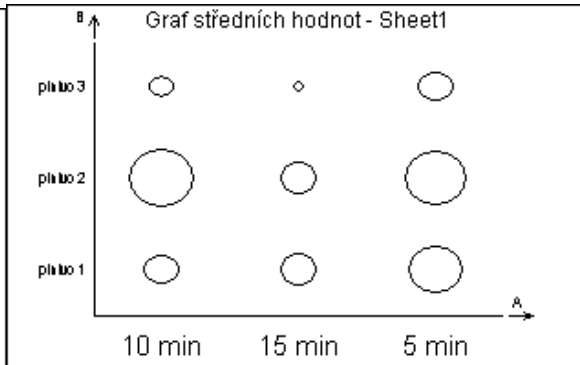


Graf 1.3 – Q-Q graf reziduí

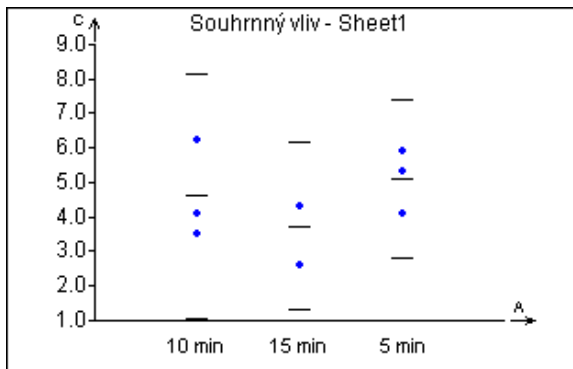
Graf 1.4 – Graf interakce



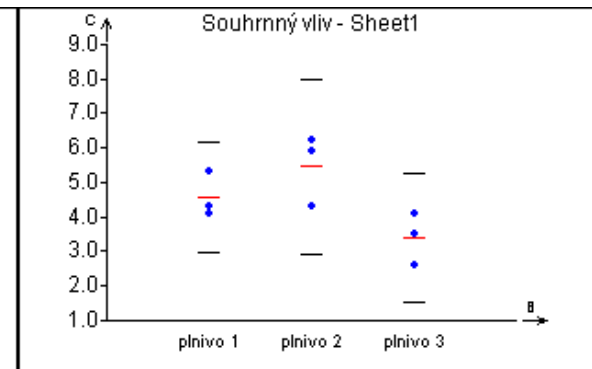
Graf 1.5 – Graf heteroskedasticity



Graf 1.6 – Graf středních hodnot



Graf 1.7 – Graf souhrnného vlivu (faktor A)



Graf 1.8 – Graf souhrnného vlivu (faktor B)

Pro posouzení normality reziduí byl sestaven Q-Q graf. Body leží přibližně na přímce, dá se tedy říci, že rezidua mají normální rozdělení bez vybočujících hodnot.

Významnost interakce byla posouzena podle grafu interakce. Body jsou v prostoru rozmístěny spíše náhodně, přímka nevykazuje rostoucí tendenci, proto se dá interakce mezi faktory A a B považovat za statisticky nevýznamnou.

Z grafu heteroskedasticity je patrné, že data vykazují mírnou nekonstantnost rozptylu, jelikož modrá křivka neparаметrického vyhlazení čtverce reziduí nekopíruje přesně černou vodorovnou přímku. Tento fakt může mít mírný vliv na spolehlivost výsledků získaných metodou ANOVA.

První porovnání vlivu faktoru A a B na data je možné provést na základě grafu středních hodnot. Z grafu je patrné, že faktory A ani faktory B příliš velký vliv nevykazují.

Lépe je významnost vlivu jednotlivých faktorů vidět z grafů souhrnného vlivu. Druh použitého plniva ovlivňuje výsledek, což je potvrzeno i zobrazením středních hodnot v grafu červeně. Naopak délka mísení tabletové směsi významný vliv nemá, protože jsou hodnoty napříč všemi čtyřmi úrovněmi faktoru značně podobné.

## II. ANOVA

Tabulka ANOVA; hladina významnosti:  $\alpha = 0,05$

zdroj variability	stupně volnosti	součet čtverců	průměrný čtverec	testovací F-kritérium	kritický kvantil	závěr	p-hodnota
délka mísení	2	2,869	1,434	4,844	6,944	$H_0$ není zamítnuta	0,085
složení	2	6,442	3,221	10,878	6,944	$H_0$ je zamítnuta	0,024
interakce	1	0,010	0,010	0,025	10,127	$H_0$ není zamítnuta	0,884
reziduální	3	1,175	0,392				
celkový	8	10,496	1,312				

Protože F-kritérium (4,844) je menší než kritický kvantil (6,944) není nulová hypotéza  $H_0$  zamítnuta a faktor A je statisticky nevýznamný. Protože F-kritérium (10,878) je větší než kritický kvantil (6,944) je nulová hypotéza  $H_0$  zamítnuta a faktor B je statisticky významný. Protože F-kritérium (0,025) je menší kritický kvantil (10,127) je nulová hypotéza  $H_0$  přijata a interakce faktorů A a B je statisticky nevýznamný.

### **Závěr:**

Cílem této úlohy bylo zjistit, které faktory (délka mísení tabletové směsi a složení tablet, případně interakce mezi nimi) mají vliv na obsahovou stejnoměrnost produktu vyjádřenou jako RSD. Pomocí dvoufaktorové metody ANOVA bez opakování bylo zjištěno, že významný vliv má pouze faktor B, tedy složení tablet, zatímco délka mísení a interakce mezi faktory se jeví jako statisticky nevýznamné.



### Úloha 3. Dvoufaktorová ANOVA s opakovaním (nevyvážená)

**Zadání:** Cílem je zjistit vliv použitého pH modifikátoru (faktor A) a jeho obsahu v tabletě (faktor B) na množství uvolněného verapamil hydrochloridu zjištěného zkouškou disoluce (čas 360 minut). Pro analýzu souboru je nutné použít nevyvážený model testu ANOVA, jelikož existuje pro každou kombinaci faktorů různý počet měření.

**Data:**

množství uvolněného léčiva (%); n = 22

kyselina	obsah pH modifikátoru v tabletě 50 mg								
citronová	52,94	52,49	51,46	52,01					
fumarová	60,22	51,78	53,18	56,45					
itakonová	52,76	51,21	52,42	52,86					
	obsah pH modifikátoru v tabletě 100 mg								
citronová	63,43	59,65	55,71						
fumarová	53,26	59,77	57,93	57,02					
itakonová	55,69	53,59	50,87						

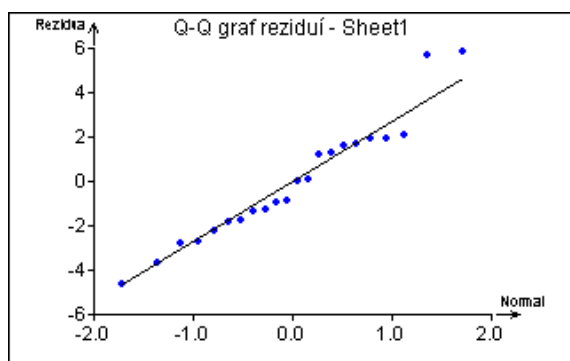
**Užitý program:** QC Expert

**Řešení:**

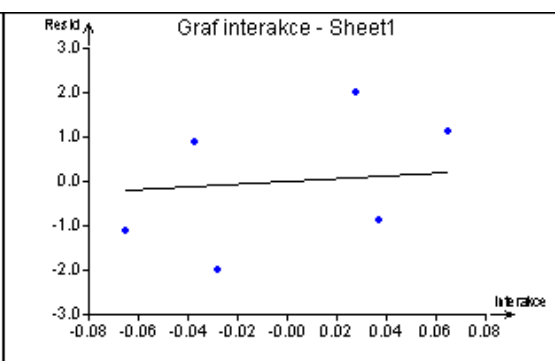
$H_0$ : Faktory A, B a interakce mezi nimi nemají významný vliv na měření v jednotlivých skupinách dat; použitý pH modifikátor (faktor A) a jeho obsah v tabletě (faktor B) nemají významný vliv na množství uvolněného léčiva v čase 360 minut.

$H_A$ : Faktory A, B, nebo interakce mezi nimi mají významný vliv na měření v jednotlivých skupinách dat; použitý pH modifikátor (faktor A) a jeho obsah v tabletě (faktor B) mají významný vliv na množství uvolněného léčiva v čase 360 minut.

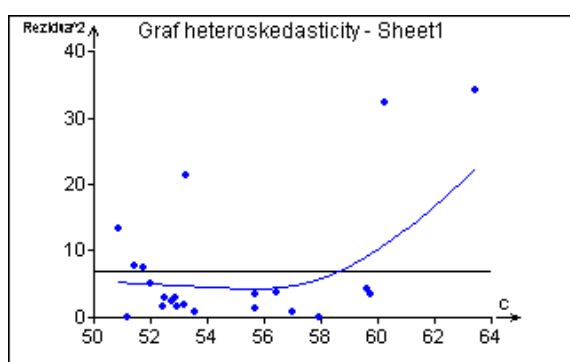
## I. Grafické diagnostiky



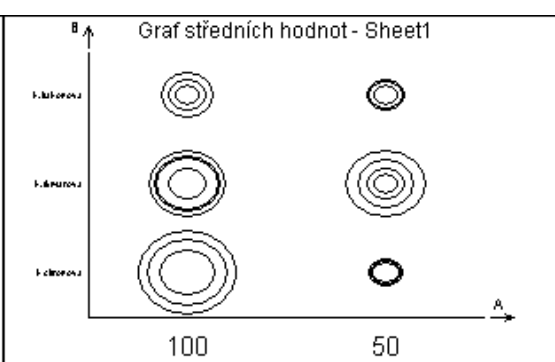
Graf 1.9 – Q-Q graf reziduí



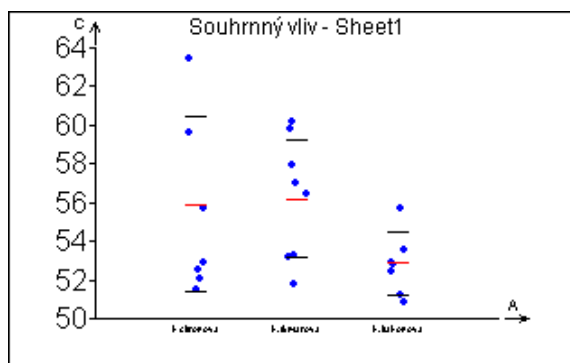
Graf 1.10 – Graf interakce



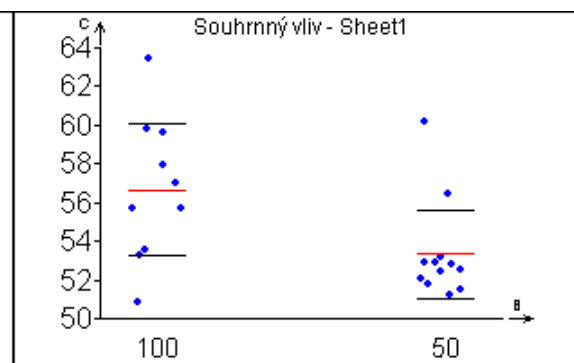
Graf 1.11 – Graf heteroskedasticity



Graf 1.12 – Graf středních hodnot



Graf 1.13 – Graf souhrnného vlivu (faktor A)



Graf 1.14 – Graf souhrnného vlivu (faktor B)

V Q-Q grafu jsou vidět dva body, které se odchyľují od přímky, což může ukazovat na asymetrii rozdělení.

Graf interakce prokázal nevýznamnost interakce mezi faktory A a B, body leží víceméně na přímce rovnoběžné s osou  $x$ .

Z grafu heteroskedasticity je patrná nekonstantnost rozptylu, což může mít mírný vliv na spolehlivost výsledků získaných metodou ANOVA.

Graf středních hodnot poskytl první porovnání vlivu faktoru A a B na datový soubor. Z tohoto znázornění se však s jistotou nedá říci, zda mají jednotlivé faktory na výsledek vliv.

Významnost vlivu jednotlivých faktorů je lépe znázorněná pomocí grafů souhrnného vlivu. Je vidět, že u různých pH modifikátorů je střední hodnota uvolněného množství léčiva rozdílná. Podobně můžeme pozorovat rozdílné střední hodnoty uvolněného množství léčiva pro různé obsahy pH modifikátorů v tabletě.

## II. ANOVA

Tabulka ANOVA; hladina významnosti:  $\alpha = 0,05$

zdroj variability	stupně volnosti	součet čtverců	průměrný čtverec	testovací F-kritérium	kritický kvantil	závěr	p-hodnota
A	2	49,706	24,853	3,638	3,634	$H_0$ je zamítnuta	0,0499
B	1	60,361	60,361	8,835	4,494	$H_0$ je zamítnuta	0,0090
interakce	2	44,054	22,027	3,224	3,634	$H_0$ není zamítnuta	0,0666
reziduální	16	109,317	6,832				
celkový	21	256,739	12,226				

Analýza dat metodou ANOVA zamítla nulovou hypotézu pro oba faktory i pro jejich interakci. Protože F-kritérium (3,638) je větší než kritický kvantil (3,634) je nulová hypotéza  $H_0$  zamítnuta a faktor A je statisticky významný. Nicméně p-hodnota se velmi blíží stanovené hladině významnosti, což značí, že tento závěr nemusí být zcela spolehlivý. Protože F-kritérium (8,835) je větší než kritický kvantil (4,494) je nulová hypotéza  $H_0$  zamítnuta a faktor B je statisticky významný. Protože F-kritérium (3,224) je menší kritický kvantil (10,127) je nulová hypotéza  $H_0$  přijata a interakce faktorů A a B je statisticky nevýznamný.

### **Závěr:**

Záměrem úlohy bylo zjistit, které faktory (druh modifikátoru a jeho obsah v tabletě) mají vliv na množství uvolněného léčiva v čase.

Pomocí dvoufaktorové metody ANOVA s opakováním (nevyvážený model) byla zjištěna statistická významnost obou faktorů, jejich interakce je nevýznamná. Druh pH modifikátoru i jeho obsah v tabletě mají na vliv na množství uvolněného léčiva v čase.