

# Statistické zpracování dat 3.přednáška

Mgr. Radmila Krkošková, Ph.D.



**SLEZSKÁ  
UNIVERZITA**

OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



# Dvoufaktorová a vícefaktorová ANOVA



# 2Faktorová ANOVA

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- Techniky testů hypotéz o rozdílech mezi skupinami, kdy rozdíly způsobuje 2 nebo více faktorů

## Příklad otázek, na které odpovídá 2F ANOVA:

- - Má typ obalu nové limonády, resp. pohlaví respondentů vliv na uváděnou pravděpodobnost prodeje limonády?
  - Která složka má větší vliv?
  - Je celkový vliv součtem vlivů jednotlivých znaků posuzovaných odděleně?



# ANOVA bez interakce, ANOVA s interakcí

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Účinky jednotlivých znaků mohou být:

- a) vzájemně nezávislé (bez interakce),
- b) závislé (s interakcí)



## 2-faktorová ANOVA bez interakce

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- **Předpoklad:** Účinky jednotlivých znaků jsou vzájemně nezávislé, sčítají se
- Každý respondent se nachází v právě jediné skupině tříděné podle 1. faktoru  $X_1$  a právě jediné skupině podle 2. faktoru  $X_2$
- Často se jedná o vyhodnocení experimentu, kdy se sleduje vliv dvou faktorů na daný znak, jeden faktor je ovlivnitelný výběrem hodnot



## 2 faktorová ANOVA - Příklad



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

benzin/řidič	Aral	Shell	Benzina	Slovnaft
A	7,5	6,8	7,2	8,0
B	7,9	7,2	8,1	8,3
C	7,9	6,5	7,8	7,2
D	7,3	7,0	7,6	7,5
E	6,9	7,3	7,8	8,2
F	7,8	6,2	6,9	8,7

Každý řidič jede s každým typem benzínu právě jedenkrát – bez opakování.



## 2-faktorová ANOVA: Součty čtverců



- Celkový součet čtverců:  $S_y = \sum_{i=1}^k \sum_{j=1}^r (y_{ij} - \bar{y})^2$
- Meziskupinový součet čtverců:  $S_{y,m} = r \sum_{i=1}^k (\bar{y}_{i\cdot} - \bar{y})^2$
- Meziblokový součet čtverců:  $S_{y,b} = k \sum_{j=1}^r (\bar{y}_{\cdot j} - \bar{y})^2$
- Vnitroskupinový součet čtverců:  $S_{y,v} = \sum_{i=1}^k \sum_{j=1}^r (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2$



## 2-faktorová ANOVA: Součty čtverců



- Základní vztah 2F ANOVA:  $S_y = S_{y,m} + S_{y,v} + S_{y,b}$

- Testová kritéria:

$$F_1 = \frac{\frac{S_{y,m}}{k-1}}{\frac{S_{y,v}}{(k-1)(r-1)}}$$

$$F_2 = \frac{\frac{S_{y,b}}{r-1}}{\frac{S_{y,v}}{(k-1)(r-1)}}$$





# Jednofaktorová ANOVA – příklad v Excelu



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Data → Analýza dat → ANOVA: Jeden faktor...

Anova: jeden faktor		Faktor benzín				
Faktor						
Výběr	Počet	Součet	Průměr	Rozptyl		
Aral	6	45,3	7,55	0,16		
Shell	6	41,0	6,83	0,18		
Benzina	6	45,4	7,57	0,19		
Slovnaft	6	47,9	7,98	0,30		
ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Mezi výběry	4,103	3	1,368	6,560	0,003	3,098
Všechny výběry	4,170	20	0,209			
Celkem	8,273	23				

p-hodnota = 0,003 < 0,05 ⇒ faktor **benzín** má na spotřebu vliv



# Jednofaktorová ANOVA – příklad v Excelu



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Data → Analýza dat → ANOVA: Jeden faktor...

Anova: jeden faktor		Faktor řidiči			
Faktor					
	Výběr	Počet	Součet	Průměr	Rozptyl
A		4	29,5	7,38	0,26
B		4	31,5	7,88	0,23
C		4	29,4	7,35	0,42
D		4	29,4	7,35	0,07
E		4	30,2	7,55	0,32
F		4	29,6	7,40	1,18

  

ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Mezi výběry	0,848	5	0,170	0,411	0,835	2,773
Všechny výběry	7,425	18	0,413			
Celkem	8,273	23				

p-hodnota = 0,835 > 0,05 ⇒ faktor **řidič** nemá na spotřebu vliv



# Dvoufaktorová ANOVA – 1. příklad v Excelu



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Data → Analýza dat → ANOVA: Dva faktory bez opakování

Anova: dva faktory bez opakování					
<i>Faktor</i>	<i>Počet</i>	<i>Součet</i>	<i>Průměr</i>	<i>Rozptyl</i>	
Aral	6	45,3	7,55	0,16	
Shell	6	41,0	6,83	0,18	
Benzina	6	45,4	7,57	0,19	
Slovnaft	6	47,9	7,98	0,30	
A	4	29,5	7,38	0,26	
B	4	31,5	7,88	0,23	
C	4	29,4	7,35	0,42	
D	4	29,4	7,35	0,07	
E	4	30,2	7,55	0,32	
F	4	29,6	7,40	1,18	

  

ANOVA						
<i>Zdroj variability</i>	<i>SS</i>	<i>Rozdíl</i>	<i>MS</i>	<i>F</i>	<i>Hodnota P</i>	<i>F krit</i>
Řádky	4,103	3	1,368	6,177	0,006	3,287
Sloupce	0,848	5	0,170	0,766	0,588	2,901
Chyba	3,322	15	0,221			
Celkem	8,273	23				



# Dvoufaktorová ANOVA – 1. příklad v Excelu - výsledek

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

**H0: faktor řidič nemá na spotřebu vliv**

**H1: faktor řidič má na spotřebu vliv**

**Výsledek: H0 nelze zamítnout**

**H0: faktor benzín nemá na spotřebu vliv**

**H1: faktor benzín má na spotřebu vliv**

**Výsledek: H0 zamítáme**

Za současného působení faktorů faktor **řidiči** nemá na spotřebu vliv,  
faktor **benzín** vliv má!



# Dvoufaktorová ANOVA – 2. příklad v Excelu



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Data → Analýza dat → ANOVA: Dva faktory s opakováním...

	benzin/řidič Aral	Shell	Benzina	Slovnaft
A	7,5	7,6	7,2	7
	7,7	7,4	7,6	7,4
	8	7,3	8,1	7,7
B	6,9	7,2	8,1	7,3
	6,7	7,4	8,5	7,6
	6,6	7,6	8,8	7,8
C	7,9	7,5	7,8	7,2
	8	7,8	7,7	7,1
	8,3	8,1	7,6	7
D	7,3	8	7,6	7,5
	7,2	8	7,8	7,7
	7,1	7,9	8	7,8
E	6,9	7,3	7,8	8,2
	6,8	7,2	8	8,1
	6,7	7	8,1	8
F	7,8	8,2	6,9	7,7
	7,7	8,4	7,5	7,7
	7,5	8,5	7,9	7,7



Faktor	Aral	Shell	Benzina	Slovnaft	Celkem
<b>A</b>					
Počet	3	3	3	3	12
Součet	23,2	22,3	22,9	22,1	90,5
Průměr	7,73	7,43	7,63	7,37	7,54
Rozptyl	0,06	0,02	0,20	0,12	0,10
<b>B</b>					
Počet	3	3	3	3	12
Součet	20,2	22,2	25,4	22,7	90,5
Průměr	6,73	7,40	8,47	7,57	7,54
Rozptyl	0,02	0,04	0,12	0,06	0,46
<b>C</b>					
Počet	3	3	3	3	12
Součet	24,2	23,4	23,1	21,3	92
Průměr	8,07	7,80	7,70	7,10	7,67
Rozptyl	0,04	0,09	0,01	0,01	0,16
<b>D</b>					
Počet	3	3	3	3	12
Součet	21,6	23,9	23,4	23	91,9
Průměr	7,20	7,97	7,80	7,67	7,66
Rozptyl	0,01	0,00	0,04	0,02	0,10
<b>E</b>					
Počet	3	3	3	3	12
Součet	20,4	21,5	23,9	24,3	90,1
Průměr	6,80	7,17	7,97	8,10	7,51
Rozptyl	0,01	0,02	0,02	0,01	0,33
<b>F</b>					
Počet	3	3	3	3	12
Součet	23	25,1	22,3	23,1	93,5
Průměr	7,67	8,37	7,43	7,70	7,79
Rozptyl	0,02	0,02	0,25	0,00	0,19



## Dvoufaktorová ANOVA – 2. příklad v Excelu - výsledek



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

ANOVA						
Zdroj variací	SS	Rozdíl	MS	F	Hodnota P	F krit
Výběr	0,69	5	0,14	2,64	0,03	2,41
Sloupce	2,08	3	0,69	13,23	0,00	2,80
Interakce	10,23	15	0,68	12,99	0,00	1,88
Dohromady	2,52	48	0,05			
Celkem	15,53	71				

Za současného působení mají oba faktory:  
**řidiči** a **benzín** na spotřebu vliv!



# Nealko nápoje - studie

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

## Faktory:

$X_1$  - Pohlaví (2 kategorie: Muž, Žena)

$X_2$  - Typ vybrané odměny (2 kategorie: Peníze nebo Nápoje)

## Znak:

$Y$  - pravděpodobnost zakoupení nové limonády  
(kvantitativní data)







Každá hodnota prvního faktoru se kombinuje s každou  
hodnotou druhého faktoru vícekrát (15x) –

## **2F ANOVA s opakováním**



# Nealko nápoje - studie

Data → Analýza dat →

ANOVA: Dva faktory s opakováním...

	peníze	nápoje
M u ž	20	75
	50	50
	0	30
	0	60
	50	100
	20	30
	50	100
	0	50
	0	90
	15	90
	0	50
	10	100
	50	100
	0	50
	50	75
Ž e n a	70	80
	0	50
	0	90
	20	90
	90	100
	0	100
	0	70
	0	100
	50	50
	0	80
	20	100
	0	50
	50	90
	100	75
	20	90



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ



# Nealko nápoje – studie – výpočet v Excelu



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Anova: dva faktory s opakováním						
Faktor	peníze	nápoje	Celkem			
<i>Muž</i>						
Počet	15	15	30			
Součet	315	1050	1365			
Průměr	21	70	45,5			
Rozptyl	500,7143	667,8571	1185,086			
<i>Žena</i>						
Počet	15	15	30			
Součet	420	1215	1635			
Průměr	28	81	54,5			
Rozptyl	1245,714	343,5714	1493,707			
<i>Celkem</i>						
Počet	30	30				
Součet	735	2265				
Průměr	24,5	75,5				
Rozptyl	855,7759	519,569				
ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Výběr	1215	1	1215	1,762	0,190	4,013
Sloupce	39015	1	39015	56,587	0,000	4,013
Interakce	60	1	60	0,087	0,769	4,013
Dohromady	38610	56	689,46			
Celkem	78900	59				



# Nealko nápoje – formulace hypotéz

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

$H_0$ : výběr odměny nemá na pravděpodobnost nákupu vliv

$H_1$ : „opak  $H_0$ “

**Výsledek:**

**$H_0$  zamítáme; typ odměny má na pravděpodobnost nákupu vliv**

$H_0$ : pohlaví nemá na pravděpodobnost nákupu vliv

$H_1$ : „opak  $H_0$ “

**Výsledek:**

**$H_0$  nelze zamítnout; pohlaví nemá na pravděpodobnost nákupu vliv**



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Obliba Coca Coly (2 faktory) - studie

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

## Faktory:

$X_1$  – Pohlaví (2 kategorie: Muž, Žena)

$X_2$  – Vzdělání (7 kategorií: 1=Z, ..., 7=VŠ)

## Znak:

$Y$  – spokojenost s cenou nápoje  
(1=velmi nespokojen až 7=velmi spokojen)



# Obliba Coca Coly (2 faktory) - studie

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Každá hodnota prvního faktoru se kombinuje s každou  
hodnotou druhého faktoru vícekrát (4x) –

**2F ANOVA s opakováním**



## Řešení úlohy v CETIL (spotřeba vody) – zadání

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Ve třech městech okresu Karviná jsme v jednotlivých dnech sledovali průměrnou spotřebu pitné vody (v m<sup>3</sup>) na jednoho obyvatele. Zjistěte, zda je průměrná spotřeba vody závislá na dni v týdnu, a je-li spotřeba v různých městech různá. Uvažujte hladinu významnosti 0,05. Zjištěné údaje jsou uvedeny v následující tabulce.



# Řešení úlohy v OETU – data



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

	Karviná	Petřvald	Bohumín
Po	0,6	0,7	0,5
Út	0,7	0,6	0,6
St	0,9	0,8	0,7
Čt	0,6	0,6	0,5
Pá	1	1,3	0,8
So	1,2	1,6	1,3
Ne	1	1,2	1,3





# Řešení úlohy v OETU – formulace hypotéz



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

- **Formulace první dvojice hypotéz:**

$H_0$ : spotřeba pitné vody nezávisí na dnu v týdnu,

$H_1$ : spotřeba pitné vody závisí na dnu v týdnu.

- **Formulace druhé dvojice hypotéz:**

$H_0$ : spotřeba pitné vody nezávisí na městě,

$H_1$ : spotřeba pitné vody závisí na městě.



# Řešení úlohy v GRETU – zadání hodnot do programu



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

gretl: ukázat data

	spotreba	mesto	den
1	0,6	1	1
2	0,7	1	2
3	0,9	1	3
4	0,6	1	4
5	1,0	1	5
6	1,2	1	6
7	1,0	1	7
8	0,7	2	1
9	0,6	2	2
10	0,8	2	3
11	0,6	2	4
12	1,3	2	5
13	1,6	2	6
14	1,2	2	7
15	0,5	3	1
16	0,6	3	2
17	0,7	3	3
18	0,5	3	4
19	0,8	3	5
20	1,3	3	6
21	1,3	3	7



# Řešení úlohy v OETU – Testování první dvojice hypotéz.



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
V OPAVĚ

gretl: specifikovat model

ANOVA

spotreba  
mesto  
den

Závisle proměnná  
spotreba

Nastavit jako výchozí

Cílová proměnná  
den

Bloková proměnná (nepovinné)  
mesto

Nápověda Vymazat Zrušit Budiz

gretl: ANOVA

Analýza rozptylu, reakce = spotreba, úprava = den:

	Součet čtverců	df	Střední kvadrát
Úprava	1,75905	6	0,293175
Blok	0,092381	2	0,0461905
Reziduum	0,220952	12	0,0184127
Úplné	2,07238	20	0,103619

$F(6, 12) = 0,293175 / 0,0184127 = 15,9224$  [p-hodnota 4,42e-005]



## Řešení úlohy v OETU – výsledek

---



SLEZSKÁ  
UNIVERZITA  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Výsledek:  $p$ -hodnota =  $4,42 \cdot 10^{-5}$  a tato hodnota je menší než hladina významnosti 0,05, proto nulovou hypotézu o nezávislosti spotřeby pitné vody na dnu v týdnu zamítáme.

Můžeme tedy tvrdit, že spotřeba pitné vody z 95 % závisí na dnu v týdnu.



# Řešení úlohy v OETU – Testování druhé dvojice hypotéz.



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

gretl: specifikovat model

ANOVA

spotreba  
mesto  
den

Závisle proměnná  
spotreba  
 Nastavit jako výchozí

Cílová proměnná  
mesto

Bloková proměnná (nepovinné)  
den

Nápověda Vymazat Zrušit Budiž

gretl: ANOVA

Analýza rozptylu, reakce = spotreba, úprava = mesto:

	Součet čtverců	df	Střední kvadrát
Úprava	0,092381	2	0,0461905
Blok	1,75905	6	0,293175
Reziduum	0,220952	12	0,0184127
Úplné	2,07238	20	0,103619

$F(2, 12) = 0,0461905 / 0,0184127 = 2,50862$  [p-hodnota 0,1230]



## Řešení úlohy v OETU – výsledek

---



**SLEZSKÁ  
UNIVERZITA**  
OBCHODNĚ PODNIKATELSKÁ  
FAKULTA V KARVINĚ

Výsledek:  $p$ -hodnota = 0,123 a tato hodnota není menší než hladina významnosti 0,05, proto nulovou hypotézu o nezávislosti spotřeby pitné vody na městě nelze zamítnout.

Z 95 % **nebylo prokázáno**, že by spotřeba pitné vody závisela na městě.





## ANOVA

1-faktorová nebo 2-faktorová  
je účinným analytickým nástrojem  
při analýze kvalitativních/kvantitativních dat  
získaných z dotazníkových šetření!





# Děkuji Vám za pozornost!!!

