

Statistické zpracování dat 2.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

Téma: Analýza rozptylu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Co je analýza rozptylu - ANOVA
- ANOVA v marketingu a managementu
- Jednofaktorová ANOVA
- Míry závislosti: determinační a korelační poměr
- Použití ANOVA v případových studiích



Co je analýza rozptylu?



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Analýza rozptylu - ANOVA –
z anglického **AN**alysis **Of** **V**ariance
(Jednofaktorová = One-way)

- ANOVA je rozšířením parametrických (a
neparametrických) testů hypotéz na více než 2 výběry



ANOVA řeší 2 problémy (dvojí interpretace):



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

1. Ovlivňuje kvalitativní faktor kvantitativní hodnoty znaku?
2. Pochází $k (\geq 3)$ vzorků ze stejné populace?

Příklady:

- (1) Má věk respondentů vliv na konzumaci daného nápoje?
- (2) Ovlivňuje vzdělání respondentů jejich názor na poslance v parlamentu?



Marketingová studie 1



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Firma chce na trhu se sycenými nealkoholickými nápoji uplatnit svůj nový výrobek (nápoj K)
- Na fakultě je bufet a několik prodejních automatů s nealko nápoji
- Fakulta má cca 4000 studentů a 400 zaměstnanců
- Byl vytvořen dotazník a na jehož základě byla získána data - vzorek 100 studentů a zaměstnanců



Marketingový výzkum oblíbenosti sycených perlivých nápojů (Studie nealkoholických nápojů)

Nový výrobek: **K**

DOTAZNÍK



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

1. Pohlaví:.....(1) Muž(2) Žena
2. Věk.....let
3. Kolik lahví/sklenic/plechovek sycených perlivých nápojů spotřebujete během obvyklého týdne?.....(Eventuálně přepočítejte na jednotku o obsahu 330 ml)
4. Kolik dnů během obvyklého týdne vypijete alespoň jednu láhev/sklenku/plechovku některého syceného perlivého nápoje?.....
5. Jak důležitá je cena při nákupu syceného perlivého nápoje?
(1) je extrémně důležitá....., (2) je velmi důležitá....., (3) je trochu důležitá.....,
(4) není ani důležitá ani nedůležitá....., (5) je víceméně nedůležitá....., (6) je velmi nedůležitá....., (7) je absolutně nedůležitá.....
6. Jak důležitá je kvalita značky při nákupu syceného perlivého nápoje?
(1) je extrémně důležitá....., (2) je velmi důležitá....., (3) je trochu důležitá.....,
(4) není ani důležitá ani nedůležitá....., (5) je víceméně nedůležitá....., (6) je velmi nedůležitá....., (7) je absolutně nedůležitá.....
7. Poté, co jste ochutnal/a **K**, ohodnoťte jeho chuť pomocí bodů na škále 0 až 100 bodů (0 bodů vyjadřuje odpornou nepoživatelnou chuť, 100 bodů fantastickou chuť).....
8. Uspořádejte následujících 5 sycených perlivých nápojů, přičemž 1 označuje váš nejoblíbenější nápoj, 2 druhý nejoblíbenější nápoj atd., 5 je nejméně oblíbený nápoj:
.....Coca Cola
.....Fanta
.....7-Up
.....Pepsi Cola
..... **K**
9. Stanovte pravděpodobnost, že byste si koupil/a 6 nebo více lahví (330 ml) **K** za cenu 30 Kč za 1 balení po 6 ks. Použijte přitom škálu 0 až 100 bodů, přičemž 0 označuje nulovou pravděpodobnost koupě (tj. koupě je absolutně vyloučena), 100 bodů značí koupě s naprostou jistotou.....
10. Jako výraz našeho ocenění vaší účasti v marketingovém výzkumu od nás obdržíte buď 30 Kč, nebo zdarma 1 balení (6 ks) **K**. Co si vyberete?
.....(1) 30,- Kč(2) 6 ks **K**.



Příklad 1



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Znak X tzv. faktor - věk studentů
(18 až 22 let: $k = 5$)
- Znak Y - konzumace limonády
(počet lahví/týden - kvantitativní data)
- H_0 : Věk nemá na konzumaci limonád vliv
(jinak řečeno: skupiny vytvořené podle věku pocházejí z jediné populace $\Rightarrow \mu_k = \mu_0$)
- H_1 : „opak H_0 “





Idea:

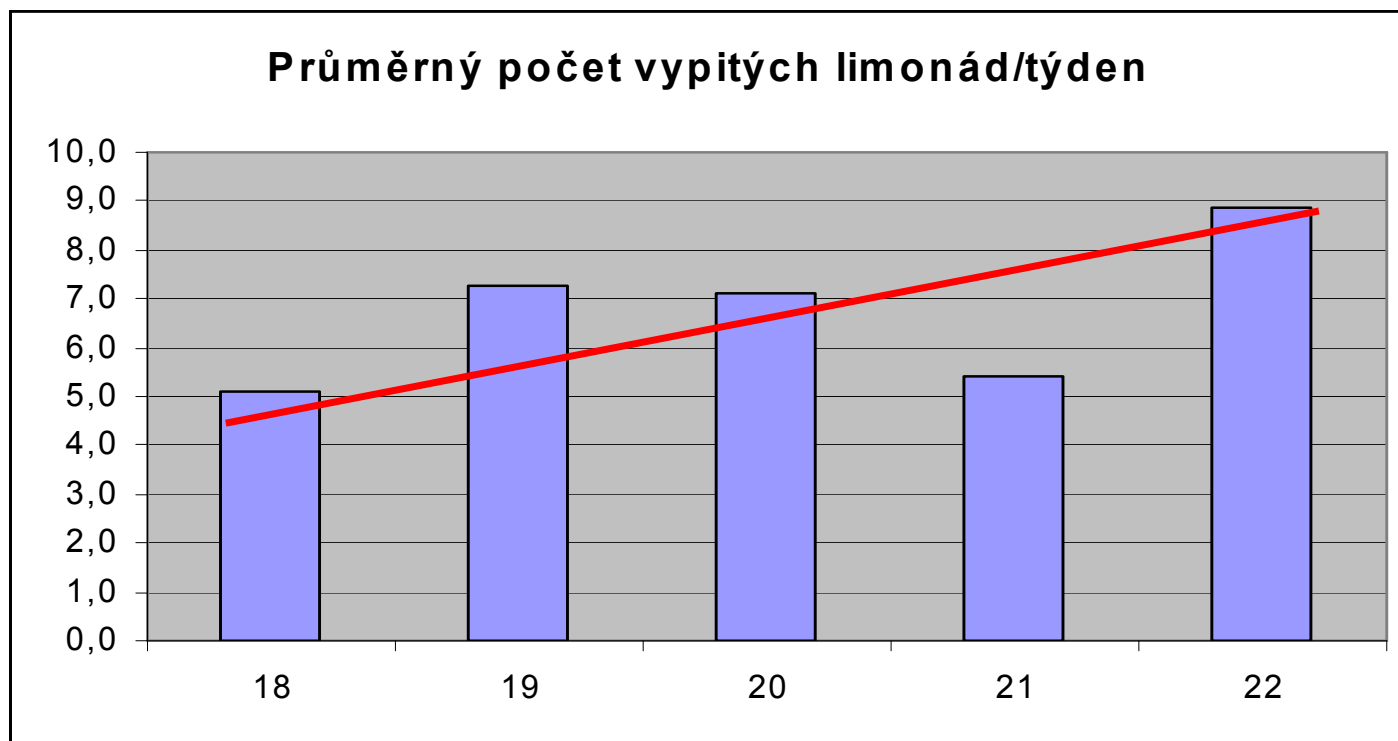
Čím větší variabilita (tj. rozptyl) mezi skupinami, tím větší vliv faktoru!!!



Počet vypitých limonád roste s věkem?!



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Jednofaktorová ANOVA - postup



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Krok 1.

Uspořádání dat a výpočty

X – kvalitativní znak (faktor), Y - kvantitativní znak

$$\bar{y} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \quad s^2 = \frac{\sum_i \sum_j (y_{ij} - \bar{y})^2}{n-1}$$

Stanovení nulové hypotézy

$H_0: \mu_i = \mu_0$ pro všechna $i = 1, 2, \dots, k$



Krok 2. Výpočet



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Celkový součet čtverců:
$$S_y = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

Meziskupinový součet čtverců:
$$S_{y,m} = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$

Vnitroskupinový součet čtverců:
$$S_{y,v} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$



Krok 2. Výpočet



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Základní vztah ANOVA:

$$S_y = S_{y,m} + S_{y,v}$$

Výpočet testového kritéria:

$$F = \frac{\frac{S_{y,m}}{k-1}}{\frac{S_{y,v}}{n-k}}$$



Krok 3. Testování



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Jestliže platí: $F > F_{1-\alpha}(k-1, n-k)$ pak

H_0 zamítáme (faktor má vliv)

jinak H_0 nezamítáme (faktor nemá vliv),

přitom $F_{1-\alpha}(df_1, df_2)$ je kritická hodnota Fisherova rozdělení

(tabelováno pro různé hodnoty df_1, df_2 a α).





Alternativně:

Pro hodnotu kritéria F vypočítáme p -hodnotu (signifikanci)

Jestliže platí:

p -hodnota $< \alpha$ pak H_0 zamítáme (faktor má vliv)

jinak H_0 nezamítáme (faktor nemá vliv)

Předpoklad ANOVA: normálně rozdělený znak Y



Řešení příkladu v Excelu



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

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

Anova: jeden faktor

Faktor

Výběr	Počet	Součet	Průměr	Rozptyl
18	9	46	5,111111	54,11111
19	18	131	7,277778	70,80065
20	21	149	7,095238	36,89048
21	15	81	5,4	44,82857
22	8	71	8,875	222,6964

ANOVA

Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Mezi výběry	95,13097	4	23,78274	0,344165	0,847153	2,510831
Všechny výběry	4560,785	66	69,1028			
Celkem	4655,915	70				

$p\text{-hodnota} = 0,847 > 0,05 \Rightarrow H_0$ nezamítáme, tzn. věk nemá na konzumaci limonád vliv!



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



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Těsnost závislosti v ANOVA



Faktor X má k kategorií, sledovaný znak Y je kvantitativní

Poměr korelace P :
$$P = \sqrt{\frac{S_{y,m}}{S_y}}$$

kde S_y - celkový součet čtverců
$$S_y = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

$S_{y,m}$ - meziskupinový součet čtverců
$$S_{y,m} = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2$$



Těsnost závislosti v ANOVA



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Poměr determinace P^2

Čím je P bližší k 1, tím je závislost sledovaného znaku na daném faktoru silnější,
čím je blíže k 0, tím je závislost slabší.



Studie spokojenosti / oblíbenosti



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Průzkum spokojenosti zákazníků v supermarketech v ČR (**Spokojenost v Hypernově, TESCO**)
- Průzkum oblíbenosti výrobku/služby (**Obliba Coca Coly**)
- 7 bodové hodnoticí škály (spokojenosti)
- Byl vytvořen dotazník a na jehož základě byla získána data - vzorek 200 zákazníků



Spokojenost v supermarketech Hypernova



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Znak *X* - faktor – věková kategorie zákazníků (1 až 5)
(1=18až21, 2=22až29, 3=30až44, 4=45až59, 5=60+)
- Znak *Y* - postoj na škále
1= velmi nespokojen až 5= velmi spokojen např. k otázkám:
 - (1) Celková spokojenost
 - (2) Příjemná obsluha
 - (3) Čistota prodejen



Hypotéza ke spokojenosti v supermarketech Hypernova



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

H_0 : věk nemá na spokojenost k otázce vliv

H_1 : „opak H_0 “



Obliba Coca Coly



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- 1. Znak X - faktor – vzdělání zákazníků (1 až 7)
(1=Z, 2=Z+, 3=SOU, 4=SOU+, 5=SŠ, 6=SŠ+, 7=VŠ)
- 2. Znak X - faktor – příjem zákazníků(1 až 9)
(1=6000Kč, 2=6001Kč až10000Kč,...,9 =24000Kč a více)
- Znak Y - na škále 1= velmi nespokojen až 5= velmi spokojen
např. k následujícím otázkám:
 - (1) Spokojenost s cenou nápoje
 - (2) Spokojenost s prestiží nápoje
 - (3) Spokojenost s kvalitou nápoje



Hypotéza k oblíbenosti Coca Coly



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

H_0 : věk nemá na spokojenost k otázce vliv

H_1 : „opak H_0 “



Sociologická studie



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Sociologický průzkum voličů v ČR
- Průzkum názorů a postojů
(5 a 7 bodové hodnoticí škály: od zcela nesouhlasím po zcela souhlasím)
- Populace 7 mil. voličů
- Byl vytvořen dotazník a na jeho základě byla získána data - vzorek 1000 voličů - občanů ČR nad 18 let



Sociologická studie



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Znak X - faktor - vzdělání voličů (ot. 15)
($k = 4$: 1=základní, 2=vyučen, 3=maturita, 4=VŠ)
- Znak Y - postoj na škále 1= velmi nesouhlasím až
5= velmi souhlasím např. k následujícím otázkám:





- (1) Lidé, které volíme do parlamentu jsou darebáci (ot. 8g)
- (2) Stát má lidem zajistit práci (ot. 8n)
- (3) Rodina je základem společnosti (ot. 12g)



Hypotéza k sociologické studii



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

H_0 : vzdělání nemá na postoj k otázce vliv

H_1 : „opak H_0 “



Řešení sociologické studie v Excelu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

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

Anova: jeden faktor						
Faktor						
Výběr	Počet	Součet	Průměr	Rozptyl		
Zakl	144	393	2,729	1,695		
Vyuč	301	886	2,944	1,500		
SŠ	424	1169	2,757	1,286		
VŠ	128	393	3,070	1,168		
ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Mezi výběry	14,609	3	4,870	3,492	0,015	2,614
Všechny výběry	1384,823	993	1,395			
Celkem	1399,432	996				
P2 =	0,010					
P =	0,102					

p -hodnota = 0,015 < 0,05 $\Rightarrow H_0$ zamítáme, tzn. vzdělání má na postoj k otázce (1) vliv!



Kruskal – Wallisova ANOVA



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- X - kvalitativní znak (faktor),
- Y - kvantitativní nebo ordinální znak

- T - testové kritérium

$$T = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{P_j^2}{n_j} - 3(n+1)$$



Kruskal – Wallisova ANOVA



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

$$T = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{P_j^2}{n_j} - 3(n+1)$$

- n - počet všech respondentů
- n_j - počet respondentů j -té kategorie
- k - počet kategorií faktoru
- P_j - součet pořadí v j -té třídě (příslušné j -té kategorii)
- H_0 - faktor neovlivňuje pořadí
- $df = k-1$
- α – hladina významnosti (obvykle 0,05)
- T - se porovnává s kritickou hodnotou $CHIINV(\alpha ; df)$



Příklad byty: Má počet pokojů vliv na cenu bytu?



xi	yij						nj
1	221	255	264	201			4
2	321	329	374	335	309		5
3	592	543	526	505	528	564	6
4	640	621	687	643	672		5
5	934	948	962	895			4
							n = 24

pořadí:

xi	yij					
1	2	3	4	1		
2	6	7	9	8	5	
3	15	13	11	10	12	14
4	17	16	20	18	19	
5	22	23	24	21		

xi	Pj	Pj ²	Pj ² /nj
1	10	100	25
2	35	1225	245
3	75	5625	937,5
4	90	8100	1620
5	90	8100	2025

Suma: 4852,5

$$k = 4$$

$$T = 22,05$$

$$\text{CHIDIST}(T; k-1) = 0,0002$$



Příklad byty: Má počet pokojů vliv na cenu bytu?



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Výpočet testového kritéria pomocí vzorce:

$$T = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{P_j^2}{n_j} - 3(n+1)$$

Výpočet p-hodnoty pomocí funkce

$$\text{CHIDIST}(T; k - 1) = \text{CHIDIST}(22,05; 4) = 0,0002$$

Nulovou hypotézu zamítáme!

Faktor Počet pokojů má vliv na Cenu!





Děkuji Vám za pozornost!!!

