Kvantitativní zpracování dat

Radoslav Škapa

Typy dat

- Neparametrické
  - Nominální (nominal) – např. pohlaví
  - Ordinální (ordinal) – např. preference vyjádřené na škálách, sociální třídy, stupeň vzdělání, toto třídění proměnných z hlediska množství obsažené informace.
- Metrické (parametrické)
  - Intervalové (interval) – např. teplota, Likertovy škály – intervaly jsou mezi stupni stejně velké. Nemá ale smysl mluvit o tom, že je např. 2x větší teplota (10 vs. 20 stupňů C). V sociálních výzkumech spíš zřídka.
  - Poměrové (ratio) – např. věk, obrat. Existuje nula. Mnoho statistických testů nerozlišuje mezi intervalovými a poměrovými proměnnými

Analýza dat

Popisná statistika

Jedno a dvourozměrná analýza

Vícerozměrné analýzy

Interpretace

Třídění 1., 2. a 3. stupně

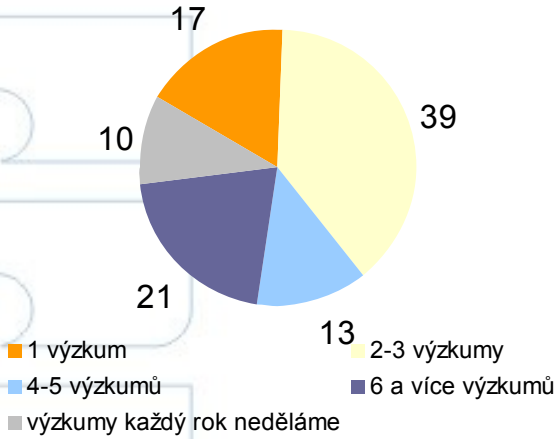
Více např. zde: www.lsvv.eu/workshop/dytrt/dytrt_prezentace.ppt

Ukazatelé polohy

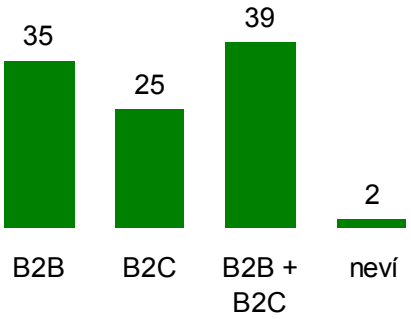
Typ proměnné	Přípustné operace
Nominální	Modus
Ordinální	Modus, medián
Intervalové	Modus, medián, průměr
Poměrové	Modus, medián, průměr

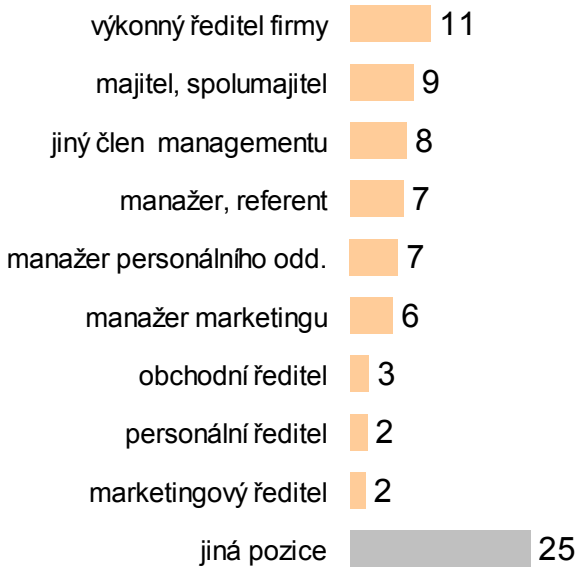
- U ordinálních by se neměl počítat průměr. U Likertových škál lze.
- Je vhodné sledovat všechny ukazatele polohy.

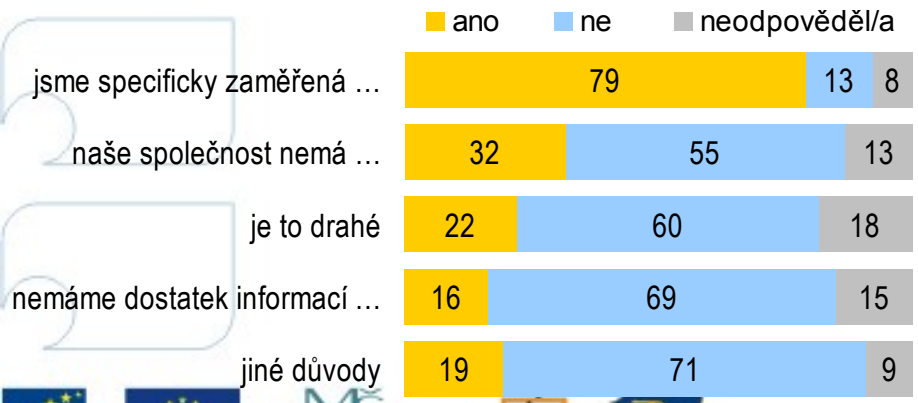
PRVOSTUPŇOVÉ TŘÍDĚNÍ – UKÁZKY

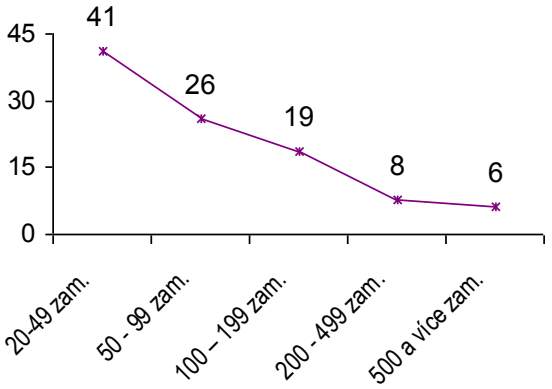


důvody nespokojenosti	počet
nebylo podle našich představ	4
neznají náš trh, nejdou do hloubky	1
byrokracie, zdržuje nás to	1









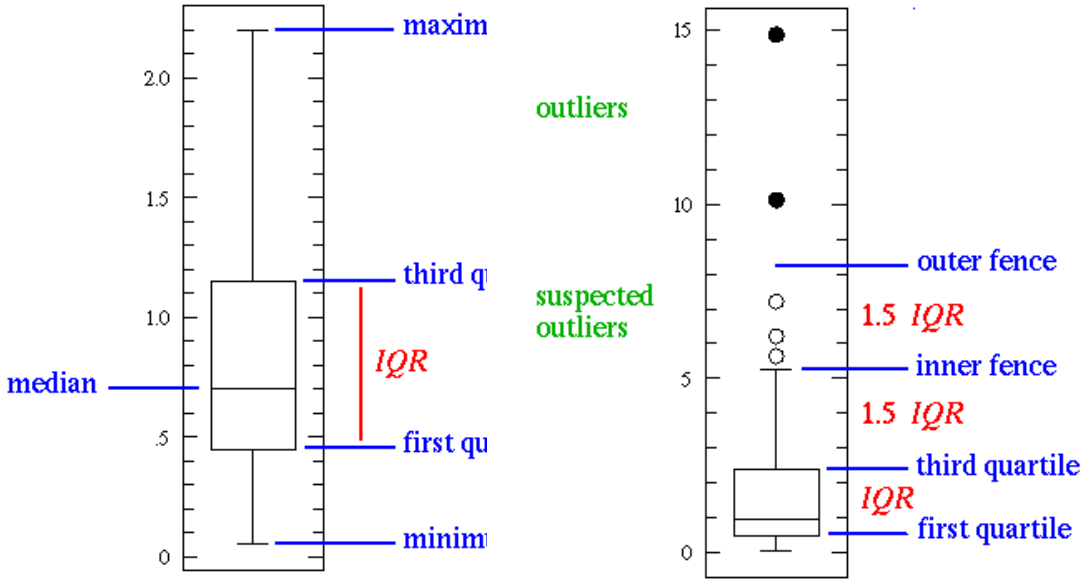
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

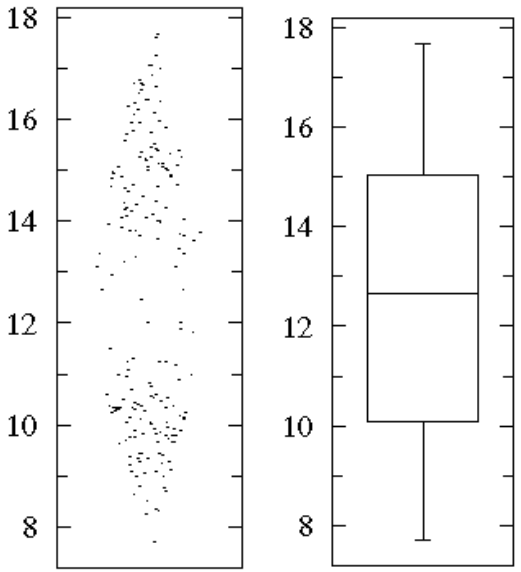
PRVOSTUPŇOVÉ TŘÍDĚNÍ – ALTERNATIVNÍ VÝSTUPY

Odpovědi na ot. 14: „Z jakého důvodu jste dosud nerealizovali žádný výzkum?“

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Krabicový diagram





http://www.physics.csbsju.edu/stats/box2.html

DRUHOSTUPŇOVÉ TŘÍDĚNÍ

- Mapuje **vzájemné souvislosti proměnných** - porovnáváme distribuci dat závisle proměnné na základě kategorií nezávislé proměnné.
- Výstupy z SPSS získáme např. pomocí procedur **General Tables, Tables of Frequencis, Crosstabs.**
- Prováděná druhostupňová třídění by měla korespondovat s našimi hypotézami a výzkumnými záměry, případně prezentovat zajímavá a významná zjištění.

esf EVROPSKÁ UNIE MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

DRUHOSTUPŇOVÉ TŘÍDĚNÍ

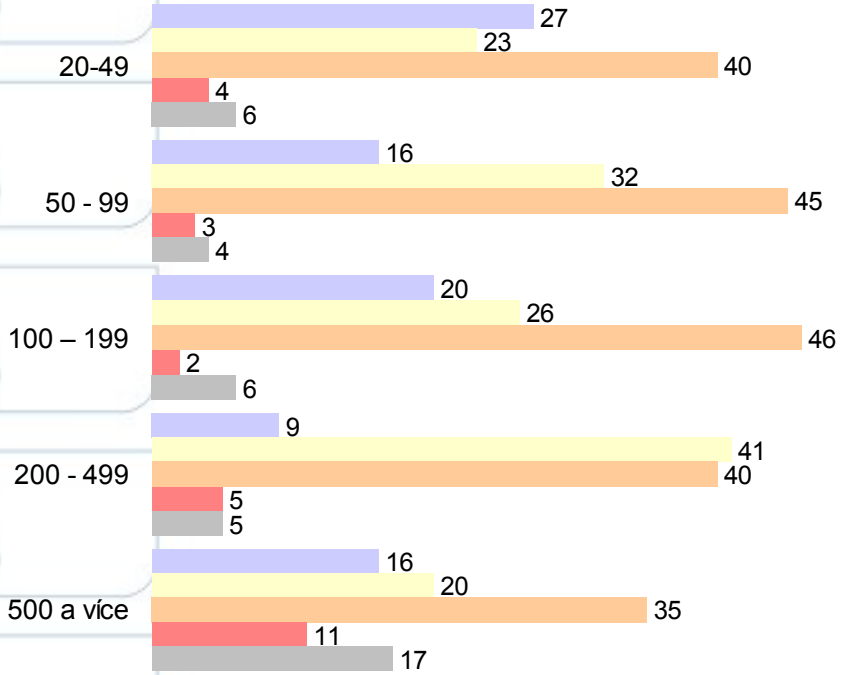
- Klasickými nezávislými proměnnými pro druhostupňová třídění jsou **demografické a socioekonomické charakteristiky respondentů** (pohlaví, věk, vzdělání, velikost obce bydliště, region, příjem, ekonomická aktivita …).
- Výstupy jsou prezentovány v různých typech grafů; záleží na typu proměnných.
- Ve výzkumu postoje firem k výzkumům byly jako nezávislé proměnné používány hlavně počet zaměstnanců organizace a působení v B2B / B2C sektoru.

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

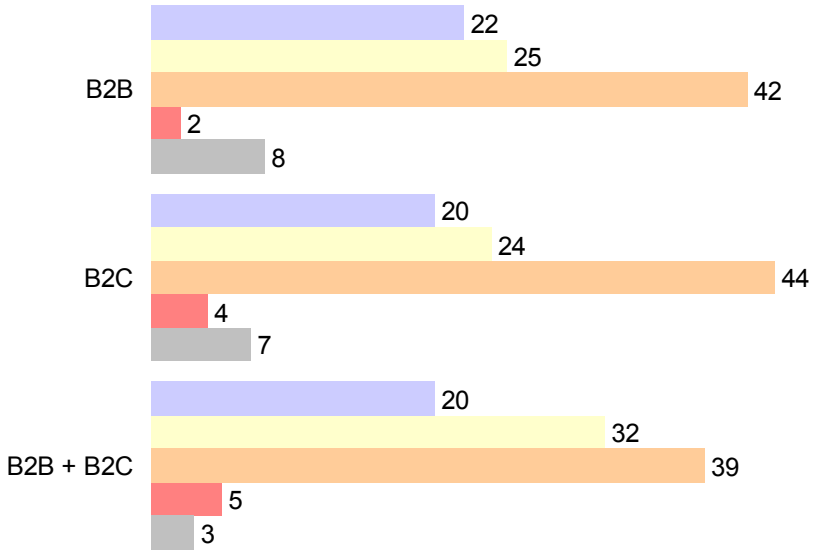
DRUHOSTUPŇOVÉ TŘÍDĚNÍ – UKÁZKA

„Můžete mi prosím říct zda, vaše firma v souvislosti s finanční a ekonomickou krizí přistoupil/a ke krácení investic do marketingového výzkumu?“ (q24)

počet zaměstnanců



podnikání v B2B / B2C segmentu



velmi jsme je pokrátili
trochu jsme je pokrátili
zůstaly na stejné úrovni jako před krizí
investice jsme zvýšili
neví, bez odpovědi

%, N = 272, provedli alespoň jeden výzkum

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Vztahy mezi proměnnými

- Nalezení vztahů je obecným finálním cílem každého výzkumu
- Dvě dimenze vztahu:
  - Velikost (síla) vztahu– hodnocení na výzkumníkovi. Obecně ve společenských vědách se za silné vazby považují už nižší koeficienty asociace (např. 0,7) něž přírodní vědy. Příklad Pearsonův produktový koeficient korelace.
  - Spolehlivost (reliabilita, pravdivost) – pravděpodobnost, že výsledek není náhodný. Spolehlivost s jakou lze výsledek zobecnit na základní soubor. Měří se pomocí „p-value“ (statistical significance) – pravděpodobnosti chyby. Např. p-value=0,05 znamená 95% spolehlivost.

Vztahy mezi proměnnými

- Z jiného pohledu: p-value=0,05 znamená např. že cca při 20 měřeních korelací nesouvisejících proměnných nám jedna vyjde spolehlivá (tzv. chyba 1. typu).
  - Čím více textů provedeme na datech, tím více „chybných“ vztahů objevíme.
- Existuje (pozitivní) vztah mezi sílou a spolehlivostí vypočteného vztahu (příklad porodnice)
- Ve stejně velkém vzorku, silnější vztahy víc spolehlivé.
- K prokázání slabých vztahů je třeba velké vzorky. (K prokázání neexistence žádného vztahu – prozkoumat téměř celou populaci). (příklad – slabě vychýlená mince).
  - $\Rightarrow$ ve velkých vzorcích i slabé vztahy budou statisticky významné – proto při interpretaci se vždy zamyslet, zda je takový vztah dostatečně silný, aby mělo smysl o něm mluvit.
- Statisticky nevýznamné výsledky nejsou publikovány.

http://www.statsoft.com/textbook/elementary-concepts-in-statistics/

Jak se počítá spolehlivost?

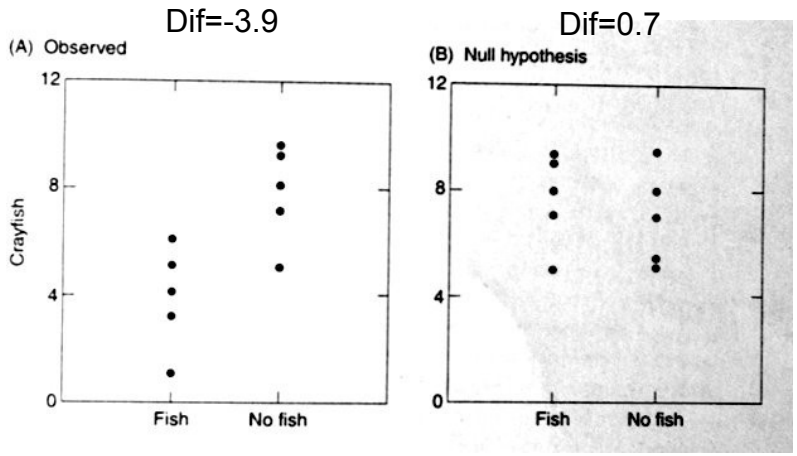


Table 5.1

Stream	Fish	Crayfish	FISH 1	FISH 2	FISH 3
1	+	1	+	+	+
2	–	5	–	+	–
3	+	3	+	+	+
4	–	7	–	–	–
5	+	4	+	–	–
6	–	8	+	–	+
7	+	5	–	–	–
8	–	9	+	–	+
9	+	6	–	+	+
10	–	9.5	–	+	–

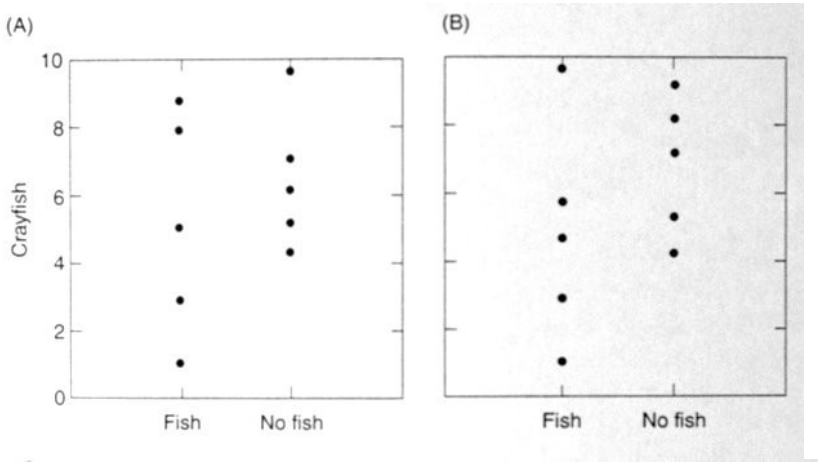
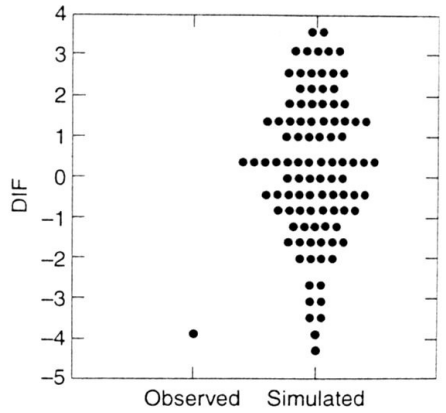


Table 5.2

Graph	DIF	Graph	DIF
1	2.70	11	2.1
2	1.90	12	2.7
3	–0.09	13	–3.1
4	2.50	14	1.5
5	–2.70	15	–1.1
6	1.70	16	–0.1
7	0.50	17	1.3
8	1.10	18	–0.5
9	1.50	19	–0.7
10	3.10	20	–0.3

Figure 5.3



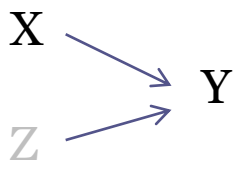
Magnusson, W. E., and G. Mourão. 2004. Statistics without math. Sinauer Associates, Londrina, Brasil

Zkreslení při použití korelací

- *Nepravá korelace*: místo (x→y) proměnné ovlivňovány třetí proměnnou

  Z →X
  Z →Y

- *Výjovová sekvence*: zdá se x→y, je z→x→y
- *Chybějící střední člen*: zdá se x→y, ale je x→z→y
- *Dvojí příčina*: závislá proměnná y má dvě příčiny, ale jen jedna, x, je zahrnuta do výzkumu



http://www.jakubholy.net/humanities/disman-soc_znalost.html

A GUIDE TO THE SELECTION OF STATISTICAL TOOLS

Number of Variables in the Analysis and Type of Data		Class of Statistics: Non-Parametric (Nominal; Ordinal)	Class of Statistics: Parametric (Interval; Ratio)
One Variable	1. Discrete	a. Percentage	(Impossible)
	2. Continuous	a. Median, Mode b. Quartile Range	a. Mean b. Std. Deviation
Two Variables	1. Both Discrete	a. Chi Square b. Phi Coefficient c. Contingency Coefficient d. Tetrachoric Corr. (2 rows x 2 columns) e. Yules' Q (2 x 2) f. Lambda	a. Correspondence analysis
	2. One Discrete and the other Continuous		a. Student t (if dependent variable is dichotomous) b. One-way ANOVA (if dependent variable is multichotomous)
	3. Both Continuous	a. Spearman Rank Correlation b. Kendalls' R c. Gamma	a. Pearson Correlation b. Simple Regression c. Eta Curvilinear Correlation

www.qualtrics.com/university/researchsuite/ebooks

Vícerozměrné statistické metody

Cíle:

- nalezení smysluplných pohledů na data popsaná velkým množstvím proměnných
- nalezení a popsání skrytých vazeb mezi proměnnými a tím zjednodušení jejich struktury
- jednoduchá vizualizace dat, kdy se v jediném grafu skrývá informace např. z 20 proměnných
- umožnění a/nebo zjednodušení interpretace dat na základě jejich zjednodušení a vizualizace

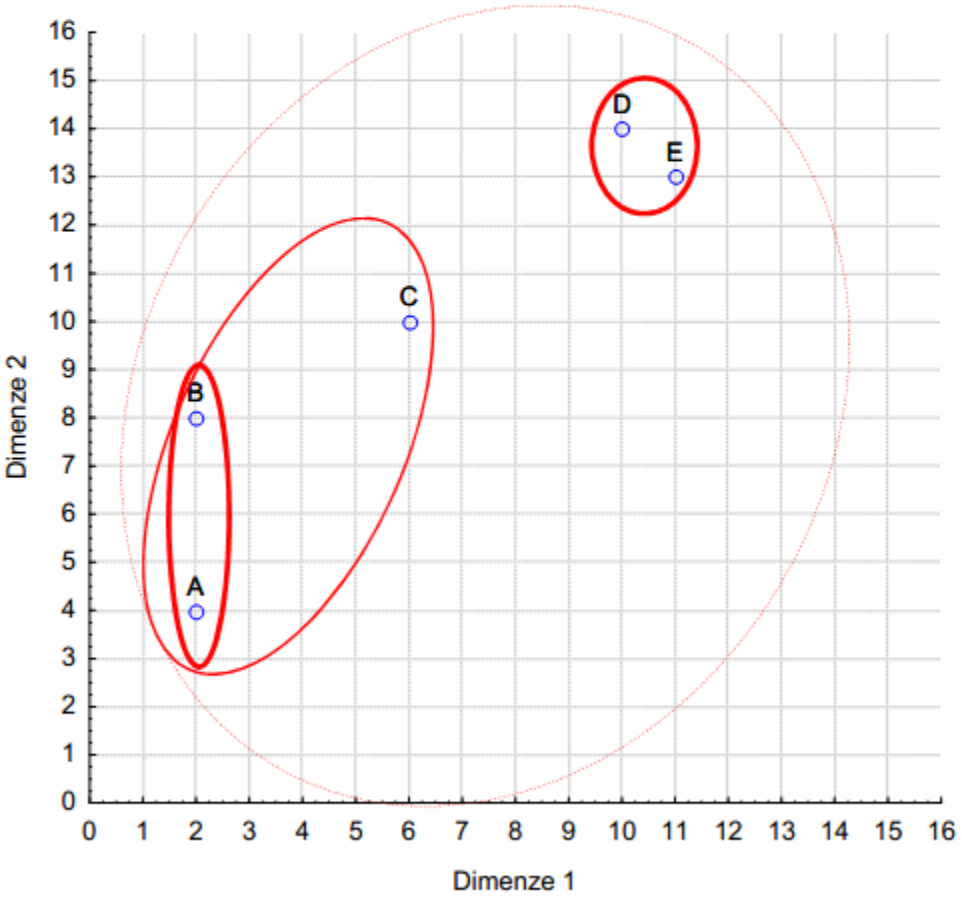
http://www.iba.muni.cz/res/file/ucebnice/jarkovsky-vicerozmerne-statisticke-metody.pdf

Vícerozměrné statistické metody

- Vícenásobná regrese *(Multiple regression)*
- Analýza hlavních komponent *(Principal Component Analysis)*
- Faktorová analýza *(Factor Analysis)*
- Shluková analýza *(Cluster Analysis)*
- Diskriminační analýza *(Discriminant Analysis)*
- Korespondenční analýza *(Correspondence analysis)*
- Kanonická korelace *(Canonical Correlation Analysis)*
- Vícerozměrné škálování *(Multidimensional Scaling)*
- Klasifikační stromy *(Classification Trees)*
- Pěšinková analýzy *(Path analysis)*
- Strukturní modelování *(Structrual equation modeling)*
- Preferenční analýza *(Conjoint analysis)*

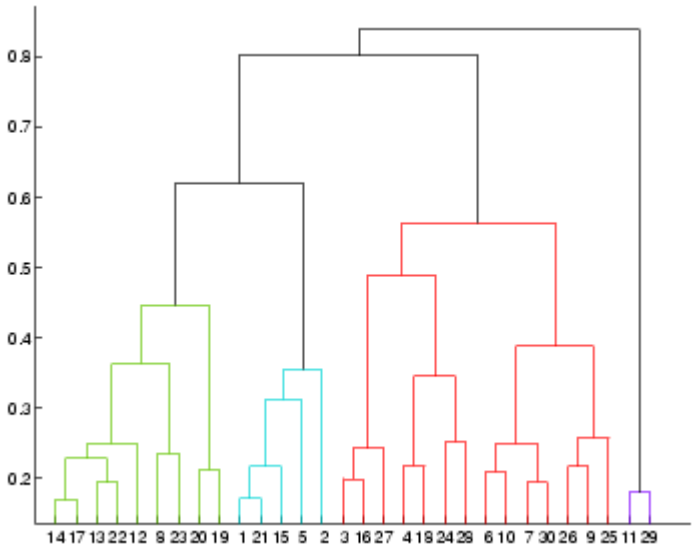
Shluková analýza: cíle a postupy

- Shluková analýza se snaží o identifikaci shluků objektů ve vícerozměrném prostoru a následnou redukce vícedimenzionálního problému kategorizací objektů do zjištěných shluků
- Existuje řada různých metod pro shlukování dat lišících se:
  - Měřením vzdálenosti mezi objekty
  - Algoritmem spojování objektů do shluků
  - Interpretací výstupů
- Každá z metod má své vlastní předpoklady výpočtu a je nasaditelná pro různé typy úloh
- Porušení předpokladů nebo nasazení chybné metody může vést k zavádějícím výsledkům

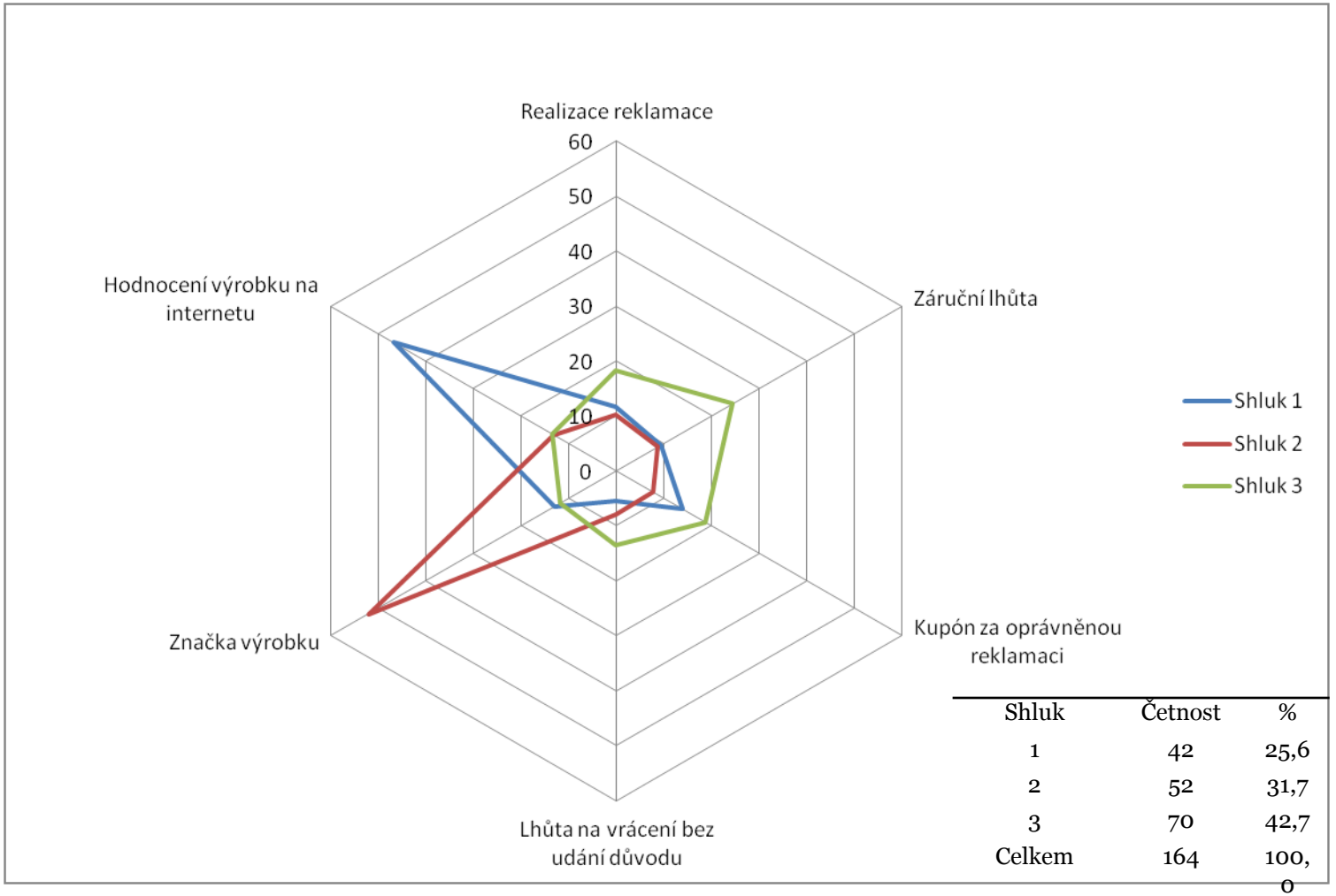


http://www.iba.muni.cz/esf/res/file/bimat-prednasky/vicerozmerne-statisticke-metody/VSM-05.pdf

Dendrogram shlukové analýzy



Důležitost atributů v % pro tři shluky



Shluk	Četnost	%
1	42	25,6
2	52	31,7
3	70	42,7
Celkem	164	100,0

Faktorová analýza

- **explorační** / konfirmační
- cíl: identifikace několika málo faktorů, které reprezentují vztahy ve větším počtu vzájemně souvisejících prom.
- metoda:
  - analýza korelací uvnitř sady proměnných
  - identifikace faktorů, různé úlohy:
    - popis vztahů mezi proměnnými pomocí faktorů
    - interpretace faktorů podle shluků silně korelovaných proměnných
  - vytvoření nových proměnných shrnujících variabilitu celé sady proměnných

http://metody.webz.cz/A07%20factor.ppt

Model pro faktorovou analýzu

- teoreticky zdůvodněný výběr proměnných
  (nikoliv výlov rybníka)
- předpoklad, že za sadou měřených proměnných je skrytá dimenze - faktor (1 - více) vysvětlující komplexnější jev
- měřené proměnné v sadě lze vyjádřit jako lineární kombinace faktorů, která nejsou přímo měřené a společné faktory zakládají některé vztahy mezi prom.

Logika dobré analýzy:

- cílem je sumarizace a simplifikace
- hledáme malý počet smysluplných - dobře interpretovatelných faktorů

http://metody.webz.cz/A07%20factor.ppt

	Creativity	Creative strategies	Strategies based on knowledge acquisition	Viability of business idea
Creativity				
I invent exceptional and surprising solutions	0.81			
My ideas are usually really unique	0.79			
When I encounter obstacles, I am able to detour around them	0.72			
I try to find novel solutions even if it is not expected from me	0.70			
I have a tremendous amount of ideas	0.69			
Creative opportunity search strategies				
I tried to find a really new business idea		0.84		
I purposefully emphasised creativity when generating the business idea		0.79		
I tried to find original and really novel ideas for a business		0.77		
I proposed and tried a lot of different ideas		0.61		
Strategies based on knowledge acquisition				
I gathered a lot of information on industries and sales etc. for the basis of the business idea			0.87	
I gathered a lot of information on markets for the basis of the business idea			0.84	
I did organised work on the business idea			0.79	
Viability				
I believe that our business idea would have a great chance of growth				0.86
I believe that the number of employees would grow rapidly with our business idea				0.82
I believe that the owners would earn fortunes with our business idea				0.78
I believe that the sales generated by our business idea would outstrip its potential customers				0.72
Eigenvalue	4.89	1.84	1.39	2.56
Percentage of variance	30.37	11.52	8.70	15.99

Notes: KMO 0.80, Bartlett's $p < 0.001$; the cut-off point was 0.50

Table I.
Exploratory factor analysis of the dependent and independent variables

Conjoitní analýza – preferenční analýza

Délka lhůt	Záruční lhůta
	Lhůta na vrácení bez udání důvodu
Transakční náklady	Realizace reklamace
	Kupón za oprávněnou reklamaci
Vnímaná kvalita (reference)	Značka výrobku
	Hodnocení výrobku na internetu

Vysavač A

Phillips

2 týdny na vrácení
Záruka 2 roky + 1 rok navíc
10% sleva na nákup při reklamaci
Reklamace zasláním poštou
Hodnocení: 1,8 bodů

Vysavač B

No name

4 týdny na vrácení
Záruka 2 roky
10% sleva na nákup při reklamaci
Reklamace u zákazníka
Hodnocení: 1,8 bodů

Vysavač C

No name

4 týdny na vrácení
Záruka 2 roky + 1 rok navíc
10% sleva na nákup při reklamaci
Reklamace zasláním poštou
Hodnocení: 2,3 bodů

Vysavač D

Phillips

4 týdny na vrácení
Záruka 2 roky + 1 rok navíc

Reklamace u zákazníka
Hodnocení: 1,8 bodů

Vysavač E

Phillips

2 týdny na vrácení
Záruka 2 roky
10% sleva na nákup při reklamaci
Reklamace u zákazníka
Hodnocení: 2,3 bodů

Vysavač F

No name

2 týdny na vrácení
Záruka 2 roky

Reklamace zasláním poštou
Hodnocení: 1,8 bodů

Vysavač G

No name

2 týdny na vrácení
Záruka 2 roky + 1 rok navíc

Reklamace u zákazníka
Hodnocení: 2,3 bodů

Vysavač H

Phillips

4 týdny na vrácení
Záruka 2 roky

Reklamace zasláním poštou
Hodnocení: 2,3 bodů

Vysavač CH

No name

4 týdny na vrácení
Záruka 2 roky + 1 rok navíc
10% sleva na nákup při reklamaci
Reklamace u zákazníka
Hodnocení: 2,3 bodů

Vysavač I

Phillips

2 týdny na vrácení
Záruka 2 roky + 1 rok navíc
10% sleva na nákup při reklamaci
Reklamace u zákazníka
Hodnocení: 2,3 bodů

Pořadí v jakém byste se rozhodli zakoupit jednotlivé vysavače:

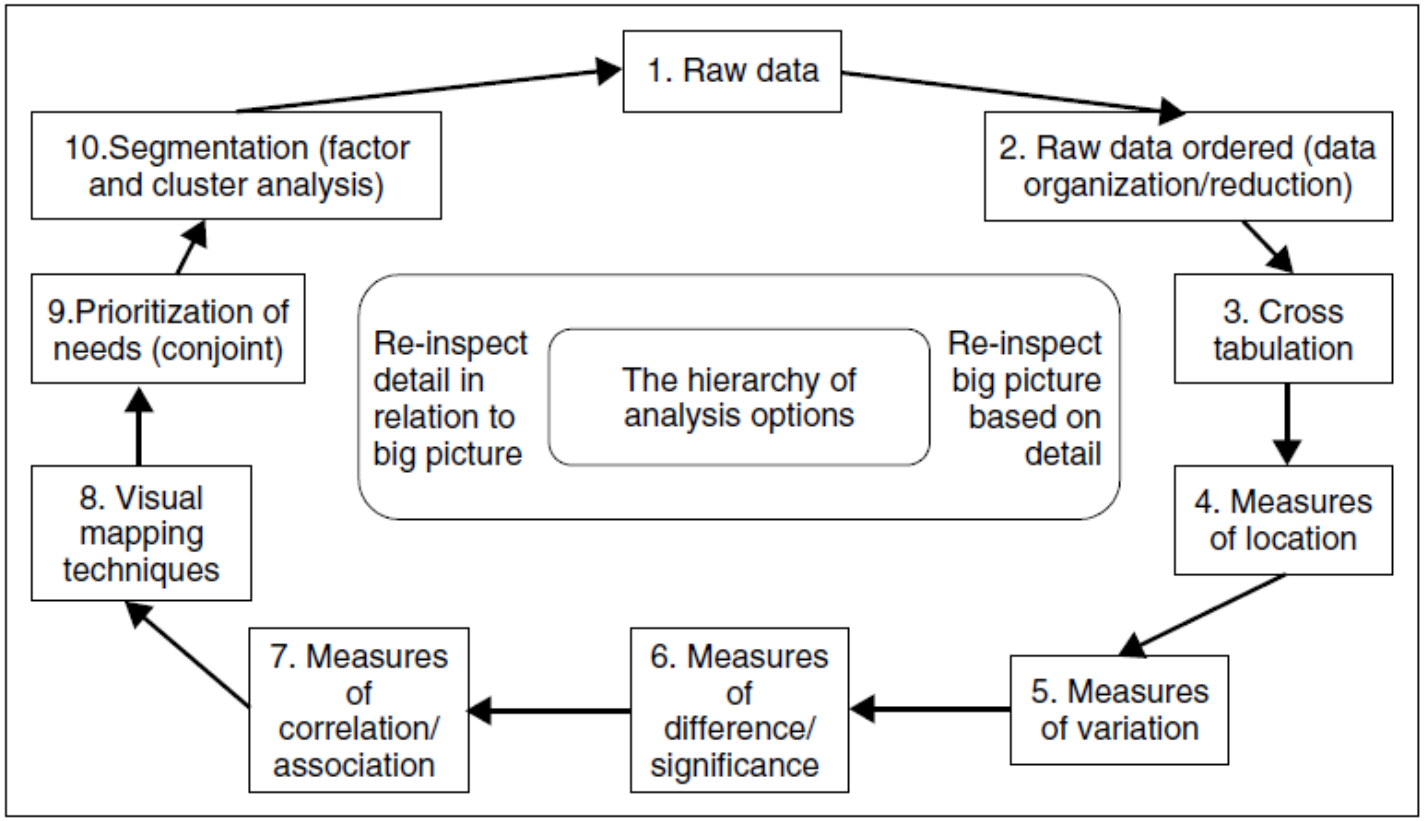
1. volba	2. volba	3. volba	4. volba	5. volba	6. volba	7. volba	8. volba	9. volba	10. volba

Conjoitní analýza

Spotřebitel se typicky nerozhoduje podle jedné vlastnosti produktu, ale podle kombinací vlastností, které jsou dosažitelné.

- Zjišťování preferencí ne podle izolovaných atributů ale podle atributů sledovaných současně ⇒
  - CONsider JOINTly
  - Věrněji se tak modeluje reálná situace nákupního rozhodování.
- Dotazovaný je nucen rozhodnout se mezi **dosažitelnými** kombinacemi vlastností (trade-offs).
- Lze ukázat, jak různé vlastnosti produktu predikují zákaznické preference po tomto produktu. Tj. jaký je s nimi spojen užitek.
  - Optimalizace stávajícího sortimentu/portfolia produktů ale i návrh nového produktu.
  - Odhad užitku lze spočítat pro jednotlivce i pro skupinu respondentů ⇒využitelnost pro segmentaci.

Příklad postupu statistického vyhodnocování dat



Smith, Fletcher: The Art & Science of Interpreting Market Research Evidence

Které analýzy v práci použít?

- Třídění 1. a 2. stupně + alespoň několik analýz 3. stupně:
  - Popisná statistika: (analýza četností, polohy, variablity)
  - Kontingenční tabulky
  - Rozdíly ve středních hodnotách (t-test, Mann-Whitney test – ordinální data)
  - Korelace (Pearson, Spearman (Kendall) – ordinální data)

- Nezapomenout na interpretaci výsledků
- Ideálně další a náročnější metody – vícerozměrná regrese, shluková analýza, diskriminační analýza, conjoint analyza, faktorová analýza. (Ty je třeba samostatně nastudovat, použít vhodně vzhledem k cíli a sestavit dotazník způsobem, abyste metodu mohli využít)

Software

- MS Excel – doplněk Analýza dat
  - **XLStatistics** http://www.deakin.edu.au/software/course.php?anchor=xlstatistics
- Statistica – licence MU
- SPSS – Multilicence MU
- Statgraphics – zaměřený spíš na průmysl. Výhodou jsou automatické komentáře k výsledkům.

Literatura

***IASTAT* - INTERAKTIVNÍ UČEBNICE STATISTIKY**

http://iastat.vse.cz/

Štatistický navigátor

http://rimarcik.com/navigator/

StatSoft, Inc. (2010). Electronic Statistics Textbook.

http://www.statsoft.com/textbook/

Petr Mareš, Ladislav Rabušic: Studijní materiály pro předět SOC708

https://is.muni.cz/auth/el/1423/podzim2005/SOC708/um/?info=1