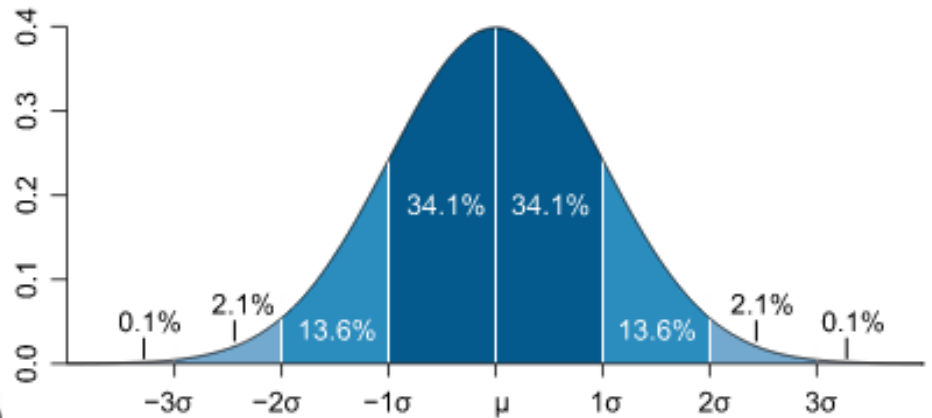


Štatistické metódy v senzorickej analýze



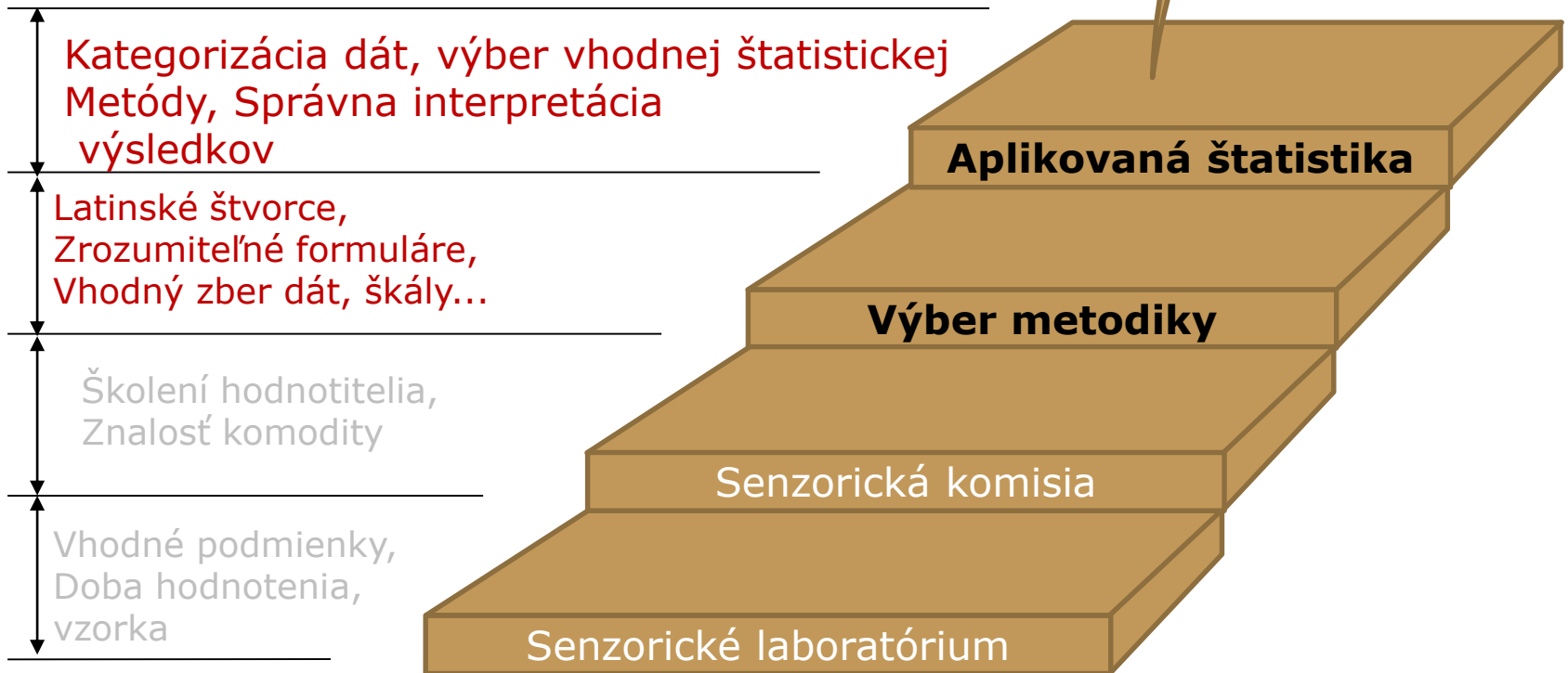
Vladimir Vietoris, PhD. Verzia 1.8.2018

Náplň prednášky

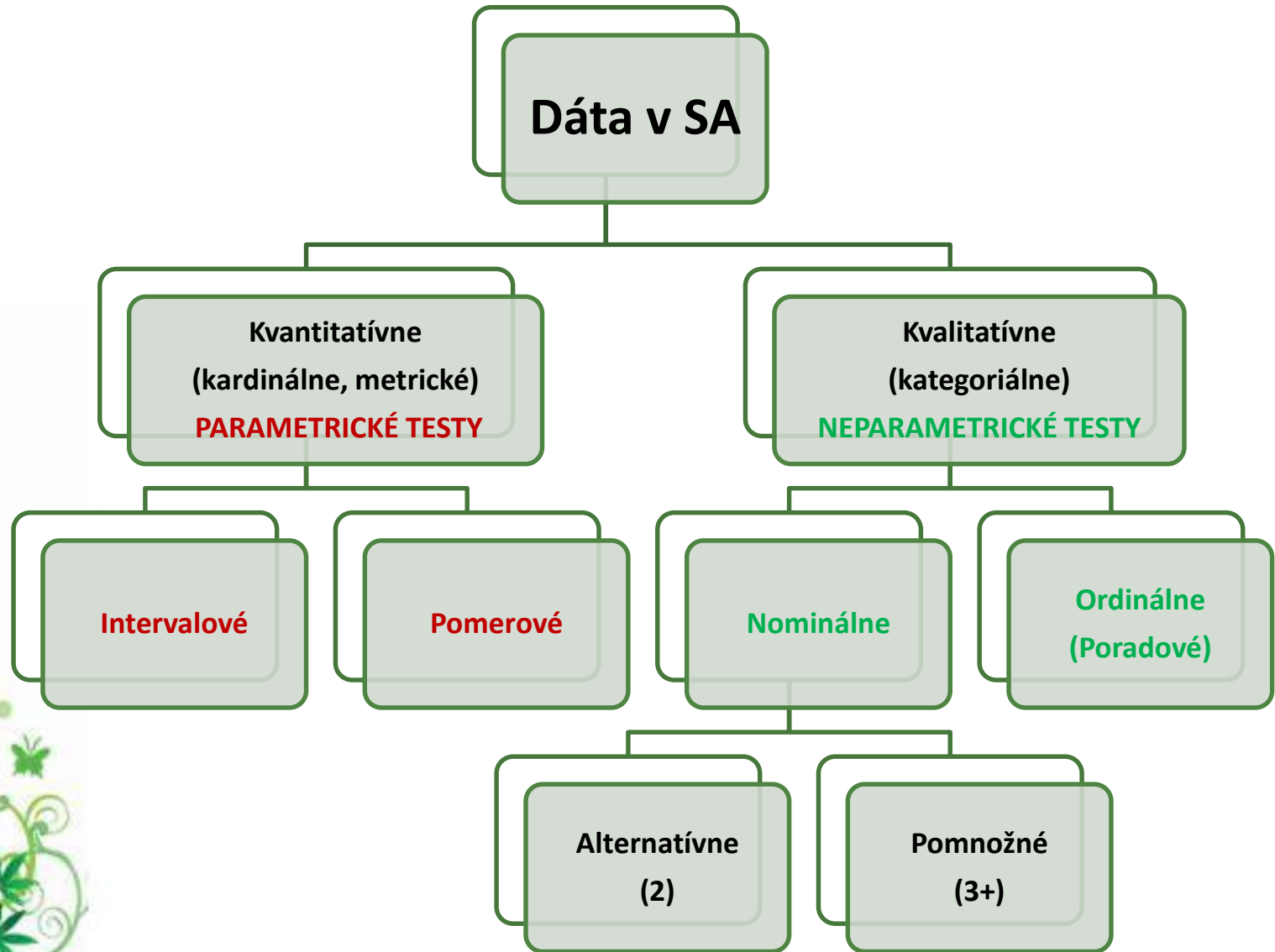
- Základné pojmy
- Rozdelenie/výber správnej metódy
- Parametrické a neparametrické m.
- Viacrozmerné metódy v SA



KOREKTNÉ VÝSLEDKY „GIGO“

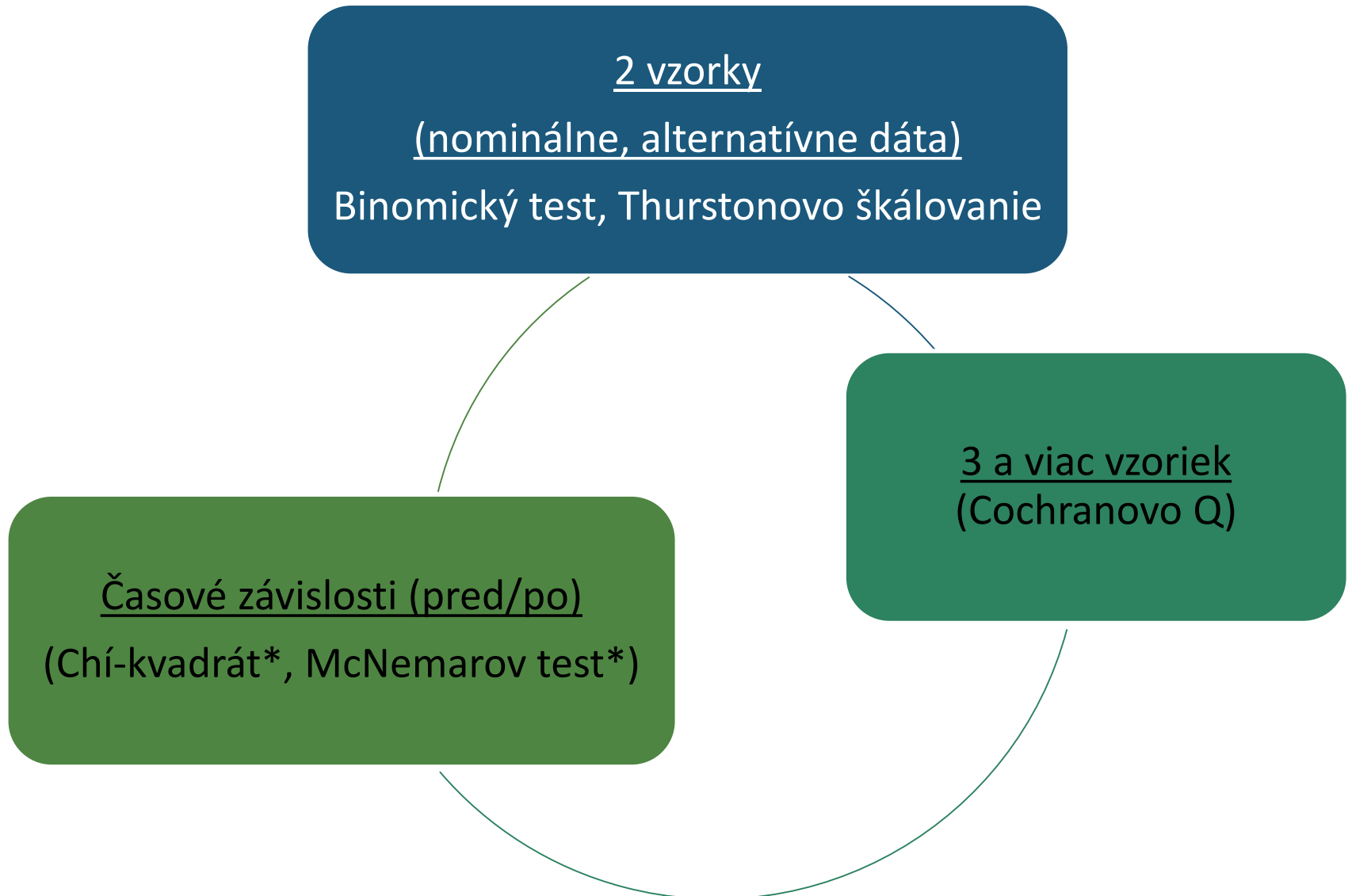


S akým typom dát pracujeme?



Kam patrí: Dáta z trianglového testu, 5 bodová Likertova škála, Celziova st. ?
Kedy je škála ešte nominálna a kedy už intervalová?

*Rozdielová metóda (Nominálne)



Definícia dôležitých pojmov

Závislé vzorky (výbery): sú vzorky hodnotené tou istou komisiou za tých istých podmienok (čas, teplota, laboratórium)

Nezávislé vzorky (výbery): sú vzorky, ktoré boli hodnotené rovnakou/rôznou komisiou za rovnakých/rôznych podmienok

Overenie Normality

- Q-Q plot
- Histogram
- Testy šikmosti a špicatosti
- Kolmogorov-Smirnov test
- Shapiro-Wilkov test
- H_0 - údaje pochádzajú z normálneho rozdelenia
- H_a – rozdelenie údajov nie je normálne

Aritmetický priemer (AP) – ukazovateľ strednej hodnoty (súčet prvkov/ich počet)
Modifikácie (windsorizovaný, prokrustovský, vážený)

Dosiahnuté skóre za čerstvosť	1	2	3	4	5
Počet hodnotiteľov ktorí prideli body	14	6	5	4	1

$$x_w = \frac{1 \cdot 14 + 2 \cdot 6 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 1}{14 + 6 + 5 + 4 + 1} = \frac{62}{30} = 2.0\bar{6}$$

- AP**
1. Nepoužívame ak je štatistické rozdelenie dát viacvrcholové
 2. Nepoužívame aj je štatistické rozdelenie dát asymetrické

Medián – stredná hodnota v súbore dát

Modus – najčastejšie sa opakujúca hodnota v súbore dát

Štandardná odchýlka vyjadruje do akej miery sú hodnoty rozptýlené od priemeru
(Aproximácia normálnym rozdelením z-skóre)

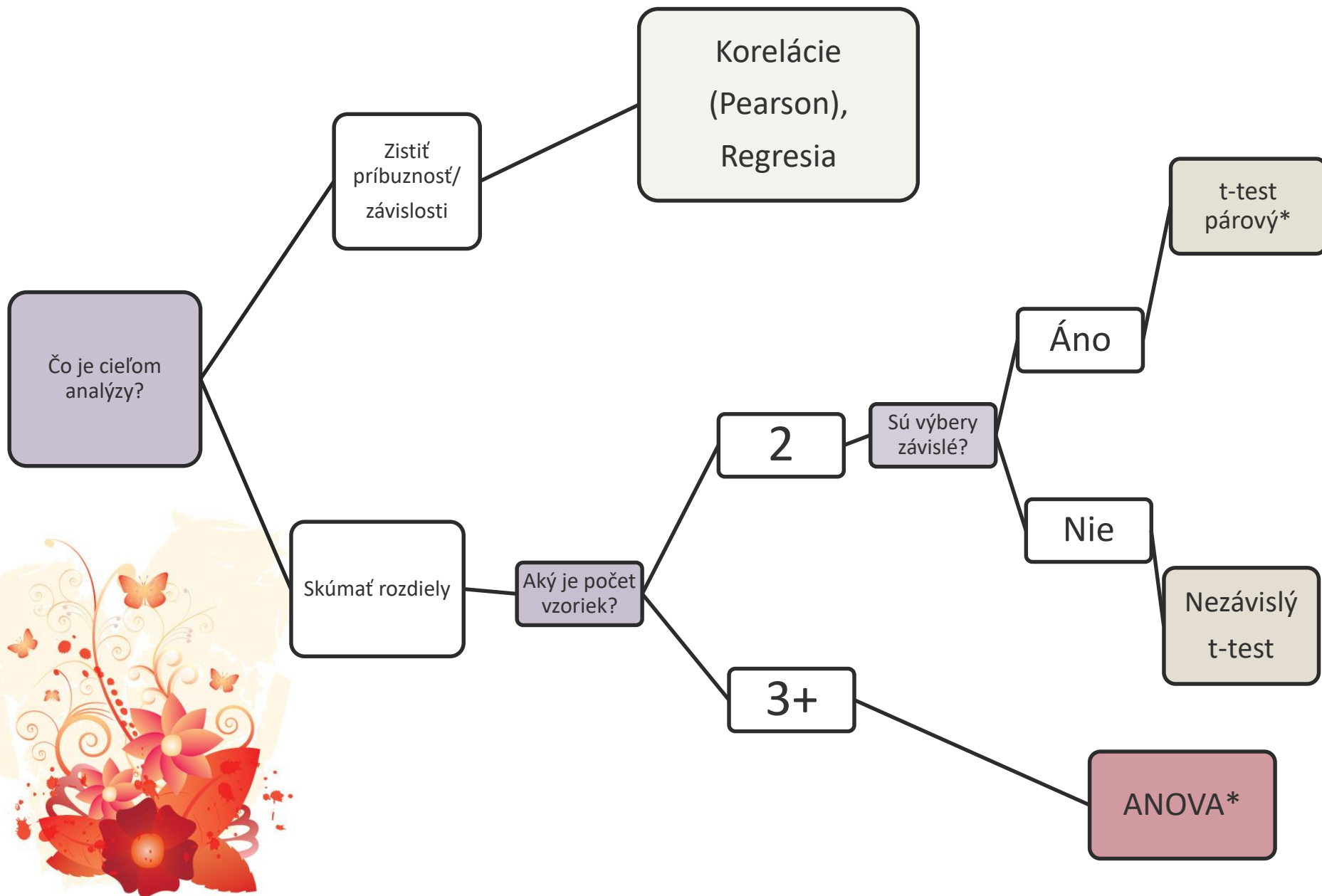
Parametrické testy :

- Sú založené na určitých predpokladoch, ktoré je potrebné splniť
- jedná sa predovšetkým o predpoklad rozdelenia náhodného výberu (Gaussovo rozdelenie)
- overujeme testom normality SHAPIRO-WILKOVÝM TESTOM

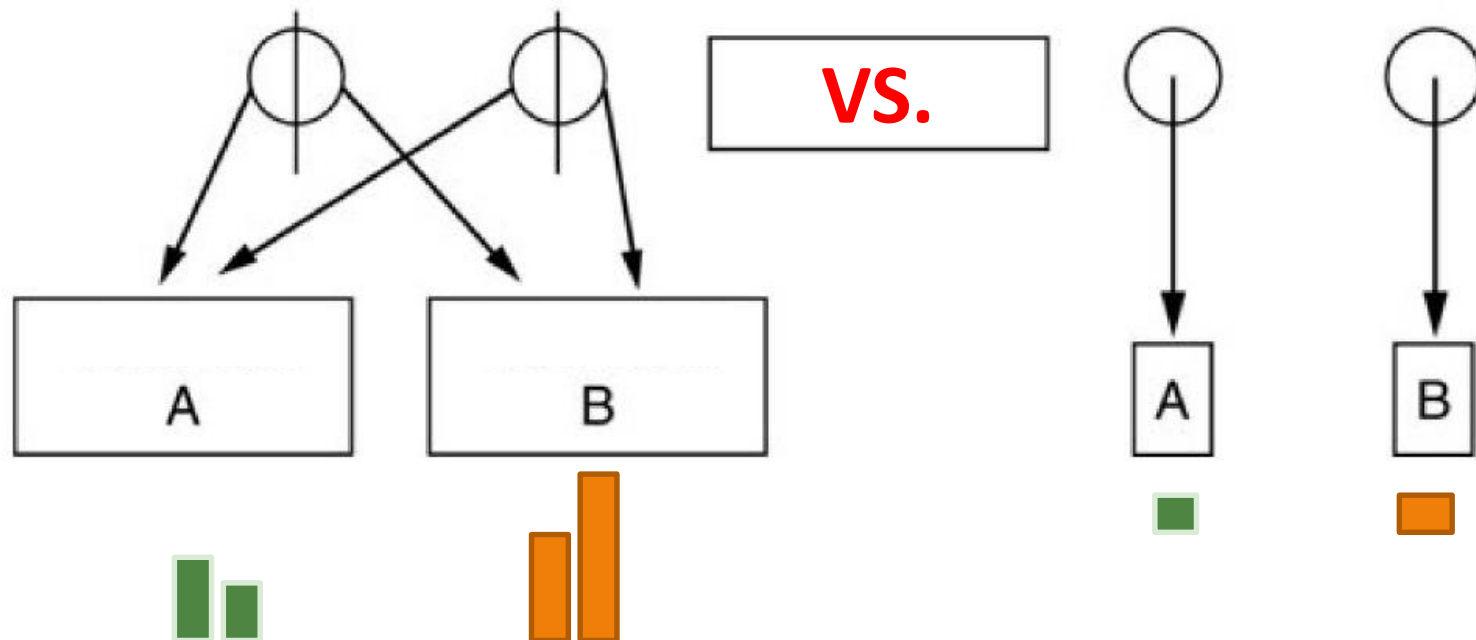
Často však existujú situácie, kedy podmienky pre použitie niektorého štandardného parametrického testu nie sú splnené:

- Malý rozsah výberu ($n < 30$)
- predpoklad jednovrcholového rozdelenia...

Výber vhodnej metódy parametrických testov



T-test a kolektív



Neparametrické testy

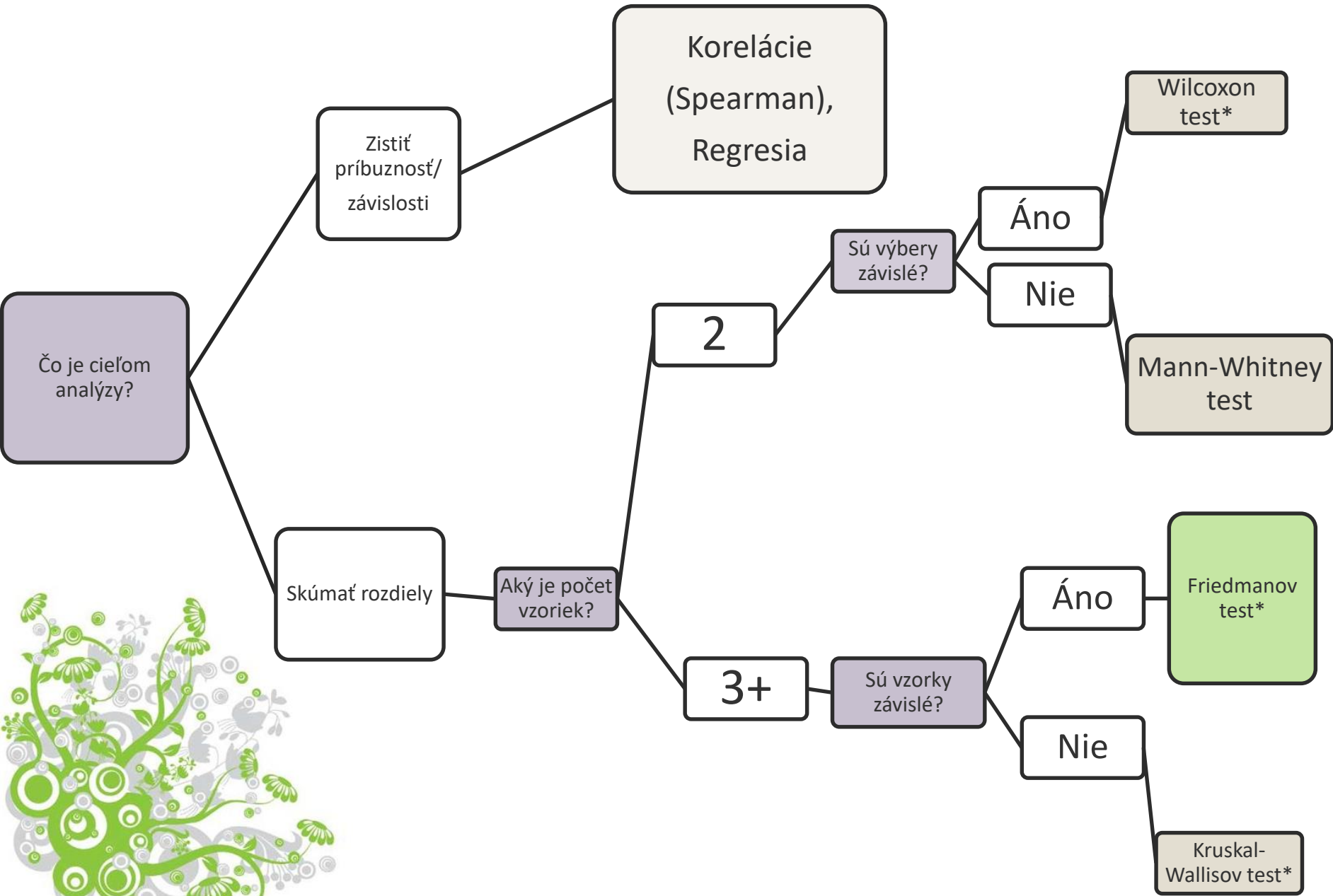
výhody:

- nie sú závislé od tvaru rozdelenia základného súboru
- Vychádzajú z opakovaní alebo poradí
- nevyžadujú typ jednovrcholového rozdelenia
- sú nenáročné na informácie (jednoduché výpočty testovacích charakteristík)
- použiteľnosť aj pre malé rozsahy výberov
- široká použiteľnosť

nevýhody:

- menšia sila testu v porovnaní s parametrickými testami
- v prípade kvantitatívnych znakov nevyužívajú celú informáciu o údajoch ALE LEN PORADIE ÚDAJOV....

Výber vhodnej metódy neparametrických testov



(intervalové dáta → poradové)

ANOVA 1F, t-test

	A	B	C	D
Adam	6,08	6,67	6,75	7,00
Eva	6,25	6,75	6,67	6,75
Bob	6,42	6,92	6,33	6,25
Iva	6,08	7,00	6,50	6,75
Edo	6,50	6,75	6,67	7,00
priemer	6,27	6,82	6,58	6,75

Prevod na poradie

Kruskal-Wallis

(1 faktorová nANOVA)

Friedman (2 faktorová nANOVA)

	Adam	Eva	Bob	Iva	Edo	
A	4	4	3	4	4	19
B	3	1,5	1	1	2	8,5
C	2	3	2	3	3	13
D	1	1,5	4	2	1	9,5

	A	B	C	D
Adam	1	11	10	18
Eva	5	14	13	14
Bob	8	16	4	6
Iva	2	18	8	14
Edo	7	12	3	18
	23	71	38	70

Profilová metóda (intervalové dáta)

Deskriptory (hodnotené znaky)

hodnotitelia

Produkty (hodnotené vzorky)

		Deskriptory (hodnotené znaky)			
		Jablková	Zeleninová	Zemitá	Cudzia
		Jablková	Zeleninová	Zemitá	Cudzia
		Jablková	Zeleninová	Zemitá	Cudzia
		Jablková	Zeleninová	Zemitá	Cudzia
ADAM		Jablková chuť	Zeleninová chuť	Zemitá chuť	Cudzia chuť...
A		6,08	6,42	6,67	6,75
B		6,25	6,08	6,75	6,50
C		6,42	6,50	6,92	6,67
D		6,08	6,50	7,00	6,75
E		6,50	6,75	6,67	7,00



Analýza hlavných
komponentov

Faktorová analýza

Viacrozmerné
metódy

Korešpondenčná
analýza

Multidimenzionálne
škálovanie (MDS)

Viacrozmerne metody

Počet premenných	Typ dát	
	Kvantitatívne	Kvalitatívne
Dve	Jednoduchá korelácia (Pearson, Spearman)	Analýza dvojrozmerných kontingenčných tabuliek (Chí, McNemar)
Viac ako dve	Analýza hlavných komponentov*	Multidimenzionálne škálovanie
	Faktorová analýza*	Korešpondenčná analýza*

Podmienky použitia viacrozmerných metód (PCA,FA)

	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_p
1												
2												
3												
4												
5...												
n												

$n > p$

opakované meranie dát (tých istých ľudí/produktov) je neprípustné
intervalové premenné

Analýza hlavných komponentov

Charakteristika

- predmetom analýzy je skupina kvantitatívnych premenných (pôvodných profilových dát)
- je metóda, ktorá umožňuje vytvárať nové premenné, ktoré sú lineárnou kombináciou pôvodných premenných
- nové premenné sa nazývajú hlavné komponenty
- Maticové operácie (transpozícia, rotácie...)
- Korelačná/ kovariančná matica

84,7%

```
> pca = princomp(data, cor=T)
```

```
> summary(pca, loadings=T)
```

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.5589391	0.9804092	0.6816673	0.37925777
Proportion of Variance	0.6075727	0.2403006	0.1161676	0.03595911
Cumulative Proportion	0.6075727	0.8478733	0.9640409	1.00000000

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4
senzorickakvalita	-0.523		0.848	
cena	-0.177	0.977	-0.120	
designobalu	0.597	0.134	0.295	-0.734
znacka	0.583	0.167	0.423	0.674

```
> plot(pca$scores[,1])
```

```
> |
```

Pozícia deskriptorov (loadings)

84,7%

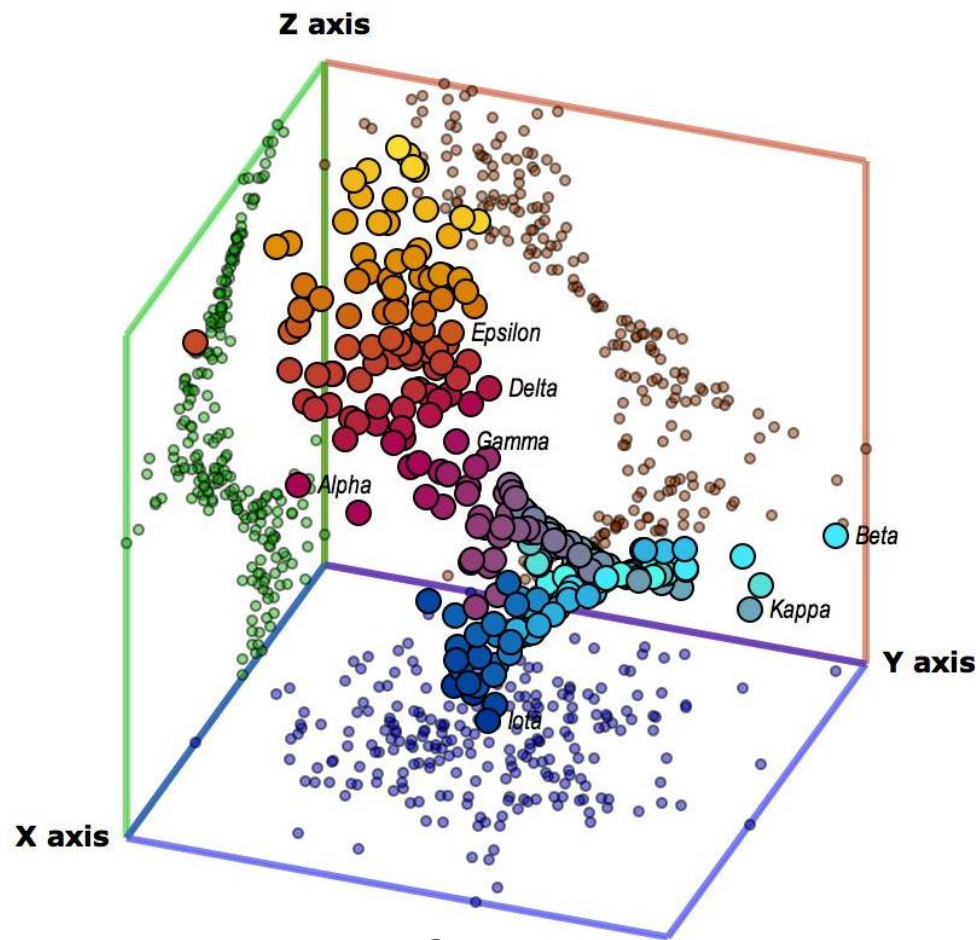
> data

	senzorickakvalita	cena	designobalu	znacka	PC1	PC2
1	6	5	3	4	1.1452093	-0.20585347
2	7	3	2	2	2.1357933	-1.82692136
3	6	4	4	5	0.3753926	-0.72038791
4	5	7	1	3	2.0936930	0.91650704
5	7	7	5	5	0.6905686	1.39486696
6	6	4	2	3	1.6681044	-1.05285950
7	5	7	2	1	2.4622708	0.79115490
8	6	5	4	4	0.8371646	-0.13681365
9	3	5	6	7	-1.5523737	0.28114087
10	1	3	7	5	-1.9363942	-1.21356043
11	2	6	6	7	-1.6817513	0.95800676
12	5	7	7	6	-0.7695088	1.62233387
13	2	4	5	6	-1.2823169	-0.56976950
14	3	5	6	5	-0.8757512	0.08674892
15	1	6	5	5	-0.9499225	0.69067064
16	2	3	7	7	-2.3601783	-1.01526413

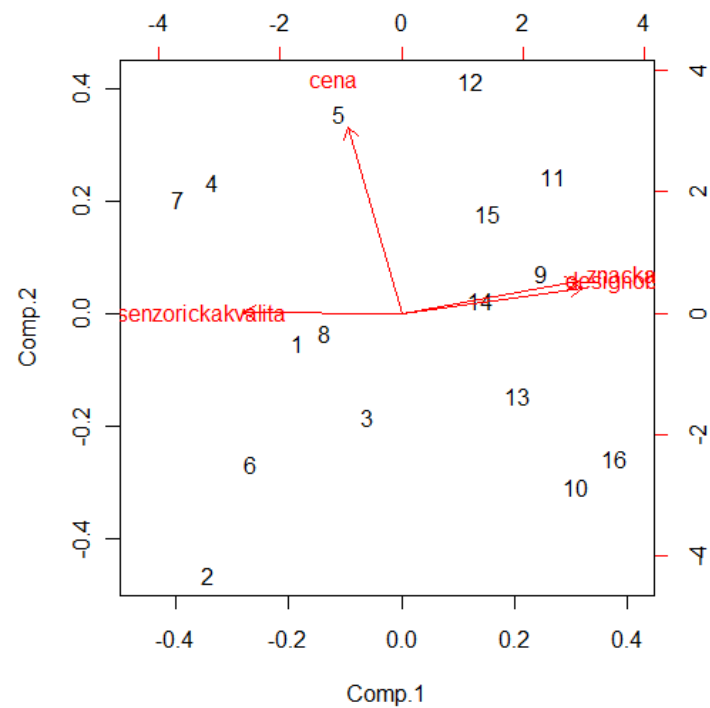
Pôvodné dáta

HK

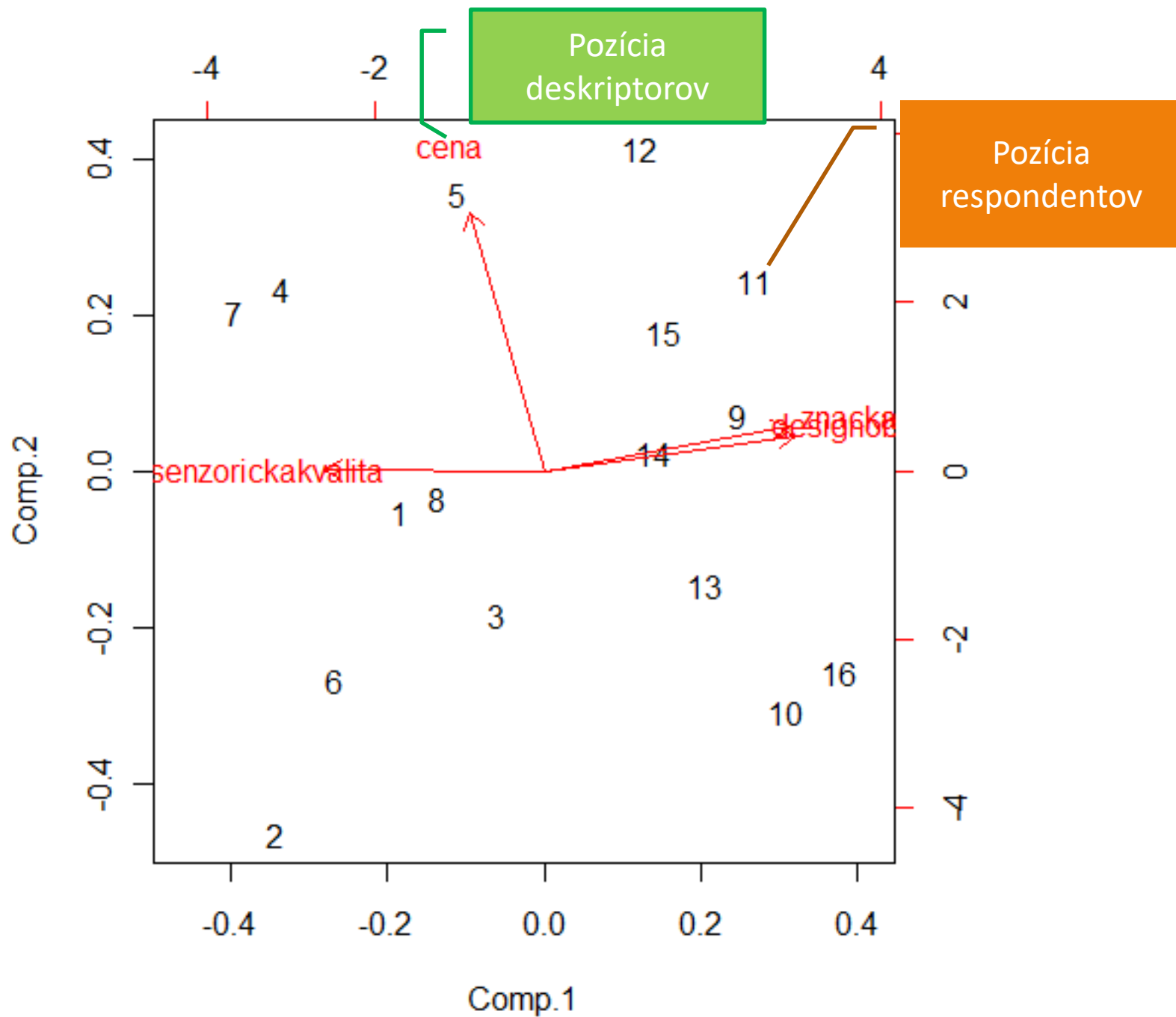
Pozícia produktov (score)

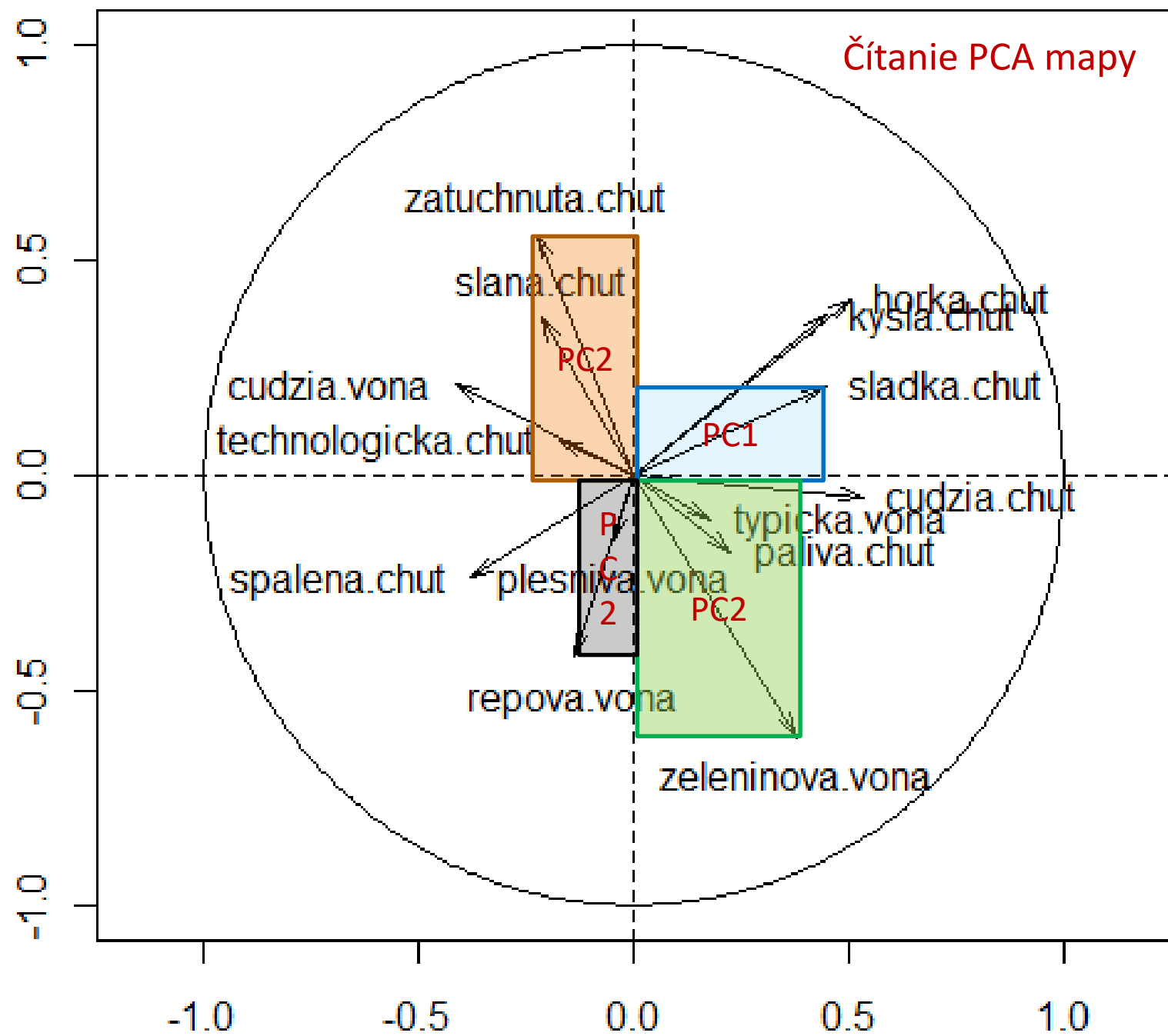


3D graf



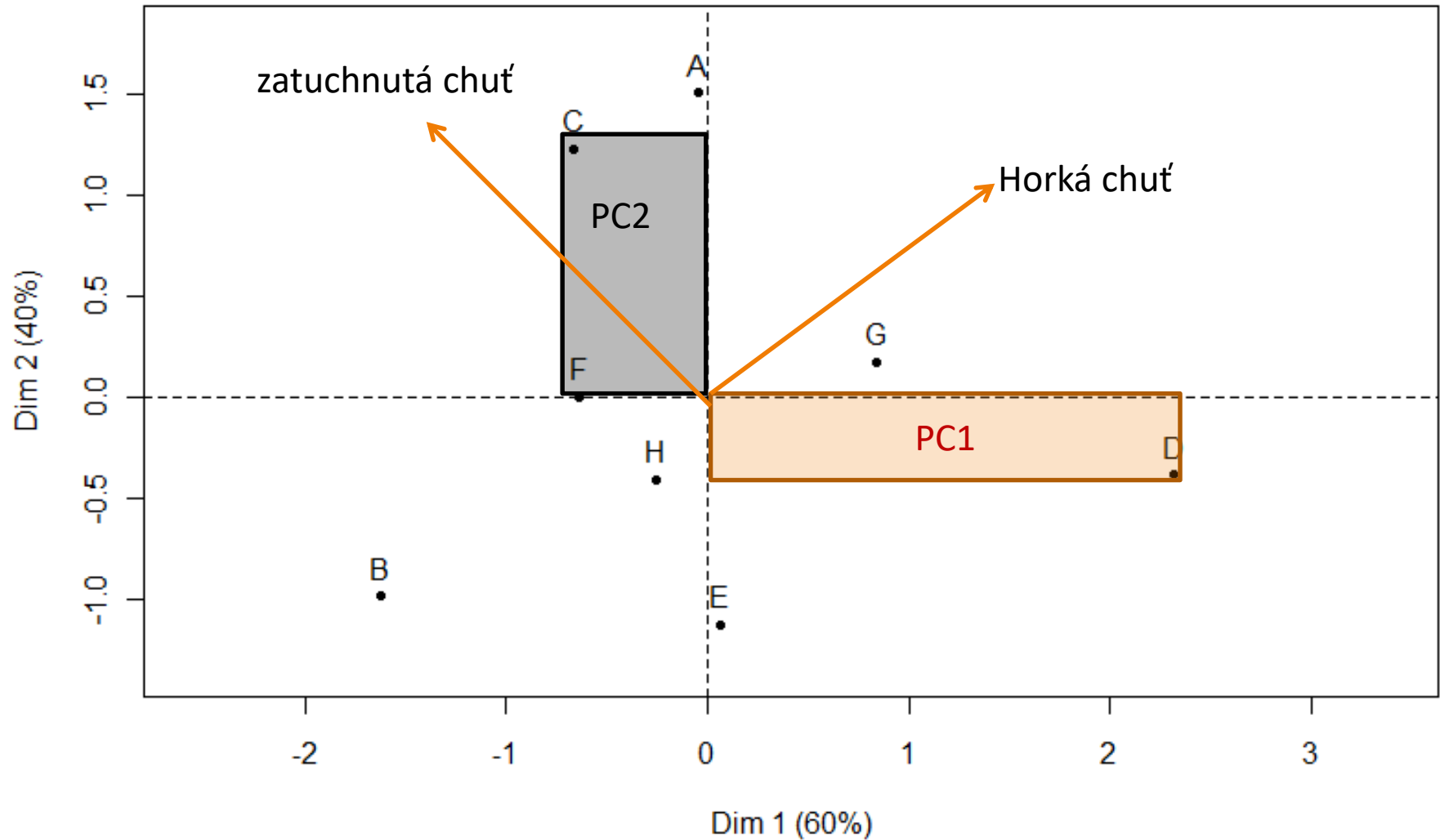
2D graf





Čítanie PCA mapy

Individuals factor map (PCA)



PROFILOGRAM



Faktorová analýza



- Hľadá skryté (latentné) premenné
- snaží sa vysvetliť ich správanie (pomenovať faktory)

Priebeh

- Redukcia dát
- Maticové úpravy (rotácia VARIMAX)
- Kaiserovo kritérium/sutinový graf → počet F
- Faktory $(-1;-0.6), (0.6;1)$

Multidimenzionálne Škálovanie (MDS)

- Obdoba FA analyzujúca **pomerové!!!**
- rozdiel je len vo vstupnej matici dát
- Výstupom je MDS mapa (identická PCA)
- v SA využívaná minimálne
- Marketing, Spotrebiteľské správanie



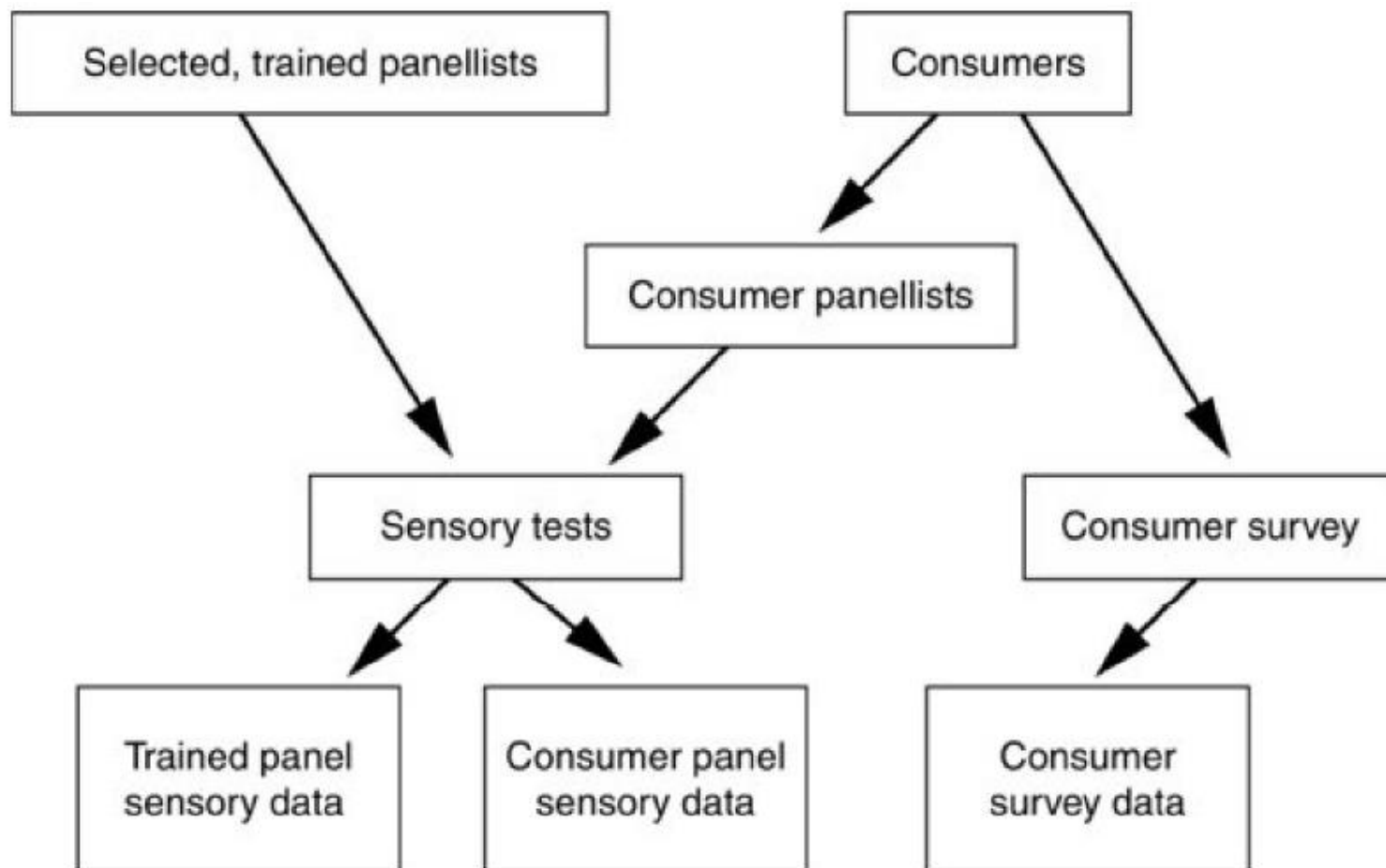
Korešpondečná analýza (CA)



Kontingenčná tabuľka sa používa k vizualizácii vzťahu dvoch štatistických premenných (opakovaní).

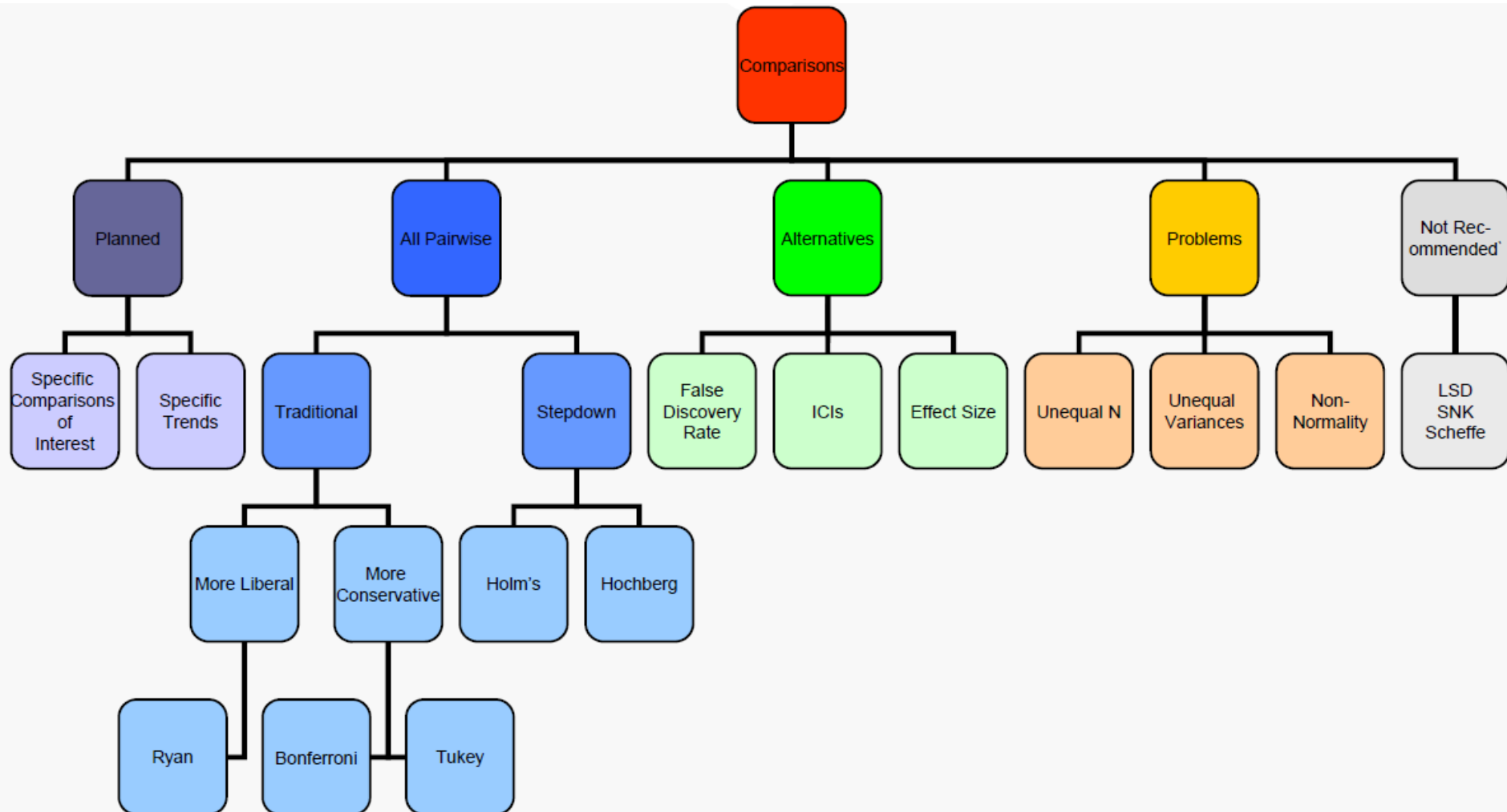
Obľuba	Svetlé pivo	Tmavé pivo	Rezané pivo
Muži	560	125	325
Ženy	55	356	59

- Idea je zobrazíť tabuľku v 2-rozmernom priestore (MAPE)
- Podobná PCA, ale s kvalitatívnymi dátami
- **Napadá Vás ako by sa dali zobrazíť kompletne výsledky ankety, nie jednotlivo po otázkach testovaných chí-kvadrátom? ☺**



Sources	Subjective nature	Error level	Analysis methods
Trained panel sensory	Low	Low	Parametric
Consumer sensory	High	High	Parametric, non-parametric
Consumer survey	High	High	Non-parametric, parametric

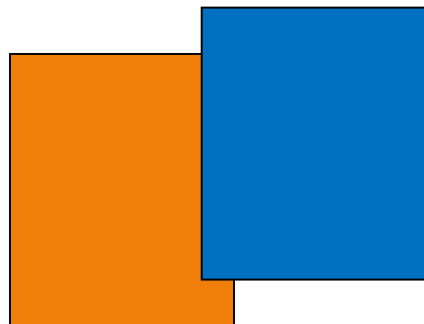
Viacnásobné porovnania



Porovnanie toho istého datasetu

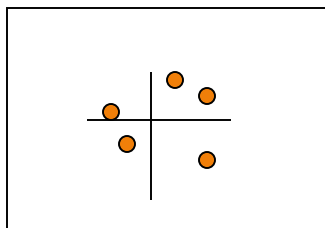
1	2	3	4	5	6	7	8
rawp	Bonf	Holm	Hochberg	SidakSS	SidakSD	LSD	BY
0.009	0.054	0.054	0.054	0.053	0.053	0.045	0.110
0.015	0.090	0.075	0.075	0.087	0.073	0.045	0.110
0.029	0.174	0.116	0.116	0.162	0.111	0.058	0.142
0.050	0.300	0.150	0.150	0.265	0.143	0.075	0.184
0.080	0.480	0.160	0.160	0.394	0.153	0.096	0.235
0.210	1.000	0.210	0.210	0.757	0.210	0.210	0.514

PREFERENČNÉ MAPOVANIE

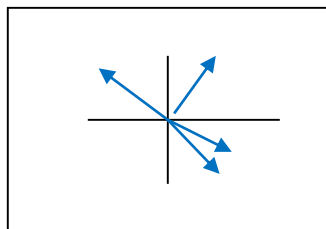


Spotrebiteľský prieskum

Laboratórne dáta

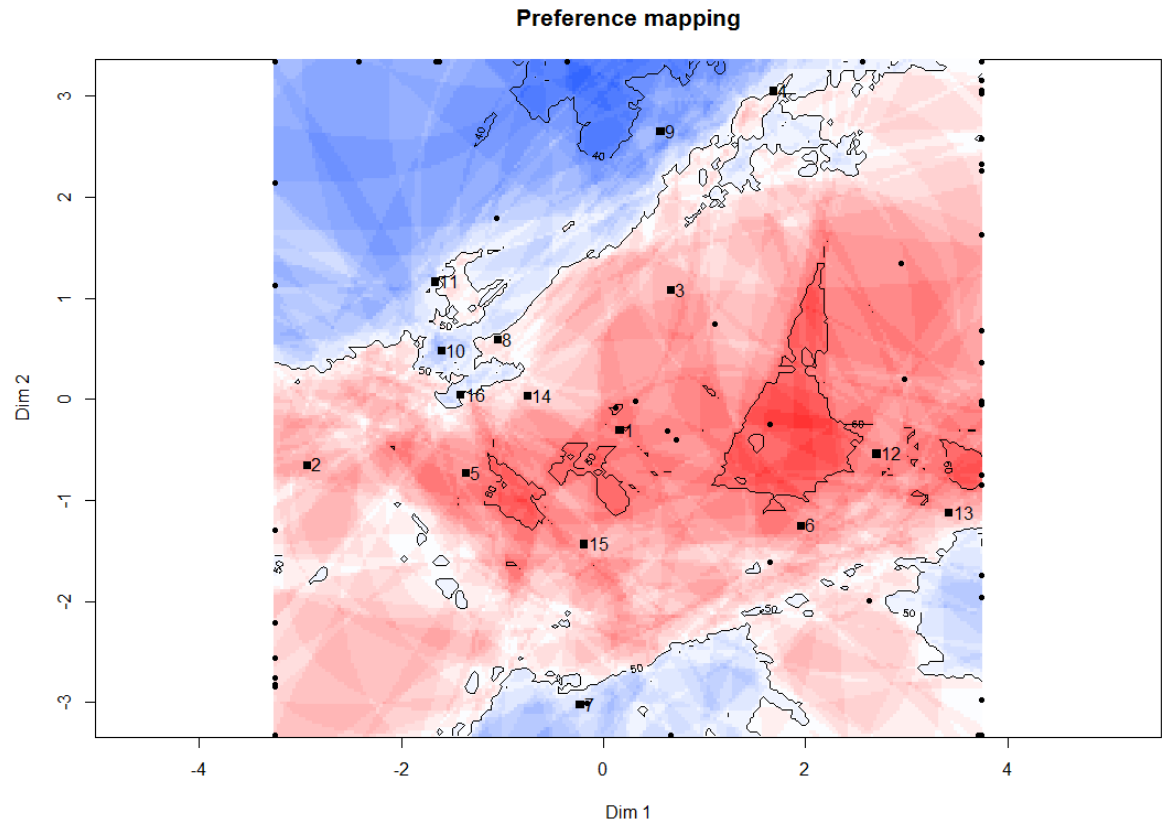
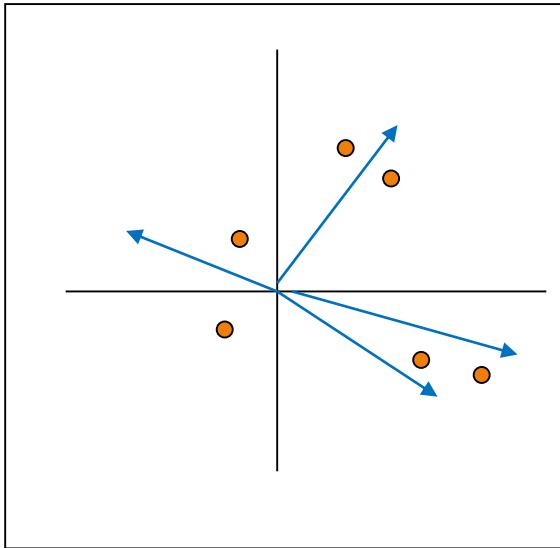


Pozícia produktov (score)
(Matica dát z profilovej analýzy)



Pozícia deskriptorov (loadings) sa vypočítajú od spotrebiteľov
(Matica dát z prieskumu)

MAPOVANIE PREFERENCIÍ (PREFERENCE MAPPING) II.



Pravidlo GIGO: „Garbage IN, Garbage OUT“.

Occamova britva: „ Ak existuje na nejaký jav viacero vysvetlení (riešení) je lepšie uprednostňovať to najmenej komplikované“.



ĎAKUJEM ZA POZORNOSŤ