KONTINGENČNÍ TABULKY

Komentované řešení pomocí programu *Statistica*

Vstupní data – transformace před vložením

	1 Město	2 Hodnocení
1	P	A
2	P	A
3	P	A
4	P	A
5	P	A
6	P	A
7	P	A
8	S	A
9	S	A
10	S	A
11	S	A
12	K	A
13	K	A
14	K	A
15	K	A
16	K	A
17	K	A
18	P	B
19	P	B
20	P	B

- Než data vložíme do tabulky ve *Statistice*, musíme si je předpřipravit.
- Označme si P – Prahu, S – Šumperk a K – Klatovy.
- Data musíme přetransformovat do původní dotazníkové podoby (obvykle v této podobě data získáváme), např. budeme mít sedm řádků dvojice P, A; čtyři S, A a šest K, A; atd.
- Anebo to můžeme udělat ve Statistice přímo…

Vstupní data – transformace po vložení I

	1 A	2 B	3 C
Praha	7	7	6
Šumperk	4	4	10
Klatovy	6	11	3

	1 Četnost hodnocení	2 Hodnocení
Praha	7	A
Praha	7	B
Praha	6	C
Šumperk	4	A
Šumperk	4	B
Šumperk	10	C
Klatovy	6	A
Klatovy	11	B
Klatovy	3	C

	1 Četnost hodnocení	2 Hodnocení	3 Město
Praha	7	A	
Praha	7	B	
Praha	6	C	
Šumperk	4	A	
Šumperk	4	B	
Šumperk	10	C	
Klatovy	6	A	
Klatovy	11	B	
Klatovy	3	C	

- Máme vloženou tabulku četností
- **Data → Přeskupování → Seskupování → Proměnné → Vybrat vše → OK → Jméno cílové proměnné:** Četnost hodnocení, **Jméno kódové proměnné:** Hodnocení → **OK**
- **Data → Proměnné → Přidat → Za:** 2, **Jméno:** Město → **OK** (anebo levým tlačítkem dvakrát poklepeme na sloupec a přidáme jednu proměnnou, a pak si ji pojmenujeme)

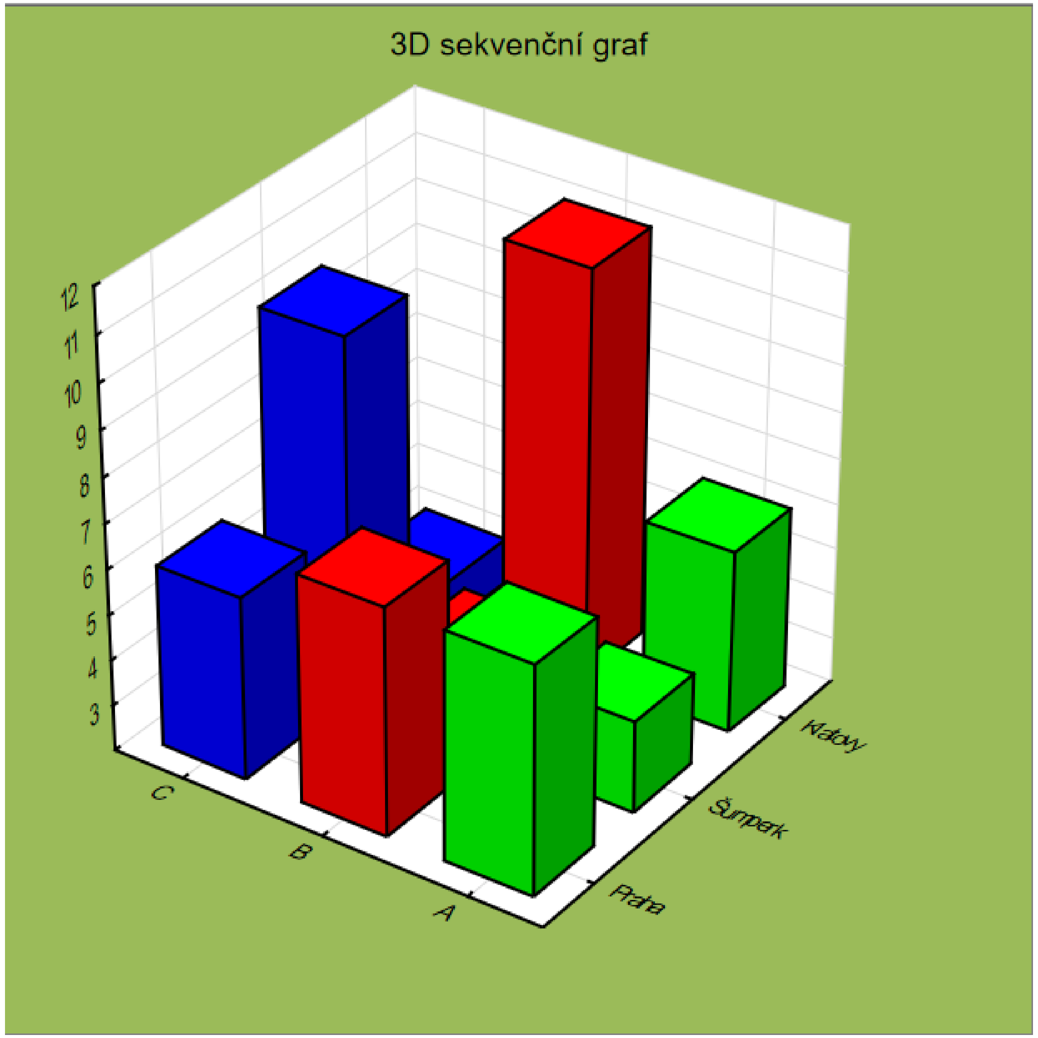
Vstupní data – transformace po vložení II

	1 Četnost hodnocení	2 Hodnocení	3 Město
Praha	7	A	Praha
Praha	7	B	Praha
Praha	6	C	Praha
Šumperk	4	A	Šumperk
Šumperk	4	B	Šumperk
Šumperk	10	C	Šumperk
Klatovy	6	A	Klatovy
Klatovy	11	B	Klatovy
Klatovy	3	C	Klatovy

- **Data** → **Jména** → v **Přenést jména případů** zvolíme **Do, Proměnné:** Město → **OK**
- **Nástroje** → **Váha** → **Proměnná vah:** Četnost hodnocení, ve **Stavu** zvolme **Zap.** → **OK**

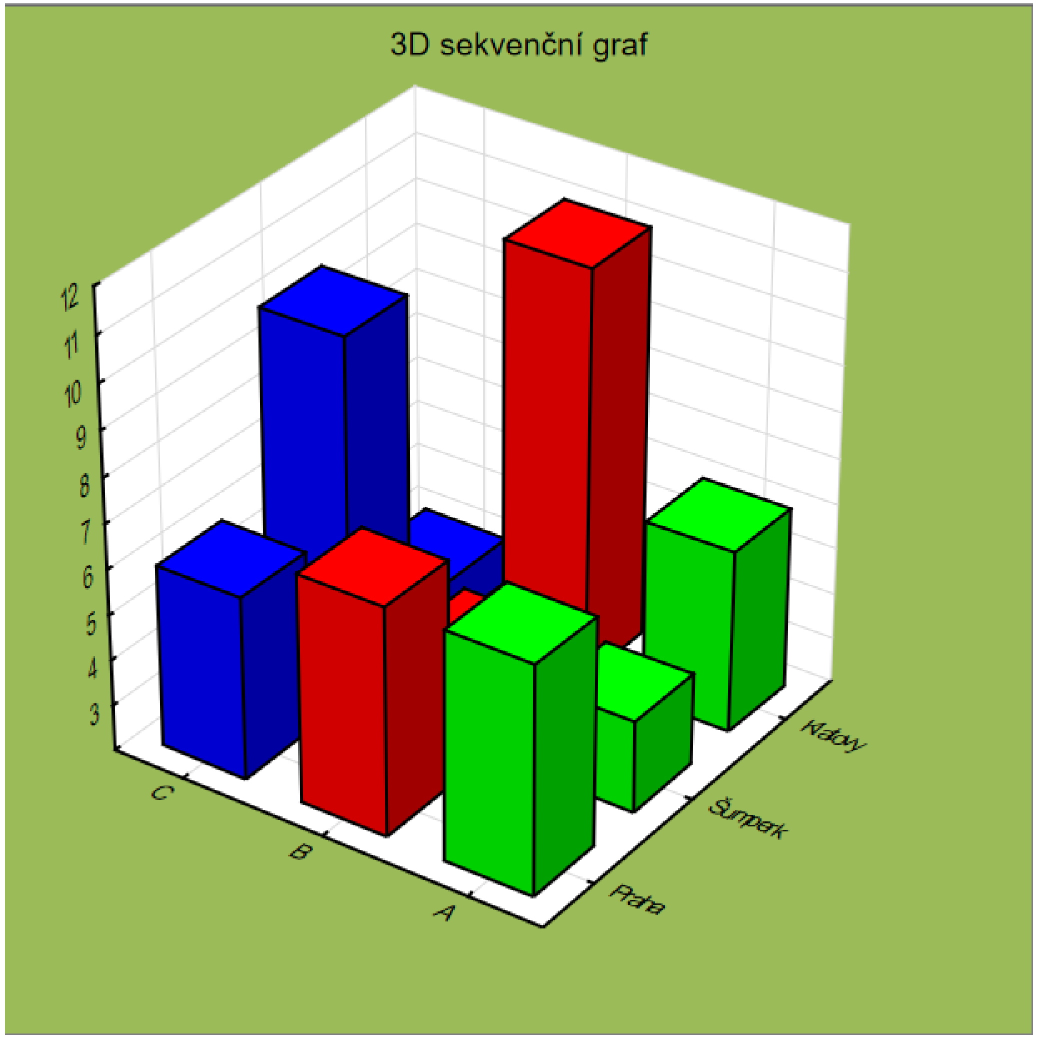
- Nyní již můžeme použít k analýze připravený balíček kontingenčních tabulek, ale nejdříve se podíváme na graf četností.

Graf naměřených četností I



- Vykreslíme si graf původní datové tabulky
- **Grafy → 3D Sekv. → Grafy zdrojových dat → Proměnné – Vybrat vše → OK → OK**
- Z grafu je patrné, že
  - salámy z Prahy mají vyrovnané hodnocení (přibližně stejný počet respondentů u všech kategorií hodnocení)
  - u salámů z Šumperka převažuje hodnocení „C“
  - salámy z Klatov byly nejčastěji hodnoceny „B“

Graf naměřených četností II



- Můžeme tedy usuzovat, že na základě senzorického hodnocení vychází nejlépe salám z Prahy, následuje salám z Klatov a nejhůře dopadl salám ze Šumperka.
- V následujícím se pokusíme zjistit, zda jsou rozdíly mezi salámy statisticky významné, nebo zda může jít pouze o náhodné odchylky. To budeme testovat pomocí χ^2 – testu nezávislosti v kontingenční tabulce.

χ^2 – test nezávislosti

- Použijeme tabulku, kterou jsme si ve *Statistice* vytvořili
- **Statistiky** → **Základní statistiky** → **Kontingenční tabulky** → **OK** → **Specif. tabulky (vyberte proměn.)** – **List1** Hodnocení, **List2** Město → **OK** → **Možnosti** → odškrtneme **Zvýraznit četn.**, naopak zaškrtneme **Očekávané četnosti**, **Reziduální četnosti** a **Pearsonův & M-T Chí-kvadrát** → **Výpočet**
- Jako výstup získáme tři tabulky
  - První je původní kontingenční tabulka
  - Druhá je tabulka očekávaných četností
  - Třetí je tabulka pozorovaných minus očekávaných četností

Kontingenční tabulka

Hodnocení	Město Praha	Město Šumperk	Město Klatovy	Řádk. součty
A	7	4	6	17
B	7	4	11	22
C	6	10	3	19
Vš.skup.	20	18	20	58

Ověření předpokladů testu

- Protože jde o test asymptotický, je třeba mít „dostatečně“ početné třídy, přesněji:
  - všechny teoretické četnosti jsou alespoň jedna,
  - nejméně 80 % tříd má četnost nejméně pět.
- Tyto předpoklady ověříme z tabulky očekávaných četností
  - Vidíme, že všechny očekávané četnosti větší, než pět, předpoklady testu jsou tedy splněny.

Souhrnná tab.: Očekávané četnosti
Pearsonův chí-kv. : 8,11269, sv=4, p=,087536

Hodnocení	Město Praha	Město Šumperk	Město Klatovy	Řádk. součty
A	5,86207	5,27586	5,86207	17,00000
B	7,58621	6,82759	7,58621	22,00000
C	6,55172	5,89655	6,55172	19,00000
Vš.skup.	20,00000	18,00000	20,00000	58,00000

χ^2 – test nezávislosti závěry

- Např. v tabulce očekávaných četností vidíme p – hodnotu 0,0875 χ^2 – testu nezávislosti, proto na požadované hladině významnosti α = 10 % zamítáme nulovou hypotézu o nezávislosti ve prospěch alternativy, tj. zjištěné rozdíly mezi salámy jsou na této hladině statisticky významné.
- Ze zjištěné p – hodnoty ovšem plyne, že statistická významnost rozdílů není úplně přesvědčivá. Kdybychom testovali na hladině významnosti α = 5 %, nulovou hypotézu už bychom nezamítali a výsledky bychom považovali za neprůkazné.

Pozorované minus očekávané četnosti

Souhrnná tab.: Pozorované minus očekávané četnosti
Pearsonův chí-kv. : 8,11269, sv=4, p=,087536

Hodnocení	**Město Praha**	Město Šumperk	Město Klatovy	Řádk. součty
A	1,137931	-1,27586	0,13793	-0,000000
B	-0,586207	-2,82759	3,41379	0,000000
C	-0,551724	4,10345	-3,55172	0,000000
Vš.skup.	0,000000	-0,00000	0,00000	-0,000000

- Na základě rozdílu pozorovaných a očekávaných četností (třetí tabulka) lze získat další představu o datech:
  - pro lepší přehlednost jsou případy, ve kterých jsou pozorované četnosti větší než očekávané, obarveny červeně, zbylé pak zeleně.
- Vidíme, že častější hodnocení než očekávané je „A" u salámu z Prahy, „C" u salámu ze Šumperka a „B" u salámu z Klatov. Potvrzuje se tím naše předchozí zjištění na základě grafu naměřených četností.

Úplným závěrem…

- Na základě senzorického hodnocení dopadl nejlépe salám z Prahy, za ním salám z Klatov a nejhůře salám ze Šumperka. Vzhledem k spočítané p – hodnotě (která je navíc pouze přibližná, neb se jedná o asymptotický test) jsou zjištěné rozdíly na hranici statistické průkaznosti.