1. cvičení

30.9.2014

Obsah

- Typy dat, SW statistika
- Popisná statistika
- Normální rozložení a důsledky nenormality pro parametrické statistické metody + simulovaný příkladový soubor s různými daty
  - T-test, ANOVA
- Korelace (s důrazem na vizualizaci klidně více proměnných pomocí maticových grafů a vliv přítomnosti odlehlé hodnoty) + simulovaný příkladový soubor s různými typy závislosti (nebo klasické kosatce)
- Kontingenční tabulky – princip vypočtu, pozorované vs. Očekávané
- Práce s binárními a kategoriálními daty – dummies, srovnání s bazální kategorií
- Grafy – základní principy grafů ve statistice

IBA MU

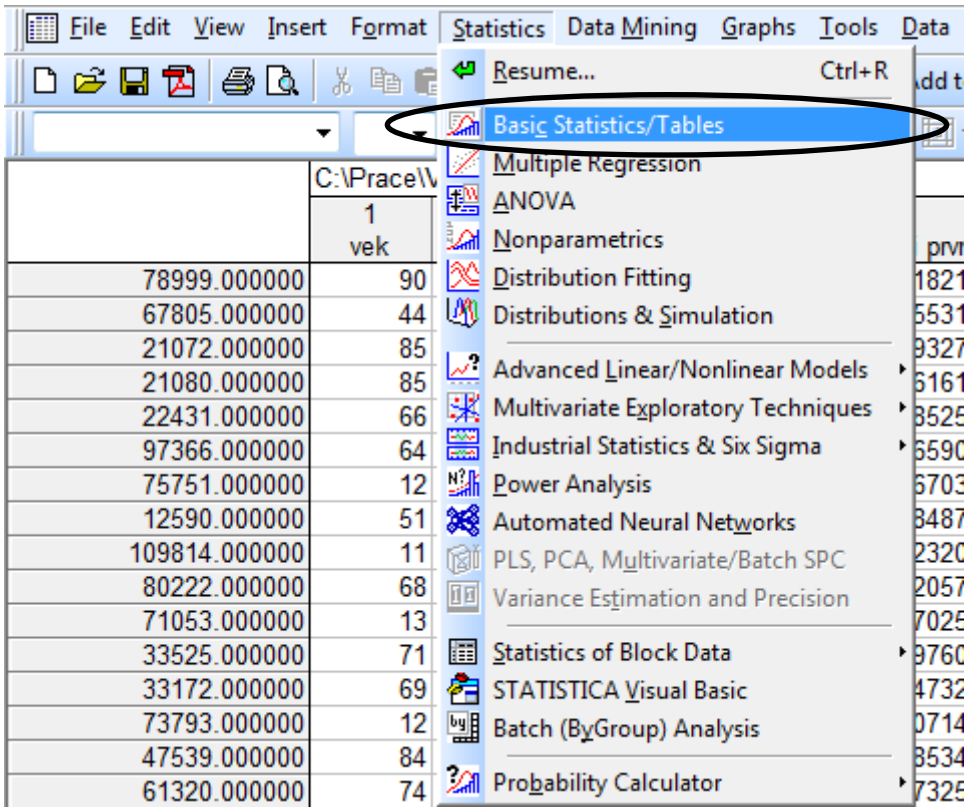
Data

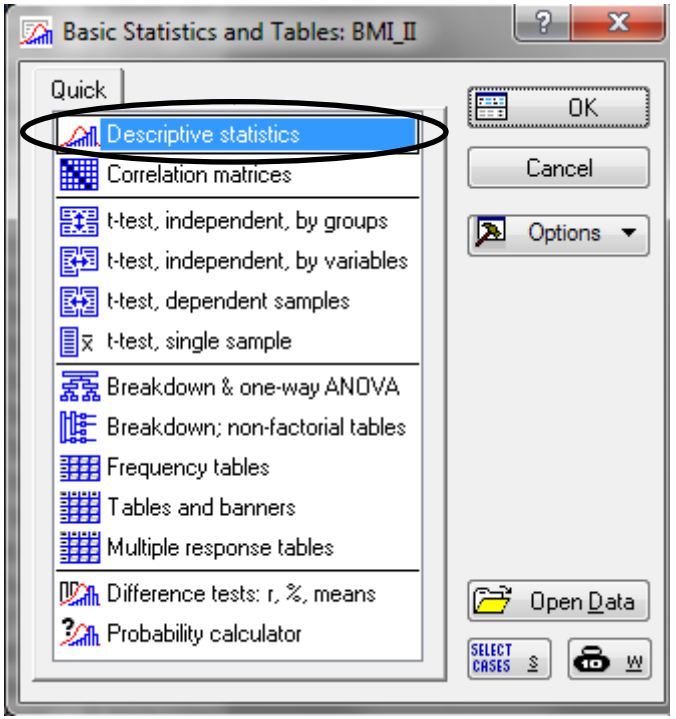
- Diskrétní
  - Kategoriální (několik kategorií: anamnéza, půdní typy,…)
  - Binární
    - speciální případ dat kategoriálních, kde máme jen dvě kategorie: muž vs. žena, výskyt vs. nevýskyt,…
- Spojitá
  - Data mohou nabývat jakýchkoliv hodnot: váha, výška, krevní tlak, nadmořská výška…
- Ordinální
  - Speciální typ dat, která jsou mezi daty spojitými a kategoriálními
  - Jedná se o kategorie, které je možno seřadit: kategorie BMI, věku, nadmořské výšky,…

Popisná statistika (1)

- Spojitá data – průměr, rozptyl, směrodatná odchylka, kvantily
- Ordinální data – kvantily, modus

Statistics → Basic Statistics/Tables → Descriptive statistics

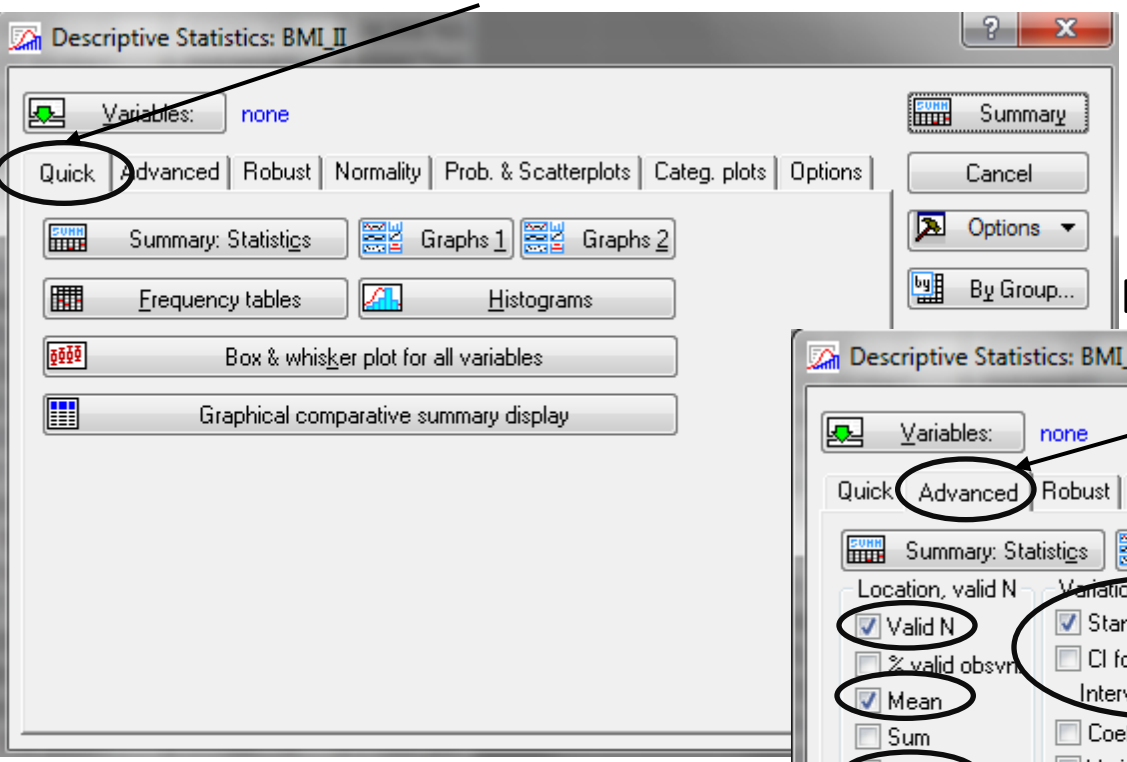




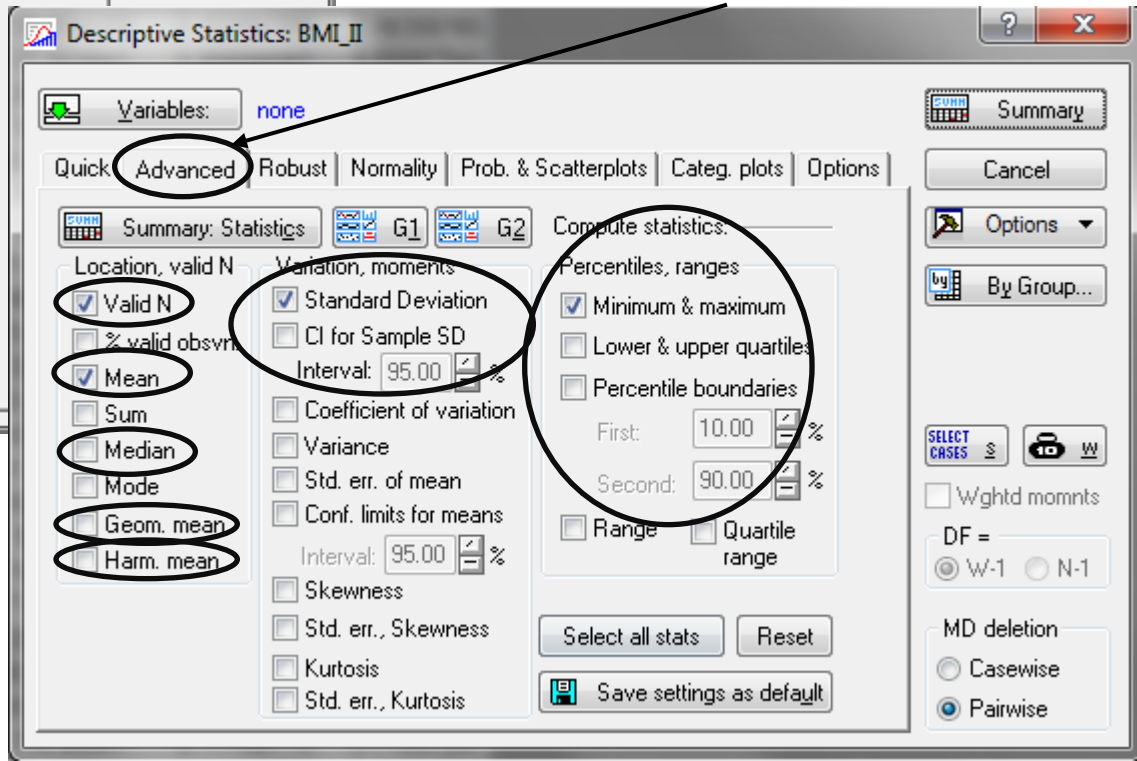
Popisná statistika (2)

- Záložka Quick a Advanced

Základní varianta – pro rychlý přehled



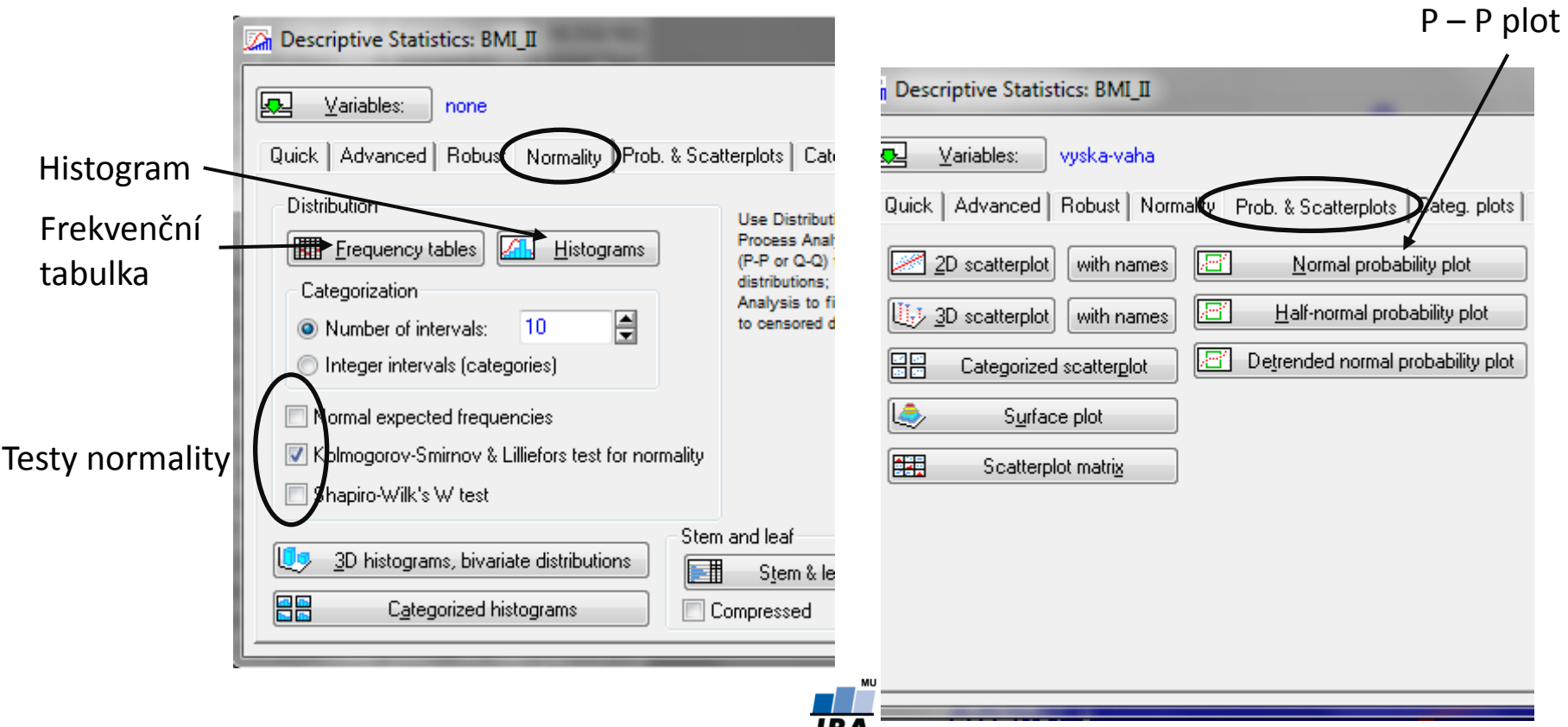
Detailnější varianta – pro pokročilé



Ověření předpokladu normality dat

- Většina statistických metod předpokládá normalitu dat → parametrické metody: t-test, ANOVA

- Záložka Normality → Testy normality, histogram, frekvenční tabulka
- Záložka Prob. & Scatterplots → P-P plot

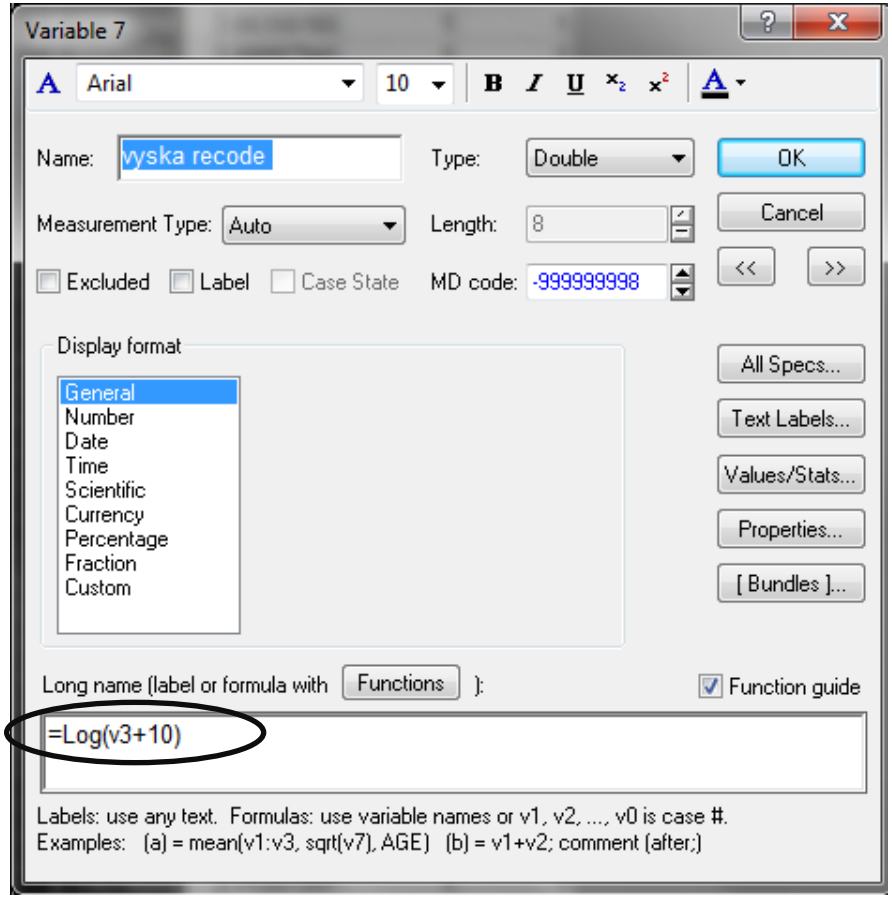


Transformace dat

- Nelineární, odmocniová, Boxova-Coxova transformace

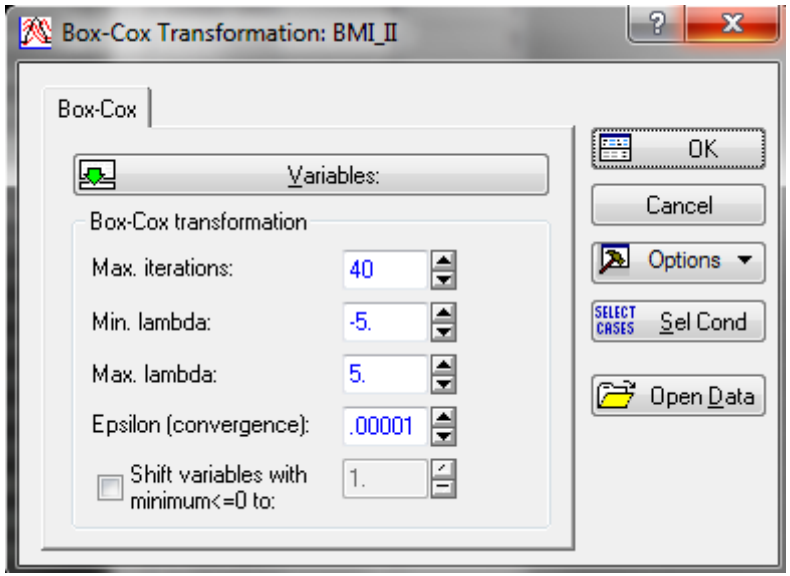
Nelineární, odmocninová transformace

Postup: Data → Variable specs…



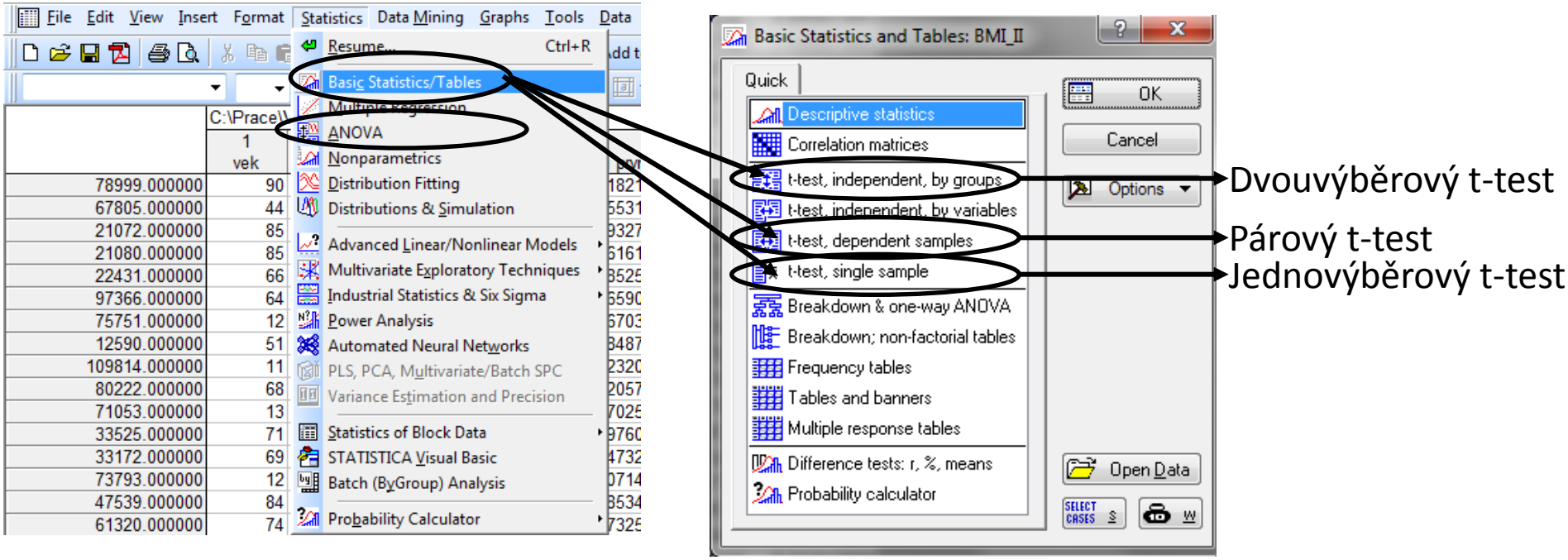
Boxova-Coxova Transformace

Postup: Data → Box-Cox transformation

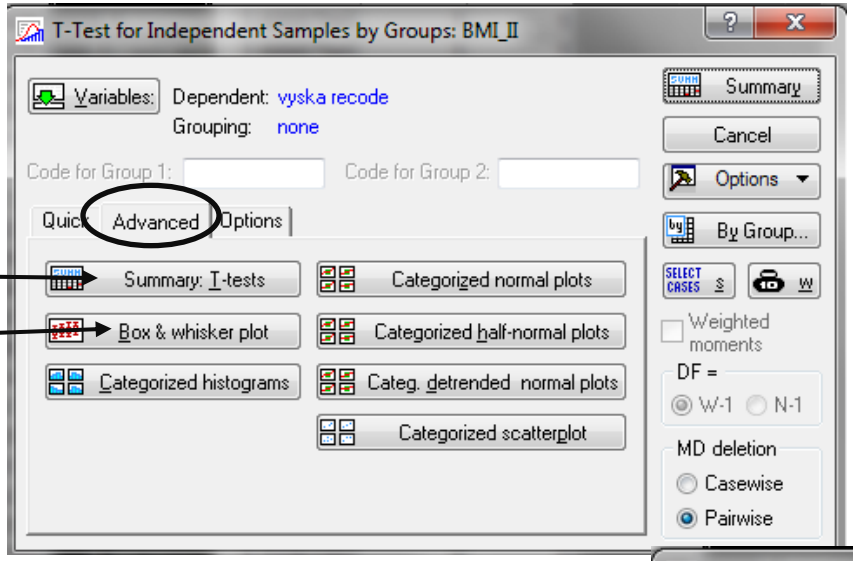


T-test, ANOVA

- Parametrické metody: obě předpoklad normality, ANOVA předpoklad homogenity rozptylu ve skupinách a nezávislost skupin



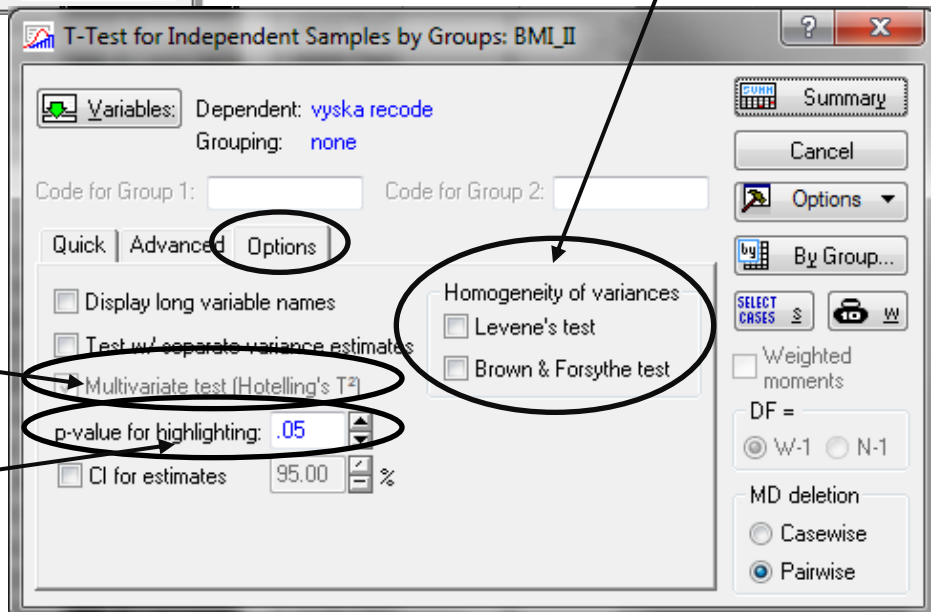
Dvouvýběrový T-test



Výsledek

Boxplot

Testy homogenity

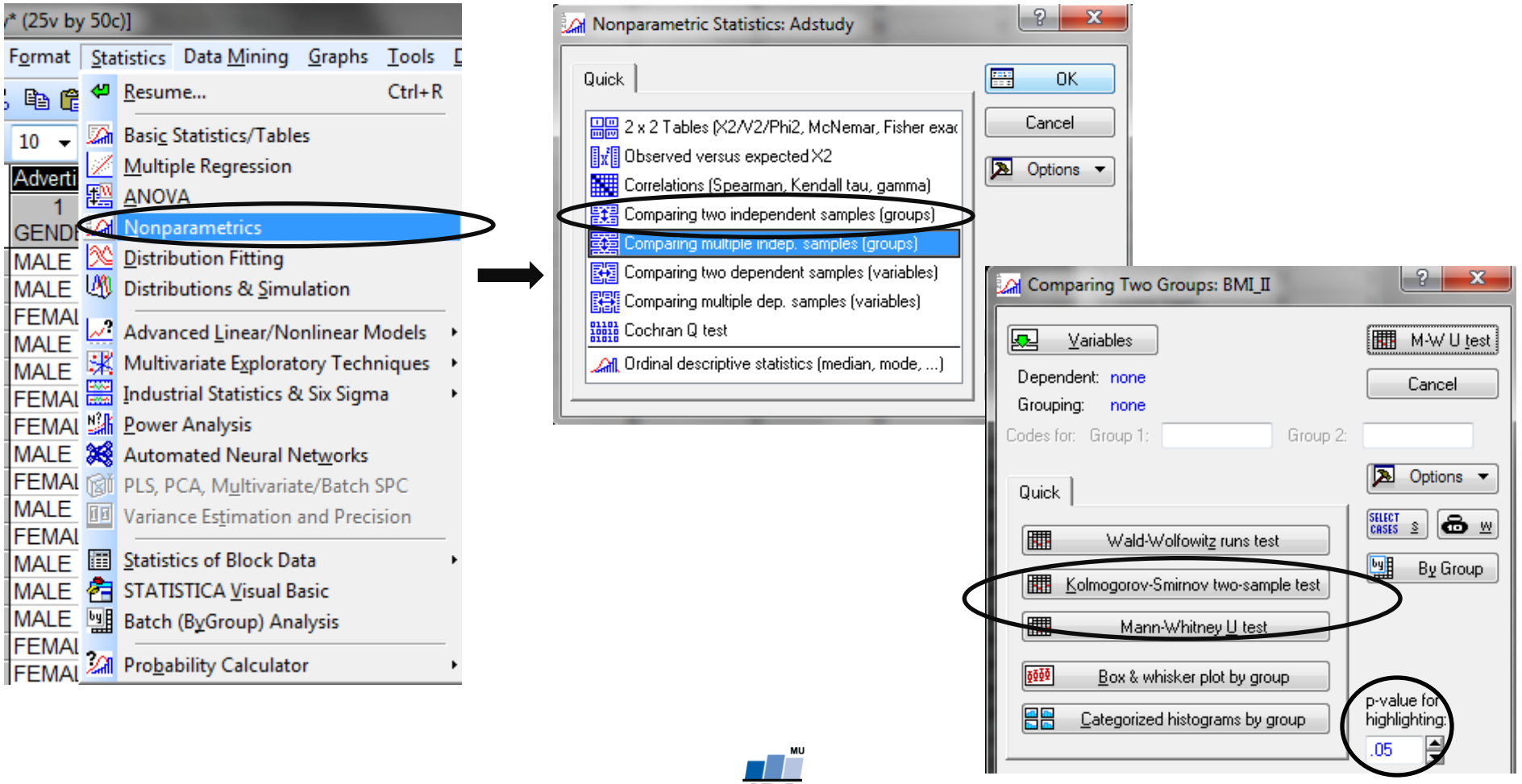


T-test pro více rozměrů

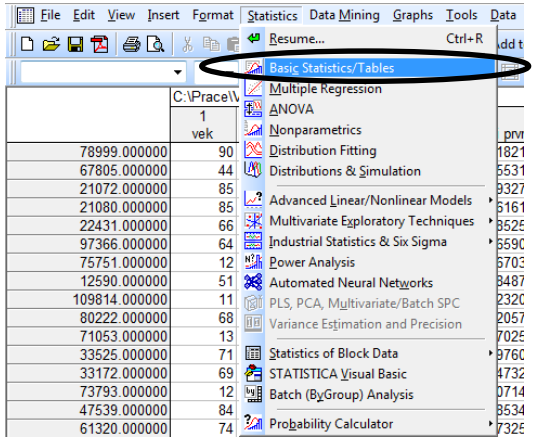
Nastavení hladiny statistické významnosti

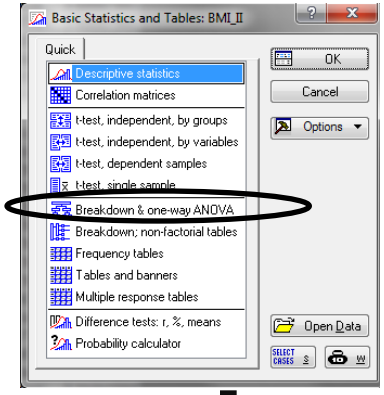
Nesplnění předpokladů pro T-test

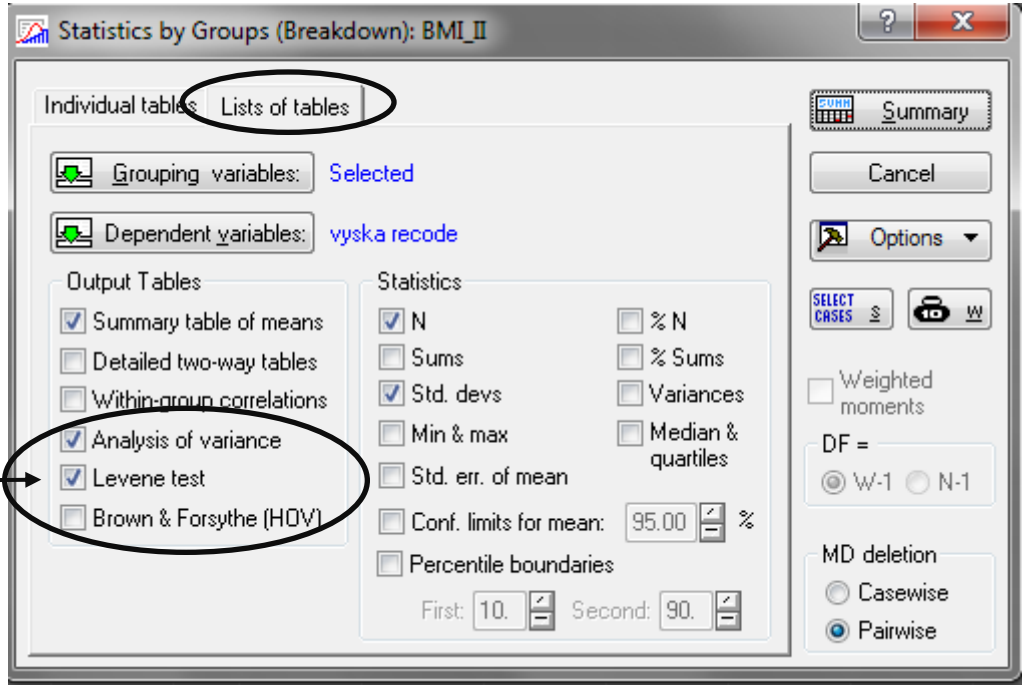
- Neparametrické testy
  - Kolmogorův-Smirnovův test
  - Mannův-Whitneyův test



Ověření předpokladu homogenity rozptylu pro ANOVu



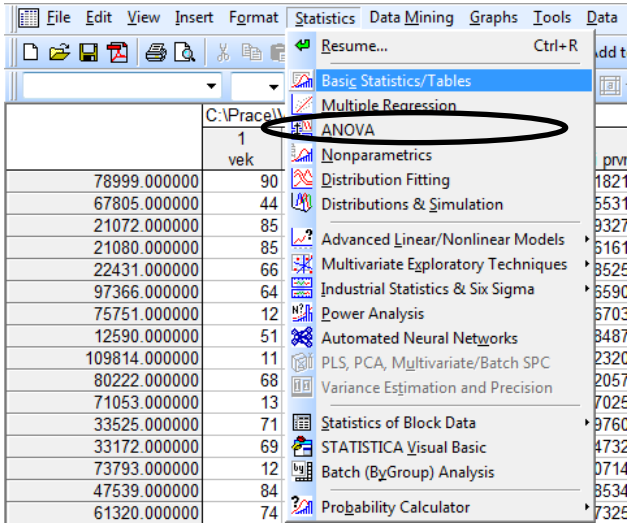


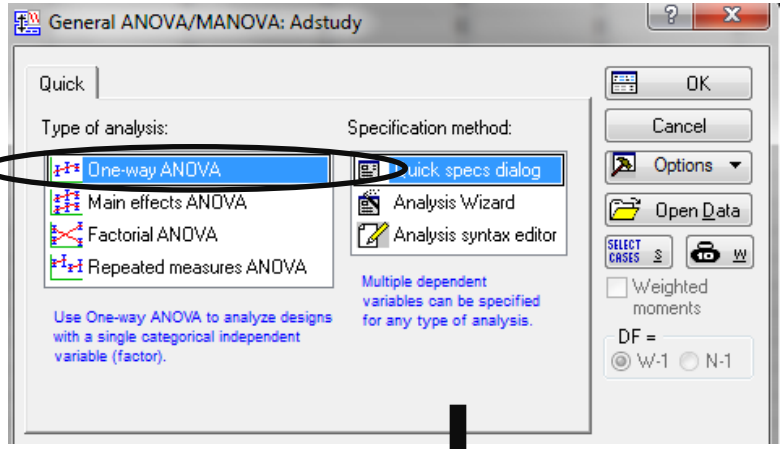


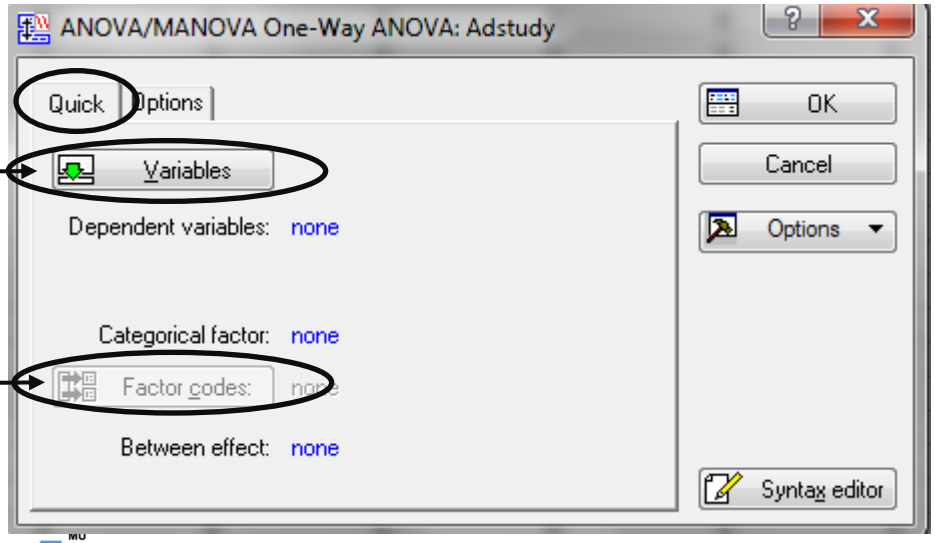
Testy homogenity rozptylu

ANOVA – analýza rozptylu

- Spojité proměnné dle kategorií
- Zda na hodnotu spojité proměnné má vliv hodnota kategoriální proměnné







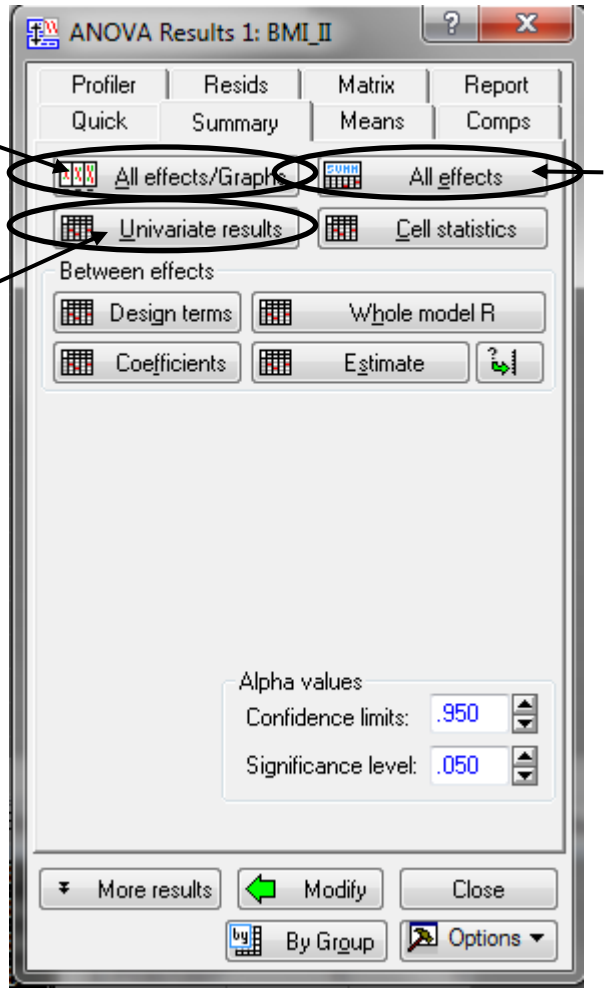
Spojitá proměnna

Kategoriální proměnná

ANOVA – výsledky 1

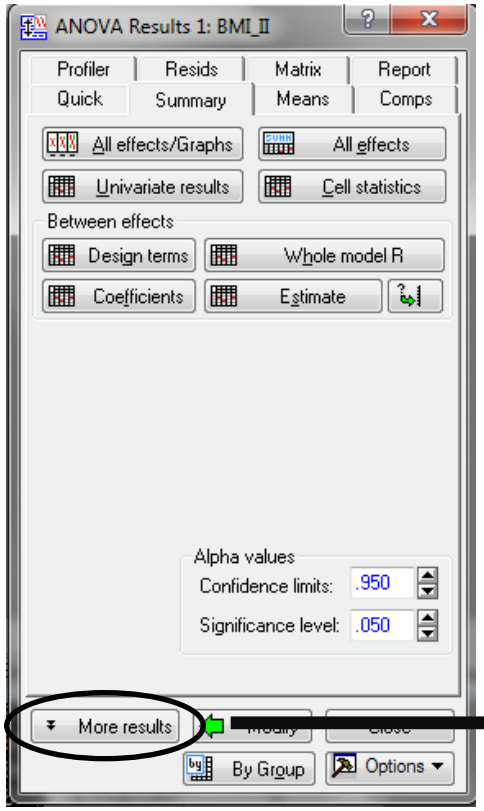
Boxploty

Výsledný vliv kategoriální proměné



Celkový výsledek

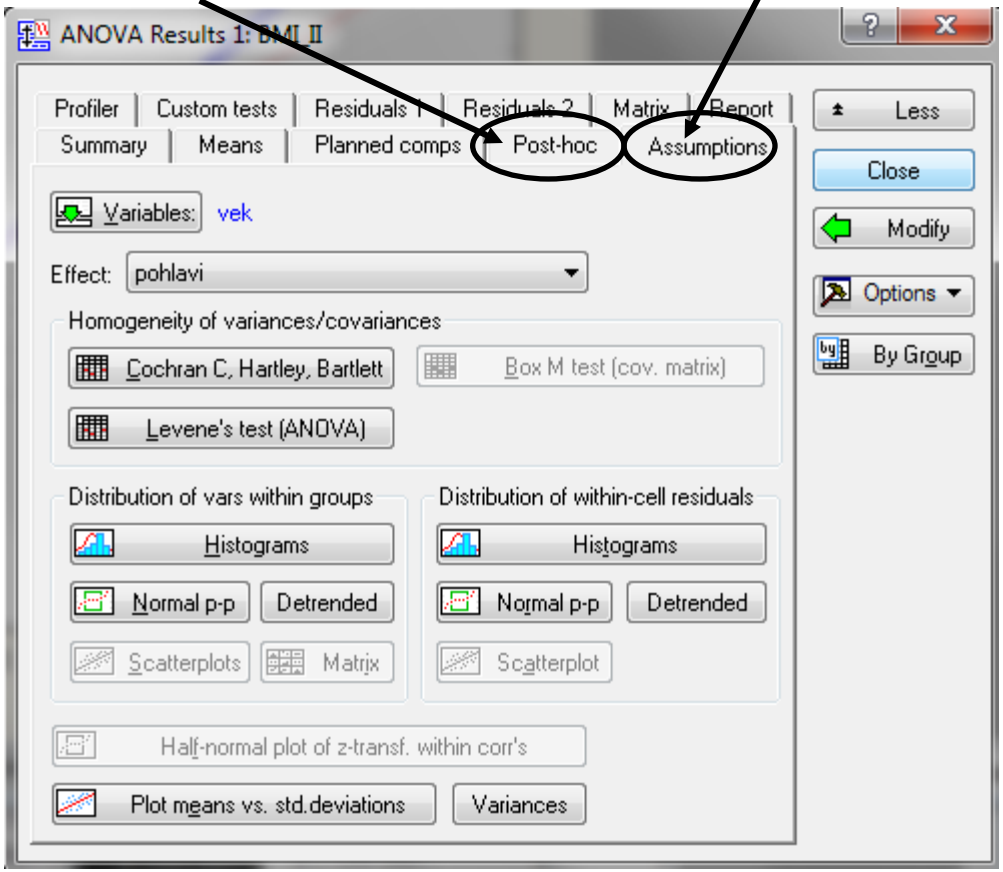
ANOVA – výsledky 2



Post-Hoc testy

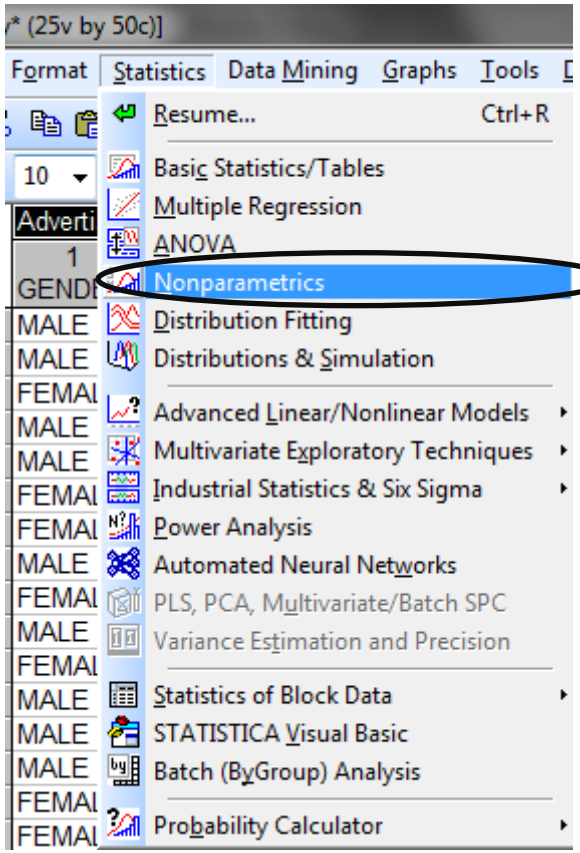
→stanovení homogenních skupin

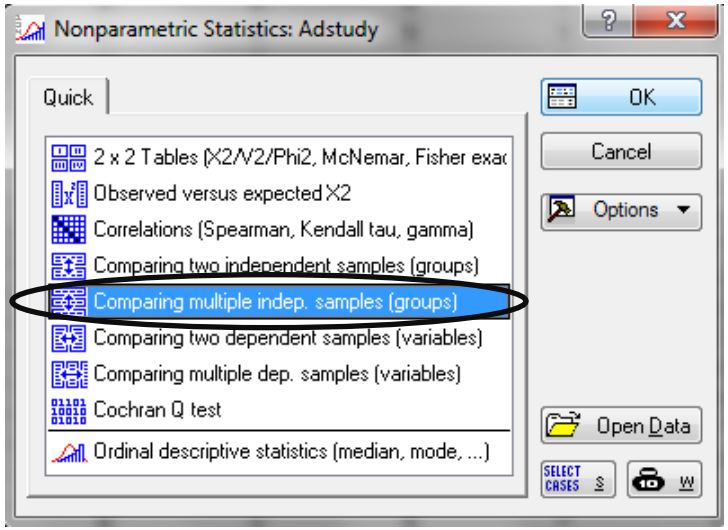
Ověření předpokladů

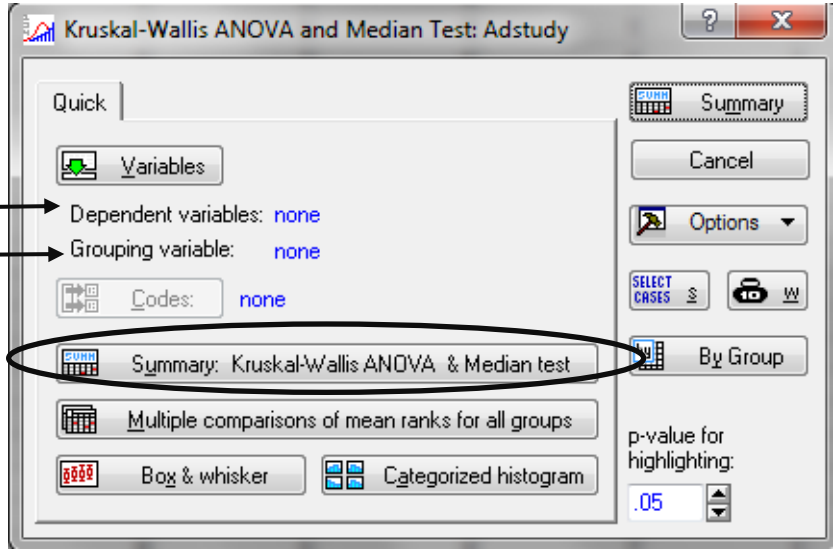


ANOVA – porušení homogenity rozptylu

- Neparametrická metoda Kruskalova – Wllisova analýza rozptylu





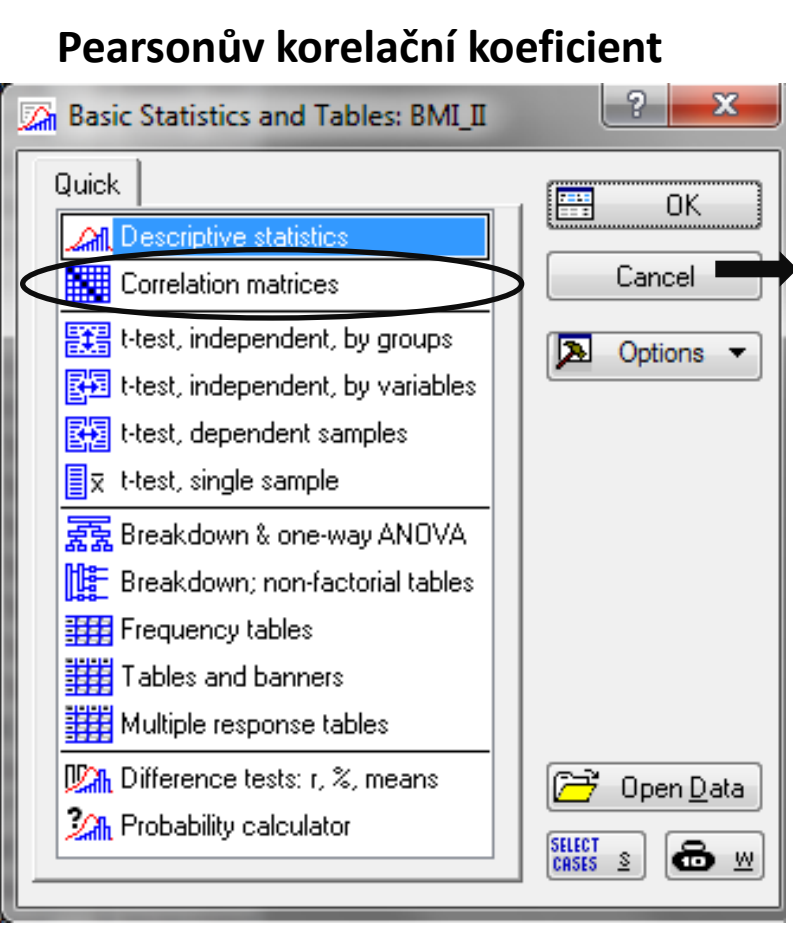


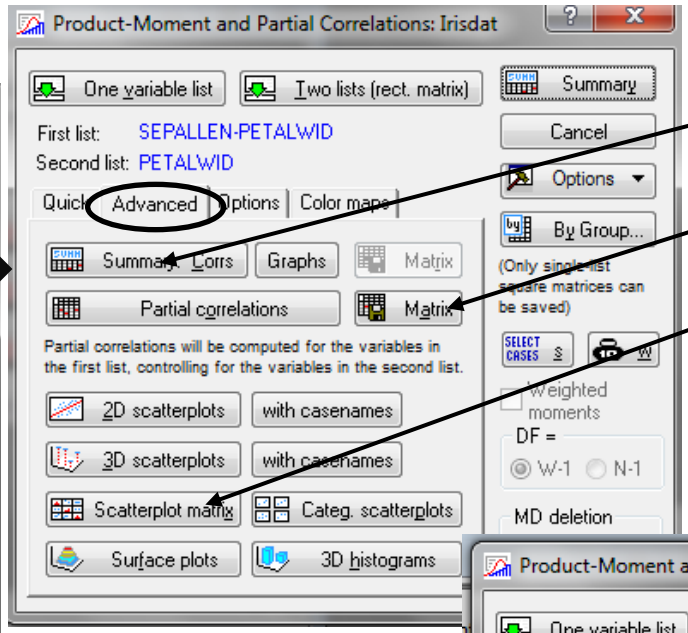
Spojitá proměnna
Kategoriální proměnná

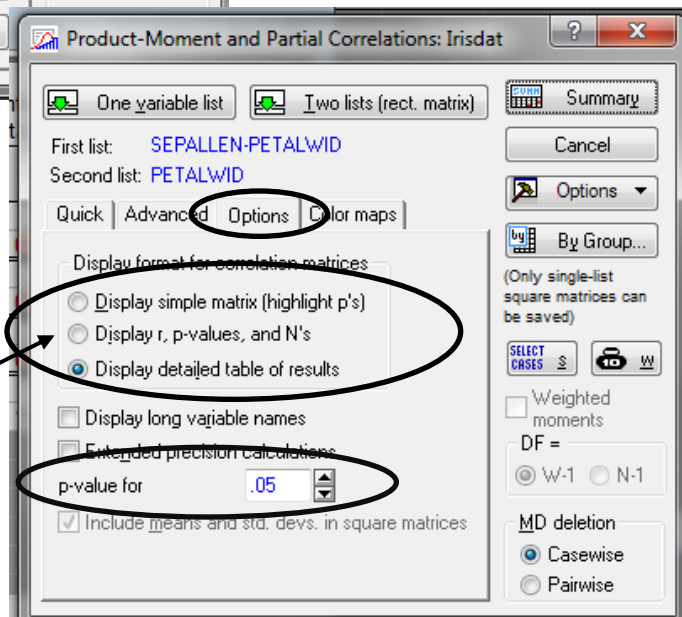
Korelace 1

- Spojité a normálně rozložené proměnné → Pearsonův korelační koeficient
- Pro ostatní proměnné → Spearmanův (pořadový) korelační koeficient

Pearsonův korelační koeficient







Celkový výsledek

Možnost uložení korelační matice

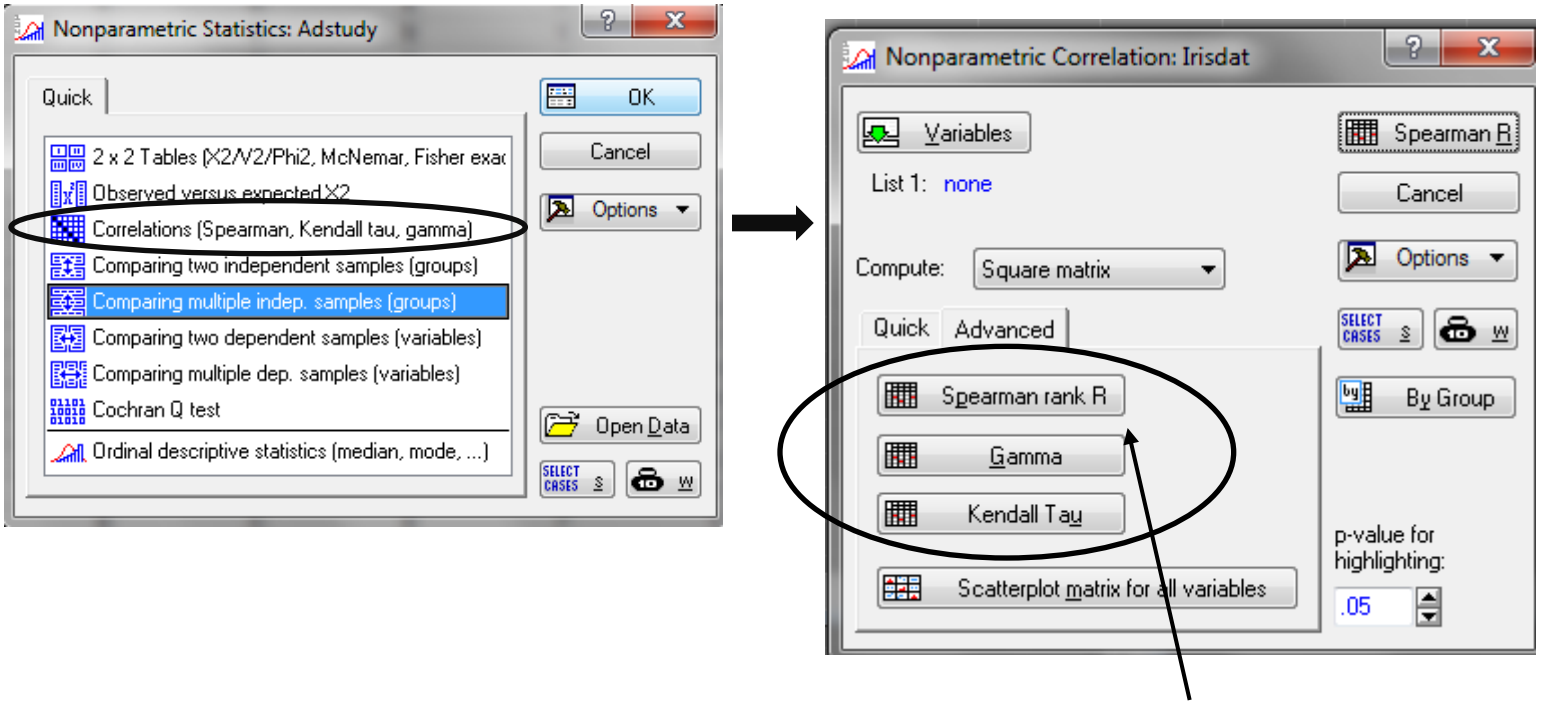
Maticový graf

Nastavení výsledné matice

Korelace 2

- Spojité a normálně rozložené proměnné → Pearsonův korelační koeficient
- Pro ostatní proměnné → Spearmanův (pořadový) korelační koeficient

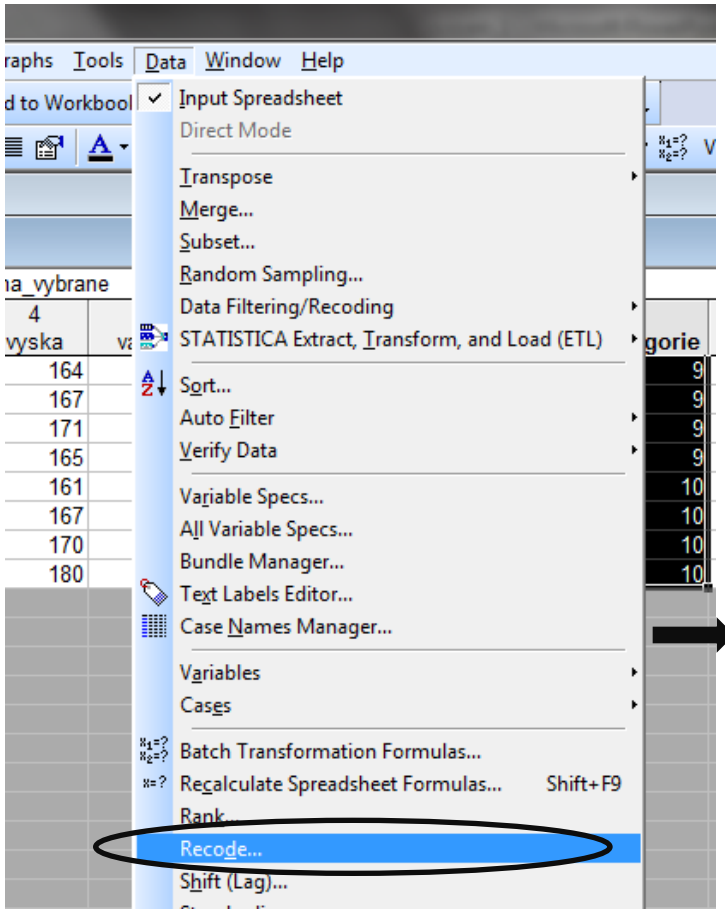
Spearmanův korelační koeficient

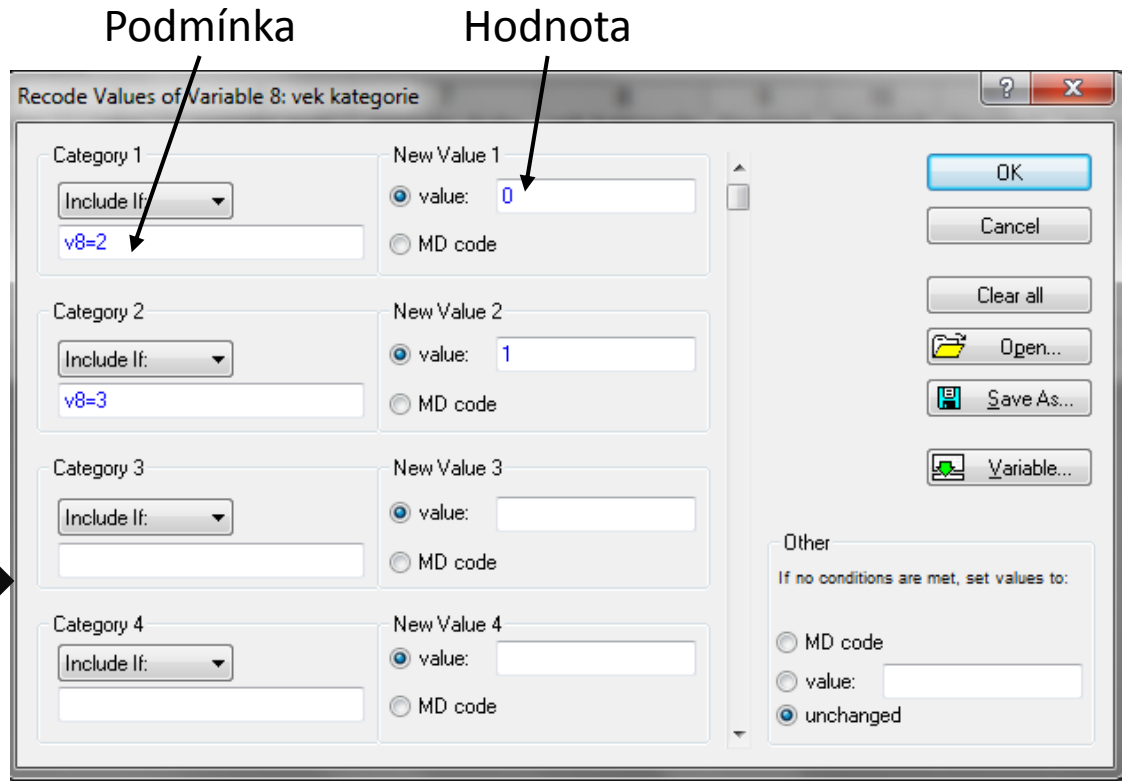


Spearmanův korelační koeficient
a jiné neparametrické korelační koeficienty

Binární a kategoriální data

- Binární → výskyt vs. nevýskyt, muž vs. žena, ano vs. ne, …
- Kategoriální → kategorie, možno převést na binární
- Srovnání s bazální kategorií → použití při logistické regresi, jedna kategorie vybrána jako základní a k ní jsou vztahovány ostatní

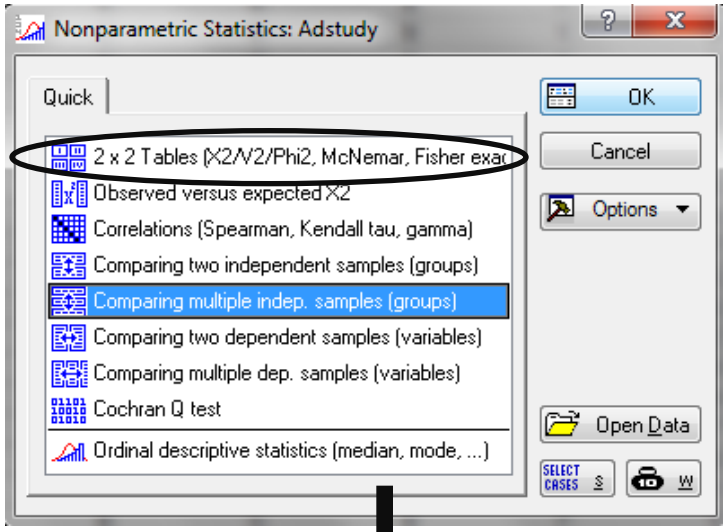


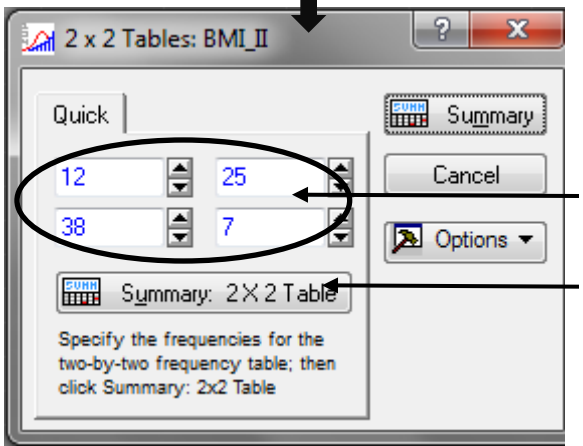


Kontingenční tabulky

- K přehledné vizualizaci vztahu dvou kategoriálních proměnných
- Excel nebo Statistika
- Chi-square test, Fisherův exaktní test,….

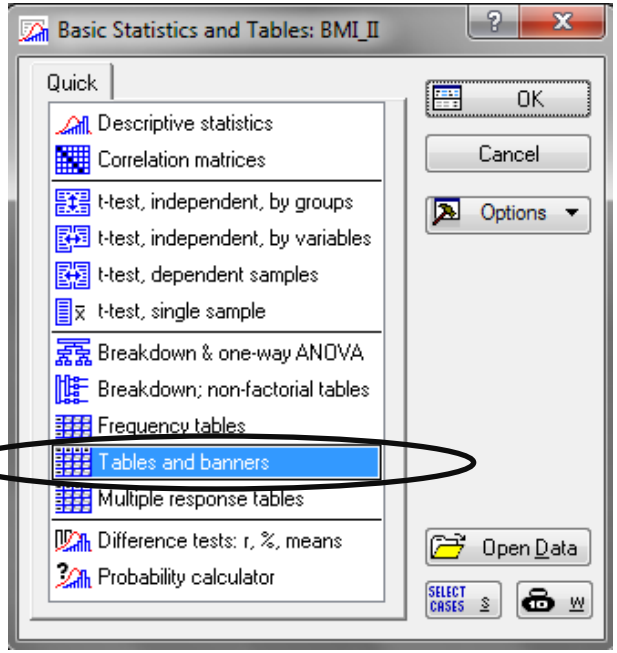
Varianta 2 x 2





vs.

Varianta n x m

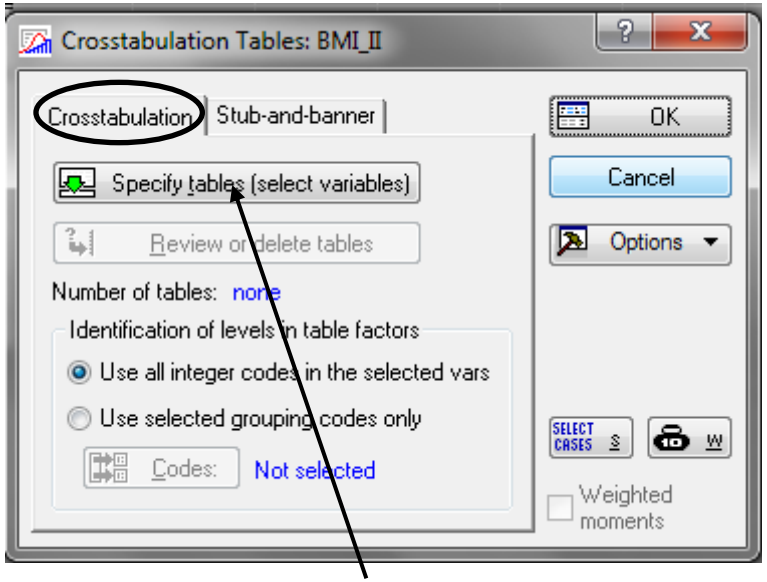


Hodnoty v kontingenční tabulce

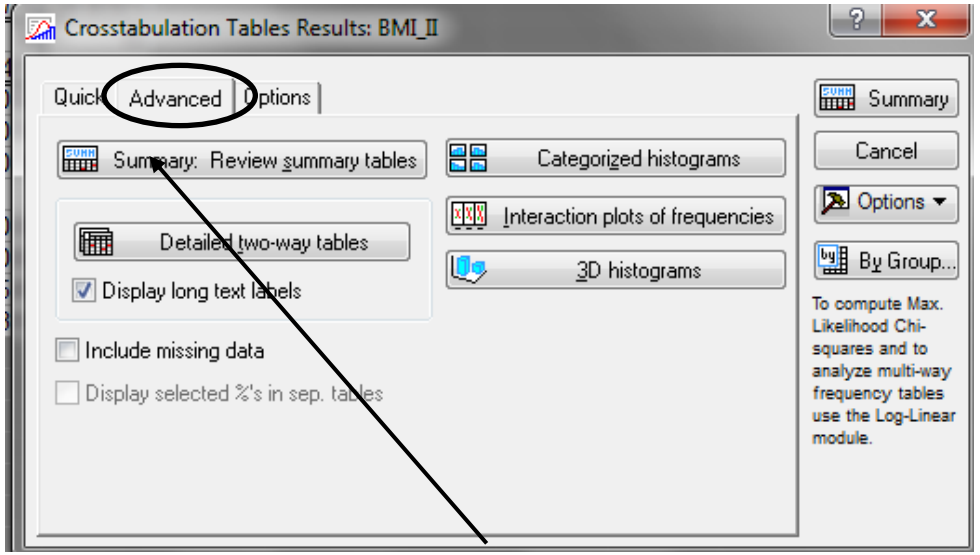
Výpočet testů

MU IBA

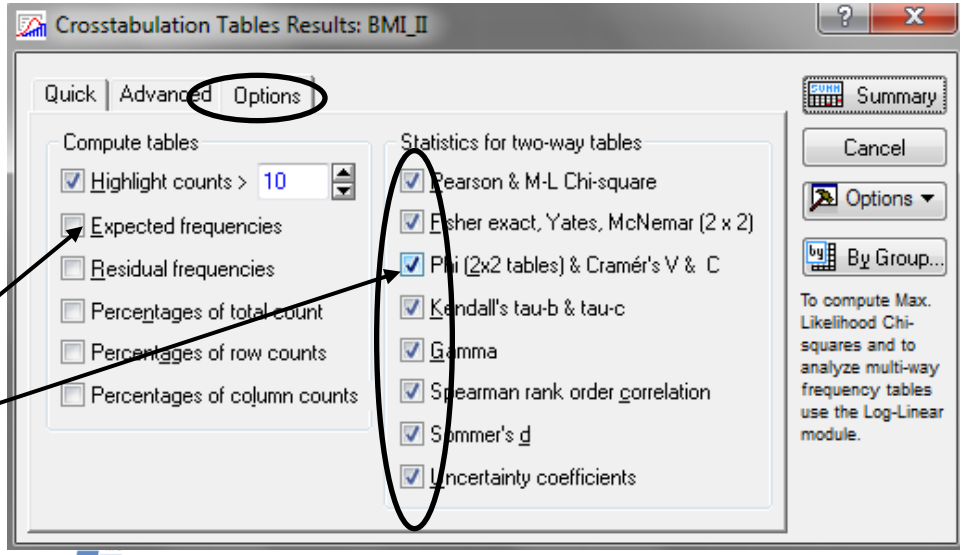
Kontingenční tabulky rozměrů n x m



Proměnné



Výpočet kontingenční tabulky



Očekáváné četnosti

Testy

IBA

Grafy ve statistice

- Histogramy, boxploty, xy grafy