

Neparametrické statistické testy a jejich SW podpora

L. Ličman, K. Langová, J. Zapletalová



**Ústav lékařské biofyziky,
LF UP Olomouc**



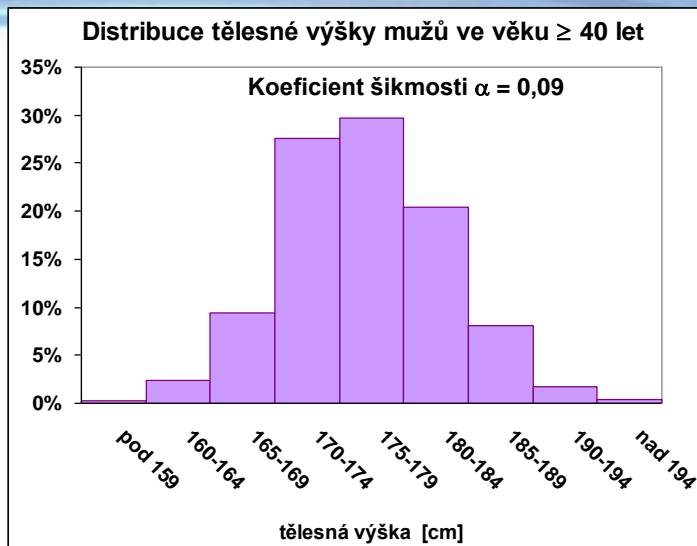
Neparametrické metody

- Pro analýzu dat, která nemají normální rozdělení
- Vhodné pro analýzu ordinálních dat
- Použití parametrů normálního rozdělení (průměr a rozptyl) není vhodné.
- Místo nich se používá medián a případně interkvartilové rozpětí (rozdíl mezi 3. a 1. kvartilem)
- Normalita dat se testuje testy normality (chí-kvadrát test dobré shody, Kolmogorov-Smirnov, Shapiro-Wilk, ...)
- Mnoho biomedicínských veličin nemá normální rozdělení (např. tělesná hmotnost)



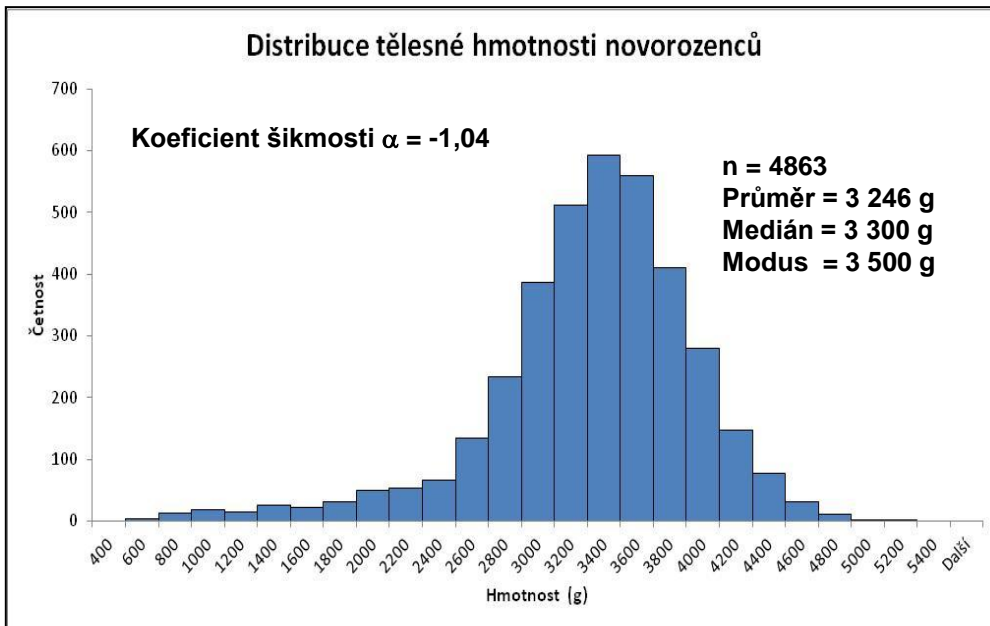
Příklady symetricky a asymetricky rozložených dat z našich klinických studií:

Symetrické rozložení

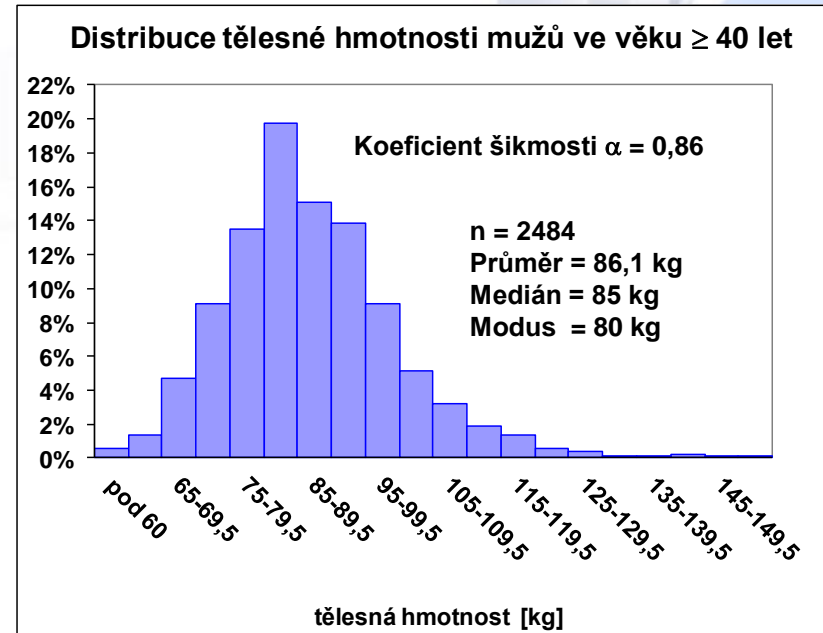


n = 2484
Průměr = 176,1 cm
Medián = 176 cm
Modus = 176 cm

Rozdělení zešikmené doleva



Rozdělení zešikmené doprava





SPSS

- IBM SPSS je standard ve statistickém zpracování dat
- Často používaný pro statistické zpracování dotazníků
- Velmi pohodlný ovládání testů pomocí formulářů
- Možnosti automatizace, je to i statistická platforma

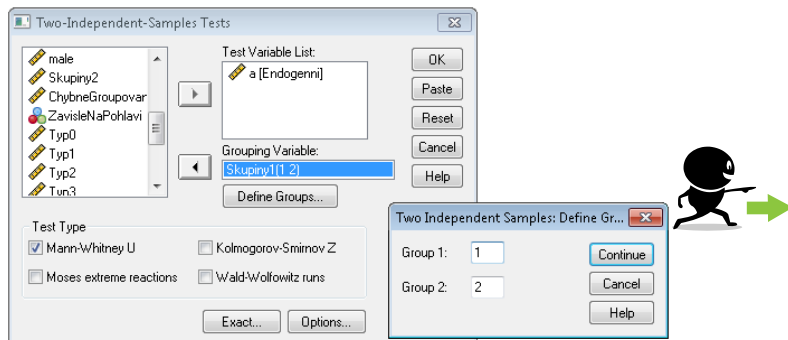


SPSS®



SPSS – Uživatelské ovládání

Uživatelské zadávání a výsledek neparametrického Mann-Whitney U-Testu



Output1 [Document1] - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output

- Log
- NPar Tests
- Notes
- Active Dataset
- Mann-Whitney Test
- Title
- Ranks
- Test Statistics

```
/M-W= Endogenni BY Skupiny1(1 2)
/MISSING ANALYSIS.
```

→ NPar Tests

[DataSet1] C:\Users\libor\SPSSv15\pro_scripty.sav

Mann-Whitney Test

| Ranks | | | | |
|-------|----------|----|-----------|--------------|
| | Skupiny1 | N | Mean Rank | Sum of Ranks |
| a | Prvni | 12 | 6,50 | 78,00 |
| | Druha | 12 | 18,50 | 222,00 |
| | Total | 24 | | |

| Test Statistics ^b | |
|--------------------------------|-------------------|
| | a |
| Mann-Whitney U | ,000 |
| Wilcoxon W | 78,000 |
| Z | -4,188 |
| Asymp. Sig. (2-tailed) | ,000 |
| Exact Sig. [2*(1-tailed Sig.)] | ,000 ^a |

a. Not corrected for ties.
b. Grouping Variable: Skupiny1

SPSS Processor is ready



Uživatelské ovládání-nevýhody

- Jsme závislí na autorech SW
- Musíme se „přizpůsobit“ jejich zadání a pojetí statistického testu
- Pokud potřebujeme výstup na míru našeho testu, přidat nějaký výsledek nebo kombinovat zároveň více testů, máme smůlu





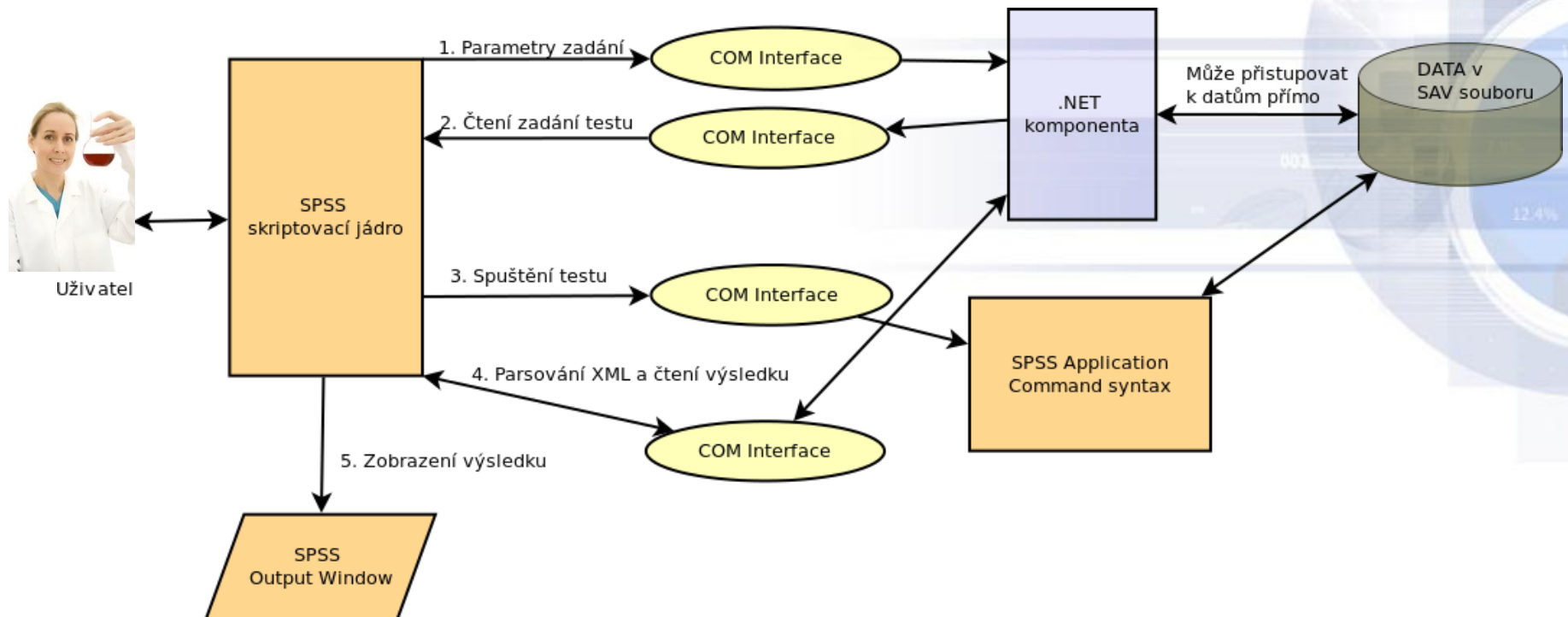
Programátorské ovládání softwaru

- **Výhodná specializace programátor a zároveň statistik**
- **V dnešní době se kombinuje data-mining se statistikou**
- **Programátorské ovládání jakéhokoliv SW umožňuje 100% využití jeho možností**



Nástroj pro automatizaci testování v SPSS

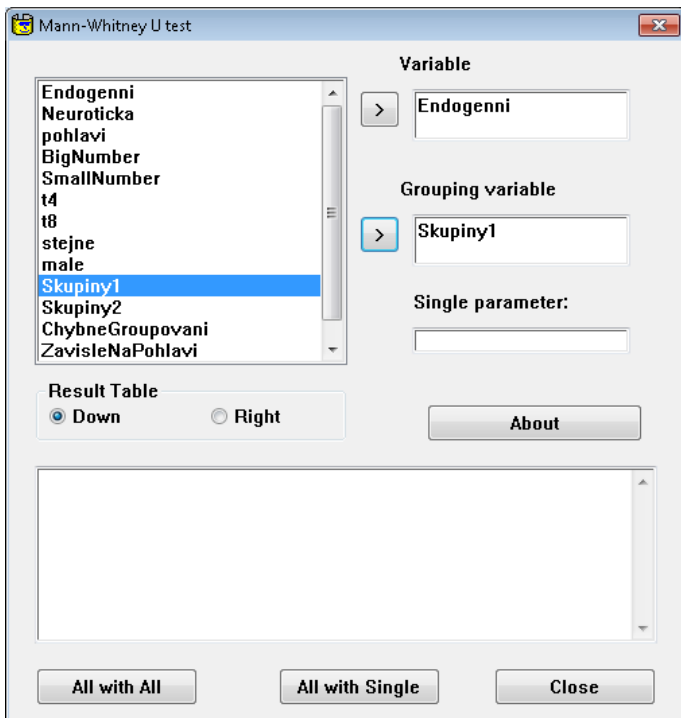
- Používá COM rozhraní
- Používá SPSS Command syntax language a SPSS macro language
- Vytvořen textový protokol pro zobrazení výsledků





Programátorské ovládání v praxi

- **Kruskal – Wallisův test s post hoc testy (Mann-Whitney) s Bonferroniho korekcí signifikance na míru konkrétním potřebám**



U test M:N (Endogenni:Skupiny1)

| U Test Info | |
|-------------------------|-----------|
| Variable | Value |
| Endogenni | Endogenni |
| Grouping variable | Skupiny1 |
| Grouping variable cases | 3 |
| Maximum number of tests | 3 |

| Case name | Variable Frequencies | | |
|-----------|----------------------|----|------------|
| | Numeric Value | N | Percent(%) |
| Prvni | 1 | 12 | 35,29 |
| Druha | 2 | 12 | 35,29 |
| Třetí | 3 | 10 | 29,41 |
| Sum | Sum | 34 | 100 |

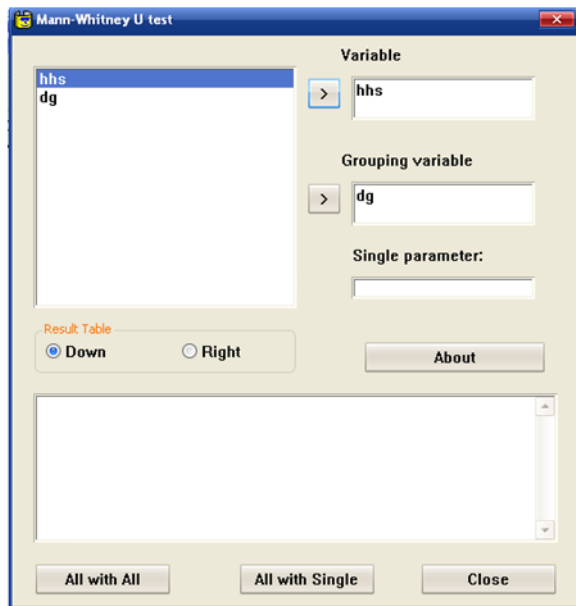
| Kruskal-Wallis Statistics | |
|----------------------------|---------|
| | Value |
| Kruskal-Wallis Chi-Square | 22,763 |
| Kruskal-Wallis Asymp. Syg. | <0.0001 |
| Kruskal-Wallis df | 2 |

| Possibility | U Test Statistics | | | | |
|-------------|-------------------|-----|--------|-------------|------------------|
| | U | W | Z | Sign 2 tail | Bonferroni corr. |
| Prvni:Druha | 0 | 78 | -4,188 | <0.0001 | <0.0001 |
| Prvni:Třetí | 0 | 78 | -3,995 | <0.0001 | ,0002 |
| Druha:Třetí | 57 | 135 | -,198 | ,843 | 1 |



Příklad využití

- Studie zabývající se dlouhodobým přežíváním Balgristovy jamky u pacientů s totální endoprotézou kyčle (ověření závislosti mezi předoperačním Harrisovým skóre a diagnózou)



| Case name | Variable Frequencies | | |
|-----------------------|----------------------|-----|------------|
| | Numeric Value | N | Percent(%) |
| primární osteoartróza | 1 | 35 | 26,72 |
| úraz | 2 | 16 | 12,21 |
| dysplastická kyčel | 3 | 74 | 56,49 |
| zánět | 4 | 6 | 4,58 |
| Sum | Sum | 131 | 100,00 |

| | Kruskal-Wallis Statistics |
|----------------------------|---------------------------|
| | Value |
| Kruskal-Wallis Chi-Square | 13,854 |
| Kruskal-Wallis Asymp. Syg. | ,003 |
| Kruskal-Wallis df | 3 |

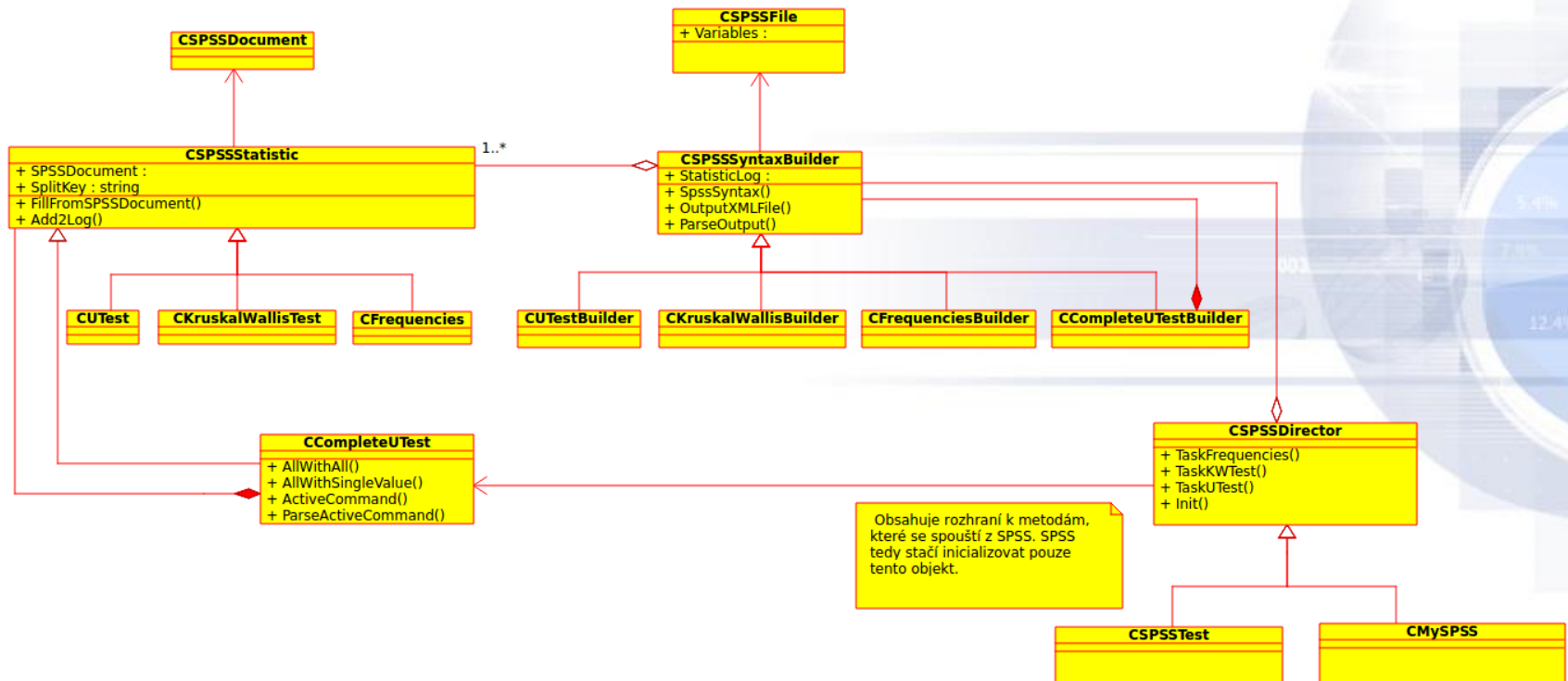
| Possibility | U Test Statistics | | | | |
|--|-------------------|---------|--------|-------------|------------------|
| | U | W | Z | Sign 2 tail | Bonferroni corr. |
| primární osteoartróza:úraz | 121,500 | 257,500 | -3,237 | ,001 | ,007 |
| primární osteoartróza:dysplastická kyčel | 1156 | 3931 | -,906 | ,365 | 1 |
| primární osteoartróza:zánět | 50,500 | 71,500 | -2,022 | ,043 | ,259 |
| úraz:dysplastická kyčel | 315,500 | 451,500 | -2,932 | ,003 | ,020 |
| úraz:zánět | 40,500 | 176,500 | -,561 | ,575 | 1 |
| dysplastická kyčel:zánět | 128,500 | 149,500 | -1,719 | ,086 | ,513 |

U test M:N (hhs:dg)

| | U Test Info |
|-------------------------|-------------|
| | Value |
| Variable | hhs |
| Grouping variable | dg |
| Grouping variable cases | 4 |
| Maximum number of tests | 6 |



- Použití OOP – jednoduché použití pro jakýkoli test
- Využití LINQ to XML pro parsování výsledků





Závěr

- **Podařilo se vytvořit komponentu, která je schopná vytvářet statistické testy v SPSS na míru řešeného problému**
- **Dá se jednoduše rozšiřovat pro další testy**
- **Může případně automaticky vytvářet SPSS soubory z různých zdrojů**
- **Může se kombinovat s R**



Závěr

- Děkuji za pozornost

