Základy biostatistiky

Práce v síti Windows NT

- **přihlásit se do sítě**
- **zřídit** svůj adresář na D:
- do něj zkopírovat data ze:
  tsntserver\home\zvara\biostat
  (otevřeno pouze pro čtení)
- stačí
  IQ1.s0, IQ1.s1, IQ.txt
  IQ3.s0, IQ3.s1, IQ4.s0, IQ4.s1
- *.txt – popis dat
  lze číst pomocí programu WordPad
  *.S0, *.S1 – data NCSS
  *.ZDB – jiný formát dat
  *.XLS – Excel

spustit **program NCSS 6.0**

- Number Cruncher Statistical System
  (to crunch – chroustat)
- zobrazení dat – vlastně dva listy Excelu 4.0
- tři okna (volí se z dolní lišty nebo Ctrl+):
  - data (Data) **Ctrl+E**
  - šablona (Template) **Ctrl+T**
  - výstup (Output) **Ctrl+W**
- **výstup** ve formátu *.RTF (lze do Wordu)
  - pamatuje si formátování
  - lze upravovat
  - umí obrázky
- uložení dat
  - běžná data (**Spreadsheet**)
    ukládat vždy aspoň dva soubory
    (*.S0, *.S1)
  - velká data formát (**Zipped Database**)
    (aspoň 500 případů nebo 50 sloupců)
    (*.ZDB)

Seznámení s NCSS

- **File**|**Open** zvol IQ1.s0
- tabulka s daty – **Sheet1**:
  - řádek – případ (žák)
  - sloupec – veličina (hodnoty znaku)
- co znamená prázdné políčko?
- nultý list tabulky – **Variable Info**
  - názvy veličin (krátké, začíná písmenem, bez mezer, možno ab_c)
  - doplňující popis (např. české označení)
  - transformace (logaritmy, podíly ...)
  - formát (počty des. míst, zápis data)
    (problémy s češtinou, nejjistější – prázdné)
  - typ (pro třídicí klíč dej 4)
  - textové vyjádření hodnot (je-li type=4)
    u POHLAVI odkaz na kod_pohl
    (kod_pohl, text_pohl) za sebou

nabídka v oknu Data

- **File**
  - načtení a uložení dat
  - import a export dat
  - tisk, ukončení ...
- **Edit**
  - kopírování, vkládání ...
  - též provedení transformace
- **Data**
  - třídění
  - možnost přípravy transformací
  - filtr
- **Analysis, Graphics**
  vlastní spouštění procedur
- **Window**
  přechod do jiného okna NCSS
- **Help** – záleží na kontextu

Graphics|Box Plots

NCSS Template (šablona)
upřesnění požadavků, kurzor přes celý řádek, k němu **dole návod**, **hodnota** se volí **nahoře**

- které veličiny (**Variable(s)**)
  - nutně jediná – přímo zvol po rozvinutí
  - může být veličin víc – **Select Variables**, rozviň a zvol (při několika **Ctrl**)
- upřesnění metody, popis obrázků ...
- případné třídění (**Grouping**)
- způsob popisu grafu (**Label Options**)
  - veličiny (**Variable Names**)
  - hodnoty (**Value Labels**)
- kam uložit případné nové veličiny
- které z nabízených metod vynechat ...
- rychlý přístup k položce (**Topic**)

Variables
Variable(s): **ZN8**
Grouping Variable: **POHLAVI**
Run – otevře se třetí okno

NCSS Output (výstup)

- dvě okna (dokumenty *.RTF):
  - běžný výstup (**Current Output**)
    po každém spuštění **Run** se **přepíše**
  - zápisník (**Log**)
    pomocí **File|Add Output to Log** lze na konec zápisníku přidat běžný výstup
- oba dokumenty (okna) lze editovat
  - vkládání poznámek
  - mazání části výstupu
  - volba fontů, velikosti, barvy
  - volba tabelátorů
- dokumenty lze ukládat jako soubory
- lze vkládat do textových editorů (Word)

- poklepáním na obrázek se získá kopie obrázku přes celou obrazovku
  po prohlédnutí vždy **zrušit** pomocí $\times$
- výstup lze modifikovat v šabloně
  - na dolní liště zelená kostka NCSS
  - **Topic|Label Options**
    Value Labels: Value Labels **Run**
    Variable Names: Var. Long Labels **Run**
  - **Topic|Vertical Tickmarks**
    Number of Major Ticks: 6 (vysvětlit)
    Number of Minor Ticks: 4 (vysvětlit)
- odečíst a poznamenat $x_{\text{min}}, Q_1, \tilde{x}, Q_3, x_{\text{max}}$ pro obě pohlaví
- **Data** (dolní lišta, žlutočervená)
  - kopie ZN8 dívek/hochů do volných sloupců
    doplň názvy, např. hoch, divka
  - **Data|Sort** (dívky)
  - spočítej $x_{\text{min}}, Q_1, \tilde{x}, Q_3, x_{\text{max}}$
  - totéž pro hochy

Data nebo **Template**
Analysis|Descriptive Stat.|Descriptive Stats

- **Variable(s)**: ZN8
  Group 1 variable: POHLAVI
- **Which Reports**
  Show Summary Section: Yes
  Show Quartile Section: Yes
  Show Stem-Leaf Section: Yes
  ostatní: No **Run**
- spočítali jsme stejné $x_{\text{min}}, Q_1, \tilde{x}, Q_3, x_{\text{max}}$?
- **Stem-Leaf Plot** (lodyha s listy) vysvětlit
- **Variable(s)**: hoch
  Group 1 variable: (none)
- uložit šablonu pro příští použití
  - **Notes**: známky chlapců
  - **File|Save Template** cviceni
  - **Run**
- **File|Add Output to Log**

Jakou průměrnou známku ZN8 musí mít nejhorší hoch, aby byla zobrazena jako odlehlá?

- $Q_3, \frac{3}{2}(Q_3 - Q_1)$
- zobraz krabicový diagram této kopie
  - **Template** ((zelená kostka)
  - **Graphics**|**Box Plot** hoch
  - **File**|**Save Template** cviceni
  - **Run**
  - **Data** 3,09 → 4,06, krabicový diagram
  - 3,09 → 4,07, krabicový diagram
- spočítej znovu popisné statistiky pro hoch
  - šablona pro popisné statistiky
  - **File**|**Load Template** cviceni
    a máme vše připraveno, stačí **Run**
  - srovnáme s původním výpočtem
    Window|**Tile Horizontal**

samostatně

- krabicový diagram pro IQ
  - upravit měřítko, popis os
  - lze změnit horní titulek?
  - lze změnit vybarvení?
- totéž s rozlišením pohlaví
  - upravit obrázek (popis os, titulek)
  - interpretovat
  - uložit (Add Output to Log)
  - podat (Data) soubor IQ4.s0
  - v čem je jiné srovnání krabicových diagramů?
  - lze změnit mez pro odlehlá pozorování?
- popisné statistiky pro IQ
  - Load Template, změň veličinu
  - interpretuj Stem-Leaf Plot
    (připomíná histogram?)