

Základy statistického zpracování údajů 2019/20

Obsah 1. hodiny (19. 2. 2020)

I. Náplň a cíle předmětu – syllabus

Předmět je koncipován jako praktický úvod do elementárních postupů statistického zpracování údajů. Látka je v hojně míře vysvětlována na příkladech. Metody statistické analýzy jsou demonstrovány na praktických cvičeních. Součástí předmětu je praktické používání tabulkového procesoru. Vyučovaná témata se mohou přizpůsobovat znalostem a potřebám studentů.

Hlavní témata

- 1) Statistické soubory a typy proměnných
- 2) Rozdělení četnosti a kontingenční tabulky
- 3) Vizualizace statistického souboru – histogramy
- 4) Vztah pravděpodobnosti a statistiky
- 5) Základní míry polohy a variability dat
- 6) Analýza závislosti proměnných
- 7) Pojem pravděpodobnostního rozdělení, binomické a normální rozdělení
- 8) Parametry populace a statistiky reprezentativních vzorků
- 9) Formulace a testování jednoduchých statistických hypotéz
- 10) Praktické metody analýzy a vizualizace dat pomocí programu Excel

II. Motivační příklad – Datový soubor `tenkrat-ve-skole`

Interpretace dat a typy proměnných. Žáci základní školy píší slohové cvičení. Na vzorku 100 žáků z různých tříd, z různých měst a různého pohlaví zkoumáme, kolikrát daný žák ve své slohové práci užil slovo „tenkrát“ a kolikrát užil slovo „tehdy“.

V souboru máme 100 záznamů, každý se skládá z pěti údajů.

- **město** kategoriální hodnota, množství hodnot není apriori omezeno
- **pohlaví** binární hodnota
- **třída** numerická diskrétní hodnota z intervalu 5..9
- **tenkrát** celočíselná nezáporná hodnota, typicky malé přirozené číslo
- **tehdy** celočíselná nezáporná hodnota, typicky malé přirozené číslo

III. Elementární praktická cvičení se souborem `tenkrat-ve-skole`

1. Otevřete soubor `tenkrat-ve-skole.xlsx` programem Excel. List „orig-data“ zkopírujte a kopii pojmenujte „data“. Dále pracujte s listem „data“. Kopie na listu „orig-data“ bude představovat zálohu původních dat.
2. Mezi městy se vyskytuje chybně slovo „Plzen“. Automatickou náhradou upravte celý sloupec tak, že tam bude správné „Plzeň“.
— Nápověda: použijte funkci „najdi a nahrad“.
3. Zjistěte, kolik různých měst je v prvním sloupci uvedeno.
— [Jak by se postupovalo v případě velkého množství záznamů?](#)

4. K tabulce přidejte další sloupec s údajem „tenkrat+tehdy”.
— Náповěda: použijte vzoreček.
5. Zjistěte, kolik žáků ve své práci vůbec nepoužilo ani slovo „tenkrát” ani slovo „tehdy”.
— Náповěda: použijte vzorečky a pomocný sloupec.
6. Kolik žáků ve své práci použilo slovo „tenkrát” vícekrát než slovo „tehdy”?
— Náповěda: použijte funkci „když” (if).
7. Jaké je procentuální zastoupení žáků podle jednotlivých měst? Znázorněte to
 - a) tabulkou,
 - b) koláčovým diagramem,
 - c) zvlášť pro dívky a pro chlapce.
8. Spočítejte celkový počet užitých slov „tenkrát” a „tehdy” pro každý ročník (třída) zvlášť. Výsledky zobrazte
 - a) tabulkou,
 - b) grafem.
9. Ve sloupci pohlaví obarvěte automaticky buňky jinou barvou pro chlapce a jinou pro dívky.
— Náповěda: použijte funkci „Podmíněné formátování” (Conditional formatting).
10. Zkopírujte celou tabulku na nový list s názvem „chlapci-Praha”. Pak tabulku zredukujte pouze na chlapce z Prahy, kteří chodí do páté nebo šesté třídy a kteří vůbec nepoužili slovo „tehdy”. Sloupce zachovejte všechny, pouze vymažte nerelevantní řádky.
— Náповěda: Nemažte jednotlivé řádky.
Představte si, že tabulka je veliká.
Použijte funkci „Řadit data” (Sort).

IV. Doporučená literatura (kterou patrně nepotřebujete)

- Pecinovský: 333 tipů a triků pro Microsoft Excel 2013. Computer Press, 2014.
- Pecinovský: Excel 2013 podrobný průvodce. Grada, 2013.

V. Požadavky na zápočet

- pravidelná docházka, plnění domácích úkolů
- v zápočtovém týdnu (= poslední týden v semestru) bude povinný písemný test
 - praktický úkol s použitím programu Excel (cca 40–50 minut)
 - vyhodnoceno bude hned na místě