

## STMG1 2010/2011 ... Úkol č. 2

**Zadání:** *Konstrukce tabulky intervalového rozdělení četností pomocí dvou různých způsobů*

*Postup:*

- 1) V tabulce *podklady.xls* najdete podle svého příjmení konkrétní proměnnou, kterou použijete k vypracování úkolu a název příslušné tabulky, ve které se daný statistický soubor nalézá (složka *tabulky*). Jedná se vždy o data vztahující se k okresům ČR jako statistickým jednotkám. V tabulkách se vyskytují i souhrny pro celé kraje – pozor, tyto hodnoty nelze použít, když náš statistický soubor tvoří okresy. Pro kontrolu – okresů je 77, tudíž byste měli mít na počátku 77 čísel ( $n=77$ ).
- 2) Vytvořte z datové řady uvedené u vašeho jména celkem 2 tabulky intervalového rozdělení četností, a to následujícím způsobem:


*Použijte u každé tabulky rozdělení četností 5 tříd ( $k = 5$ )*

- a) u první tabulky intervalového rozdělení četností použijte pro výpočet hranic intervalů pevnou šířku třídy odvozenou z rozpětí souboru ( $h = R/k$ )
- b) u druhé tabulky intervalového rozdělení četností použijte pro výpočet hranic intervalů metodu stejného počtu hodnot v každém intervalu, tzn. hranice intervalů budou tvořeny kvantily: vymezení **5-ti** tříd získáte jako
  1. (minimum – 20%-ní kvantil)
  2. (20% – 40%-ní kvantil)
  3. (40% – 60%-ní kvantil)
  4. (60% – 80%-ní kvantil)
  5. (80%-ní kvantil – maximum).
- 3) Vypočítejte střední hodnotu jednotlivých tříd u obou tabulek.
- 4) Pomocí nástroje *Histogram (Nástroje – Analýza dat – Histogram)* a vámi vypočítaných hranic intervalů dotvořte tabulky rozdělení četností a histogramy *absolutních* četností s kumulativní křivkou *relativních* četností pro obě možnosti (a, b). Graf a tabulka ve stavu, ve kterém ji Excel „vrátí“ není úplně správná, musíte je ještě upravit, tzn. odstranit přebytečné řádky „Další“ v grafu a upravit osy grafu (měřítko vedlejší osy y a popisky osy x – aby se popisek vztahoval přesně k tomu místu, kde dané číslo na ose leží!! – například použijte pro popisky osy x vámi vypočítané středy jednotlivých tříd)

**Zadání:** *Vyjádření prostorového rozložení zkoumaného jevu v mapě ve formě kartogramu*

V šabloně *Mapa ČR.doc* je mapa okresů České republiky. Zpracujte do ní rozmístění hodnot vámi zpracovávané proměnné (jednotlivým okresům je přiřazována barva výplně podle předem stanovené legendy na základě hodnoty proměnné v daném okrese). Použijte hranice intervalů z metody a) i b) (tzn., nejdříve vytvoříte mapu pomocí dat z a), pak druhou mapu pomocí dat z b)). V obou šablonách, kromě přiřazování výplně okresům, také **doplňte** aktuální název mapy, popis proměnné, hranice intervalů, **relevantní sumarizující charakteristiku** pro celou ČR (suma, nebo průměr, nebo jiná – závisí na typu Vaší proměnné), autora (vaše jméno, příjmení, studovaný obor) a metodu výpočtu intervalů (konstantní šířka intervalů odvozená z rozpětí souboru, stejný počet hodnot v každém intervalu). Pro zjednodušení je vše co máte doplnit nebo změnit v šabloně vyznačeno **červeným písmem** – po doplnění vašich dat **změňte barvu písma zpět na černou**).

### Postup:

- 1) Aktualizujte název mapy (název mapy musí vždy obsahovat informaci co, kde a kdy)
- 2) Aktualizujte název proměnné a jednotky, v kterých je uvedena (např. *Počet obyvatel na km<sup>2</sup> plochy okresu*, *Podíl orné půdy na celkové rozloze okresu v %*); může se částečně opakovat i v názvu mapy. V zásadě byste měli vycházet z údajů ve zdrojových tabulkách
- 3) Vypočítejte relevantní charakteristiku vyjadřující intenzitu jevu v celé ČR (záleží od proměnné, někdy je to suma, někdy průměr,...) a aktualizujte tuto informaci v mapě
- 4) Doplněte své jméno, příjmení, obor, metodu konstrukce intervalů
- 5) Barvy můžete použít ty, které jsou přednastavené v legendě (zelená škála), nebo si můžete vytvořit vlastní barevnou škálu. Potom ale musíte dodržet pravidlo – čím intenzivnější jev, tím tmavší barva (nejlépe škála v jedné barvě od světlé po tmavou). Pokud tedy chcete, změňte si barvy v legendě.
- 6) Aktualizujte hranice intervalů v legendě mapy – v jedné mapě získané metodou *a*), v druhé mapě metodou *b*)
- 7) Přiřaďte každému okresu barvu z legendy podle toho, do kterého intervalu daný okres patří (kliknete na okres pravým tlačítkem myši - *Formát automatického tvaru – Barvy a čáry*, rychle funguje tlačítko *Kopírovat formát* ).
- 8) Po vytvoření obou map **vizuálně zhodnoťte**, zda se pro grafické znázornění vaší proměnné hodí spíše zařazení okresů do intervalů s konstantní šířkou (metoda *a*)): *vhodné* – pro rovnoměrně rozložená data – se souměrným platykurtickým rozdělením četností, *není vhodné* – pro data se šikmým a / nebo leptokurtickým rozdělením četností (tzn., pokud máte v jedné třídě většinu okresů, a v ostatních jen málo, v mapě převládá 1 barva). Pokud není vhodné, zhodnoťte, jestli je vhodnější použití kvantilů jako hranic intervalů (metoda *b*): zde může být problémem šířka intervalů v legendě – některé intervaly mohou být úzké v kontrastu s jinými velmi širokými, což může stěžovat čtení mapy (pokud daný okres podle barvy v legendě spadá do příliš širokého intervalu, lze si jen velmi špatně představit skutečnou hodnotu znázorňovaného jevu v okrese). Pokud ani tato metoda nevyhovuje, zkuste navrhnout jiné řešení výpočtu intervalů pro váš kartogram. **Hodnocení a odpovědi uveďte do textu emailu, ve kterém budete posílat vyřešený úkol.**

*Můžete se nechat inspirovat příkladem (příklad.pdf).*

Poznámka k datům v tabulkách:

ležatá čárka (-) v tabulce v místě čísla značí, že jev se nevyskytoval

tečka (.) v místě čísla značí, že údaj není k dispozici nebo je nespolehlivý

křížek (x) značí, že zápis není možný z logických důvodů

*Co odevzdávat:* Emailem na [veronika.smolkova@osu.cz](mailto:veronika.smolkova@osu.cz) (v předmětu emailu uvádějte číslo úkolu) pošlete řešení: **první část úkolu** v souboru *vaseprijmeni.xls* (přejmenovaný na vaše příjmení); **druhou část úkolu** ve **dvou** souborech, pojmenovaných podle použité metody třídění dat do intervalů: *vaseprijmeni\_a.doc*, *vaseprijmeni\_b.doc*; doporučuji odevzdat nejdříve první část úkolu, až pak, pokud bude správně, zpracovat a odevzdat druhou část.

*Termín odevzdání správného řešení:* do konce zápočtového týdne – tzn. **do 17. 12. 2010** – musí být odevzdáno správné řešení úkolu, takže pokud si nejste jisti, odevzdávejte úkoly radši v předstihu, abyste stihli provést případné opravy. Na pozdější odevzdání úkolu nebude brán zřetel (přijdete o možnost vykonání zkoušky).

Veronika Smolková

1. 11. 2010

*Pozn.: Zdroj dat k úkolu – Český statistický úřad*