Posudek vedoucího na bakalářskou práci

Autorka bakalářské práce: Eva Pirohová

Název bakalářské práce: Popisná statistika v R s aplikací na reálná data

Cílem bakalářské práce Evy Pirohové je představit základy popisné statistiky a možnosti její implementace v R s využitím reálných dat. Tyto cíle byly splněny postupně v kapitolách 1, 2 a 3.

Konkrétně v kapitole 1 je představeno R a RStudio a základní příkazy jazyka R. Tato kapitola představuje ilustrační a motivační úvod do možností jazyka R a práce v prostředí RStudio. Nelze ji považovat za vyčerpávající. To však ani nebylo cílem práce (vzhledem k tomu, že se jedná o práci bakalářskou) a vedlo by to k výraznému přesažení „povoleného" rozsahu práce.

Za velice pěknou a zdařilou považuji kapitolu 2, kde je hezký úvod do různých typů statistických proměnných (kvantitativních spojitých, kvantitativních diskrétních, ordinálních, nominálních) a práce s nimi (výpočet měr polohy, variability apod.). Na této kapitole velmi oceňuji, že pro každý typ proměnné (kvantitativní, ordinální, nominální) je přístup pro výpočet měr polohy a variability provedený odděleně. Kladně také hodnotím to, že se studentka vypořádala s naprogramováním mnoha vlastních funkcí (pro výpočet např. měr variability nominálních a ordinálních proměnných apod.). Diskuze je doplněna i představením relativně bohaté palety grafů. Vždy je také uvedeno, s využitím jakých Rkovských funkcí je možné jednotlivé charakteristiky či grafy implementovat v softwaru R.

Kapitola 3 představuje praktickou implementaci uvedených přístupů na zajímavém reálném datovém souboru. Postupy v této kapitole jsou vesměs v pořádku. Tato praktická část poskytuje hezký návod, jak přistupovat k popisné analýze dat v R, počínaje načtením vlastních dat, odstraněním chybějících pozorování, pokračující výpočtem požadovaných charakteristik a získání grafických výstupů. Jediné, co mírně postrádám, je ještě detailnější interpretace výsledků z věcného hlediska a to z důvodu, aby bylo zřejmé, co výsledky prakticky znamenají (v kontextu daného datového souboru) a že nejsou samoúčelné. Přitom použitý datový soubor je velmi hezký a přímo vyzývá k podrobnější věcné/praktické diskuzi a interpretaci výsledků.

Kvalita matematického značení má s ohledem na to, že se jedná o bakalářskou práci, nadprůměrnou úroveň. Proměnné jsou značeny správnými symboly a kurzívou, význam použitého značení je téměř vždy vysvětlen a použité značení je konzistentně používáno v celé práci.

Jako v každé práci se i v této vyskytlo několik chyb a nejasností (ale tomu nelze zamezit v žádné práci). Jako příklad je možné uvést divné výsledky ohledně variačního a mezikvartilového rozpětí u ordinální proměnné na str. 34. Úvod do práce s R a RStudiem (v kapitole 1) je jen základní, tj. neposkytuje podrobnější pohled do možností R a RStudia. V tomto ohledu je však třeba mimo jiné zohlednit to, že se autorka s R a RStudiem seznámila a začínala pracovat teprve v okamžiku výběru tématu bakalářské práce a neměla tudíž možnost získat dostatečný nadhled pro prezentaci možností R a RStudia.

Celkový dojem z bakalářské práce je pěkný. Studentka přistupovala k jejímu zpracování s pečlivostí a se zaujetím, půjčila si a pročetla též i několik knih v angličtině, aby práci se softwarem R porozuměla.

Prosím, aby studentka v průběhu obhajoby reagovala na následující dotazy (řadím je v pořadí dle priority):

- Prosím vysvětlete, jak správně počítat variační a mezikvartilové rozpětí pro ordinální proměnnou. Okomentujte i výsledky (a jejich správnost) pro variační a mezikvartilové rozpětí na str. 34.
- Je funkce na.omit() na str. 30 implementována optimálně? Nedojde (při tom, jak byla tato funkce užita v bakalářské práci) k vyřazení i těch pozorování, jež mají chybějící údaje u proměnných, jež se v analýze ve finále nepoužijí, a to i přesto, že u těch proměnných, jež se v analýze použijí, jsou

údaje u těchto pozorování kompletní? Jaký by byl optimálnější způsob vyřazení chybějících pozorování?

- Na str. 20 říkáte, že *nejvhodnější* charakteristikou variability pro kvantitativní proměnnou je variační rozpětí a kvartilové rozpětí. Z čeho usuzujete, že se jedná o *nejvhodnější* charakteristiky?

Cíle bakalářské práce Evy Pirohové byly splněny. Bakalářskou práci hodnotím jako zdařilou a nadprůměrně kvalitní, byť se v ní místy vyskytují i drobné nesrovnalosti. Bakalářskou práci proto **doporučuji k obhajobě** a v případě úspěšného průběhu obhajoby a schopnosti studentky reagovat v průběhu obhajoby na dotazy doporučuji hodnotit tuto práci stupněm **výborně**.

30. května 2017,

Mgr. Milan Bašta, Ph.D.
Katedra statistiky a pravděpodobnosti, FIS, VŠE, Praha