Cvičení 4 - Diskrétní náhodné veličiny

Datový soubor `head.txt` obsahuje antropometrické údaje mladých dospělých lidí, převážně studentů vysokých škol z Brna a Ostravy (Králík, nepublikovaná data).

`sex`	pohlaví jedince (m - muž, f - žena)
`sexor`	sexuální orientace (op - výlučně opačné pohlaví, sa - ostatní)
`obra`	existence staršího bratra (yes - ano, no - ne)
`body.H`	výška postavy (v mm)
`head.L`	délka hlavy, vzdálenost mezi body *glabella* a *opisthocranion* (v mm)
`head.W`	šířka hlavy, vzdálenost obou bodů *euryon* (v mm)
`bigo.W`	šířka dolní čelisti, vzdálenost obou bodů *gonion* (v mm)
`bizyg.W`	šířka tváře, vzdálenost obou bodů *zygion dakryon* (v mm)

1. Zpracujte datový soubor `head.txt`.
   (a) Obsahuje nějaké chybějící hodnoty? Kolik jich je?
   (b) Kolik chybějících hodnot obsahuje proměnná `sexor`? Jaké části datového souboru to odpovídá?
   (c) Liší se četnosti chybějících hodnot u proměnné `sexor` mezi muži a ženami (proměnná `sex`)?
   (d) Řádky s chybějícími hodnotami odstraňte a dále pracujte s úplným souborem.
2. Zkoumejte rozdělení výšky postavy (proměnná `body.H`).
   (a) Zobrazte data pomocí boxplotu a histogramu a výsledky interpretujte.
   (b) Jaká je průměrná výška postavy?
   (c) Jaká je nejmenší uvedená výška postavy?
   (d) Objevuje se v datech někdo, kdo měří více než 2 metry?
   (e) Jaká je výška postavy 5 procent nejmenších, resp. 10 procent nejvyšších lidí?
3. Existuje souvislost mezi délkou a šířkou hlavy?
   (a) Své tvrzení podpořte vhodným grafem.
   (b) Své tvrzení podpořte vhodnou číselnou charakteristikou.
4. Liší se šířka dolní čelisti mezi muži a ženami?
   (a) Své tvrzení podpořte vhodným grafem.
   (b) Své tvrzení podpořte vhodnými číselnými charakteristikami.
5. Existuje souvislost mezi pohlavím a existencí staršího bratra?
   (a) Své tvrzení podpořte vhodným grafem.
   (b) Své tvrzení podpořte vhodnou číselnou charakteristikou.
6. V rámci studie poměru pohlaví u lidí z roku 1889 bylo na základě záznamů z nemocnic v Sasku zaznamenáno rozdělení počtu chlapců ve čtrnáctičlenných rodinách. Mezi 6 115 rodinami s 12 dětmi byla pozorována četnost chlapců. Údaje ze studie jsou uvedeny v souboru `chlapci.txt`.
   (a) Najděte rozdělení pravděpodobnosti popisující počet chlapců.
   (b) Jaká je pravděpodobnost narození chlapce?
   (c) Porovnejte pozorované a očekávané četnosti. Data společně s modelem vhodně zobrazte (vizualizujte).
   (d) Nakreslete grafy odhadnuté pravděpodobnostní a distribuční funkce.
   (e) Na základě modelu spočtěte, jaká je pravděpodobnost, že náhodně vybraná rodina má přesně 9 chlapců?
   (f) Na základě modelu spočtěte, jaká je pravděpodobnost, že náhodně vybraná rodina má 6, 7, nebo 8 chlapců?
   (g) Na základě modelu spočtěte, jaká je pravděpodobnost, že náhodně vybraná rodina má nejvýše 4 chlapce?
   (h) Na základě modelu spočtěte, jaká je pravděpodobnost, že náhodně vybraná rodina má alespoň 8 chlapců?