

Statistika (STT): Cvičení 2

1. Najděte v datovém souboru `eyecolorgenderdata.csv` numerickou proměnnou (tj. proměnnou v poměrovém měřítku), která nejvíce koreluje s proměnnou "watchtv". Pokud možno, zvažte u některých proměnných (podle tvaru jejich rozdělení), zda je před analýzou transformovat, tj. např. vytvořit novou proměnnou jako $\log_2(\text{Proměnná} + 1)$. Závislost těchto dvou proměnných vyjádřete graficky i číselně a výsledky interpretujte.
2. Najděte v datovém souboru `eyecolorgenderdata.csv` numerickou proměnnou (tj. proměnnou v poměrovém měřítku), mimo "height", u které lze vidět vliv pohlaví na její polohu nebo variabilitu. Polohu i variabilitu této proměnné pomocí vhodných charakteristik a grafu/ů pro obě pohlaví vypočtete a porovnejte.
- 3.* Pro následující data, popisující věkovou strukturu a pohlaví nemocných, nakreslete graf/y a spočítejte charakteristiku/y závislosti mezi pohlavím a věkem nemocných a výsledky interpretujte. Je tedy mezi nemocnými (velký) rozdíl mezi věkovou strukturou mužů a žen? (Nápověda: lze spočítat jen např. "Pearson's coefficient of contingency" nebo i nějakou pořadovou míru korelace?)

Počet osob s laboratorně prokázaným onemocněním COVID-19 ke dni: 22. 2. 2022 v 8.24 h.:

Věková skupina	Ženy	Muži	Celkem
0-14	280 595	296 826	577 421
15-24	210 658	210 790	421 448
25-34	247 323	245 748	493 071
35-44	330 898	297 720	628 618
45-54	315 891	282 185	598 076
55-64	192 993	182 897	375 890
65-74	121 417	113 023	234 440
75-84	69 379	51 887	121 266
85+	33 513	15 137	48 650

- 4.* Převedte proměnnou "year" na numerickou proměnnou nazvanou např. "yearNum" a najděte vhodný regresní model závislosti "miles" na "yearNum" včetně grafického znázornění. Výsledky interpretujte.