

10 Kontingenční tabulky – domácí práce.

Rkové soubory na Isu.

data_vlasy.RData ... učitel statistiky R.D. Snee sbíral data na své domovské univerzitě a zaznamenal pohlaví studentů, barvu vlasů a barvu očí.

data_tuzky.RData ... naše data o pastelkách, analyzujeme tloušťku a ostrost (nezlomených) pastelek.

data_divizna.RData ... smyšlená data o výskytu divizen na 30 plochách v sezónách 2018 a 2019. Plocha byla označena jako „květ“, pokud na ní rostla alespoň 1 kvetoucí divizna, jako „list“, když divizna sice rostla druhým rokem, ale nekvetla, a jako „růžice“, pokud na ploše byly jen prvoroční přízemní růžice.

Termín odevzdání pátek 20. prosince.

① **Data_vlasy**: pomocí Rkových příkazů vytvořte kontingenční tabulku a její marginální součty:

- a) pohlaví (řádky) x barva vlasů (sloupce)
- b) barva vlasů x barva očí.

Do textu vložte příkazy i výsledné tabulky.

② Pomocí χ^2 testu testujte obě tabulky. Stejný test používáme pro dvě různé hypotézy: o nezávislosti a o homogenitě (o struktuře). Přiřaďte ke každé tabulce rozumnou hypotézu a slovně formulujte výsledky. Pojmy „nezávislost“ a „homogenita“ rozvedte: nezávislost čeho, homogenita čeho?

③ **Data_tuzky**: sestrojte čtyřpolní tabulku, vyslovte hypotézu, testujte a formulujte odpověď.

④ **Data_tuzky**: spočtěte Cramérův koeficient asociace (stačí podle vzorečku, není nutné stahovat extra knihovnu). Je tedy asociace mezi tloušťkou a ostrostí tužky? A jaké v číslech vidíme „tendence“, co ta asociace může popisovat, které vlastnosti „chodí spolu“?

⑤ **Data_divizna**: zajímá nás, jestli struktura populace divizen na sledovaných plochách byla srovnatelná v roce 2018 s rokem 2019. Jaký test použijete? Jakou hypotézu test testuje? Jak interpretujeme výsledek?