

MAD - pracovní list 2

Průzkumová analýza a vizualizace dat

1. část - trocha povídání na úvod a společná práce

Vizualizace dat je jednou z možností, jak si vytvořit představu o souboru více-rozměrných dat. To nám pomáhá identifikovat vektory, které vybočují z ostatních (odlehlá pozorování), identifikovat různé struktury v datech (shluky,...), či udělat si představu o vícerozměrné normalitě dat, která je jedním ze standartních (a nutno dodat, že většinou nezbytných) předpokladů pro další práci s daty.

Mezi základní metody zobrazování vícerozměrných dat patří **zobecněné rozptylové diagramy** a **symbolové grafy**.

Z webu si stáhněte soubor `lide.txt` a zdrojový kód `vizualizace.R` a otevřete si jej v **R**. Načtěte data ze souboru `lide.txt`.

• Zobecněné rozptylové grafy

Pro dvourozměrný náhodný výběr termín rozptylový graf značí vztah mezi první a druhou složkou. Jedná se tedy o bodový graf dvojic pozorování. Pro vícerozměrný vektor pozorování už není možné takový graf zakreslit (dokonce ani pro troj-rozměrný vektor nemá takový graf valný smysl, i když jej zakreslit můžeme, neboť takový obrázek je již značně zkreslený). Zobecněný rozptylový graf je tedy bodový graf všech dvojic složek náhodného výběru. Vzniká tak vlastně matice obrázků (stačila by její trojúhelníková část).

• Symbolové grafy

Symbolových grafů existuje velké množství. Mezi nimi si uved'me **profily**, **polygony** a pro zajímavost například **Chernoffovy obličejové znaky**.

Obecně nelze říci, že by nějaká metoda byla vysloveně lepší, či horší než jiná. Záleží to na situaci, charakteru dat a také vkusu.

2. část - samostatná práce v R, "opakování statistiky"

Z webové stránky si stáhněte soubor `IQ.txt`. Načtěte z něj data do programu R. Pro takto získaná data proveďte

- vizualizaci: vykreslete zobecněný rozptylový graf, profily, hvězdicové polygony a Chernoffovy obličejové diagramy.
- zjistěte sumární statistiky pro jednotlivé složky vektorů výběru.

- pro jednotlivé složky náhodného vektoru vytvořte histogramy.
- nyní se zaměříme na data z TESTU 2, TESTU 4 a IQ. Pro tato data vytvořte okékový graf.
- proložte regresní přímku daty pro IQ (vysvětlující proměnná) a TESTEM 2 (závislá proměnná).