

1. průběžná písemka z MAS10c - vzorová

Datový soubor `Dotaznik.txt` obsahuje data o 300 respondentech sesbíraná formou individuálních dotazníků ohledně životního stylu. Obsahuje tyto proměnné:

<code>Pohlavi</code>	pohlaví respondenta (M – muž, Z – žena)
<code>Vek</code>	věk respondenta
<code>BMI</code>	hodnota BMI daného respondenta
<code>SrdecniTep</code>	klidová tepová frekvence daného respondenta
<code>Koureni</code>	kouří daný respondent (Ano – kuřák, Ne – nekuřák)
<code>PocetDeti</code>	počet dětí daného respondenta

1. (10 bodů)

- Kolik pozorování a kolik proměnných datový soubor obsahuje?
- Pro každou proměnnou uveďte jakého je datového typu (nominální, ordinální, diskrétní, spojitá).
- Obsahují data nějaké chybějící hodnoty? Kolik jich je?
- Řádky s chybějícími hodnotami odstraňte.

2. (10 bodů) Vyberte si jednu kategoriální proměnnou a tu prozkoumejte.

- Uveďte a interpretujte vhodné číselné charakteristiky.
- Hodnoty této proměnné vhodně vizualizujte. Výsledek opět nezapomeňte okomentovat.

3. (10 bodů) Existuje souvislost mezi věkem respondenta a jeho hodnotou BMI?

- Své tvrzení podpořte vhodným grafem, který nezapomenete okomentovat.
- Své tvrzení podpořte vhodnou číselnou charakteristikou. Její hodnotu interpretujte.

4. (10 bodů) Modelujte proměnnou `PocetDeti` pomocí Poissonova rozdělení.

- Odhadněte jeho parametr λ z dat.
- Porovnejte pozorované a očekávané četnosti. Data společně s modelem vhodně zobrazte.
- Nakreslete graf odhadnuté pravděpodobnostní funkce.

5. (10 bodů) Modelujte proměnnou `SrdecniTep` pomocí normálního rozdělení.

- Odhadněte jeho parametry z dat a interpretujte jejich hodnoty.
- Na základě modelu spočítejte, jaká je pravděpodobnost, že respondent bude mít klidovou tepovou frekvenci vyšší než 60.
- Na základě modelu spočítejte, jaká je pravděpodobnost, že respondent bude mít klidovou tepovou frekvenci mezi 65 a 75.
- Na základě modelu spočítejte, jaká je klidová tepová frekvence, kterou má 30 procent lidí s nejvyšší tepovou frekvencí.