

## MA012 – otázky k písemné zkoušce

1. Analýza rozptylu (ANOVA): jednoduché / dvojně třídění (bez / s interakcemi) – předpoklady, rovnice odpovídajícího modelu, efekty, hypotézy a řetězec submodelů, ANOVA tabulka a interpretace; metody mnohonásobného porovnávání – princip a testované hypotézy, interpretace Scheffého a Tukeyovy metody; princip ANOVA – rovnice rozkladu variability na součty čtverců, stupně volnosti, testovací statistika, zápis ANOVA tabulky.
2. Neparametrické testy: uspořádaný náhodný výběr a (průměrná) pořadí; znaménkový test – výpočet pořadí, podmínky, nulová hypotéza; jednovýběrový Wilcoxonův test – výpočet pořadí, podmínky, nulová hypotéza; dvouvýběrový Wilcoxonův test – výpočet pořadí, podmínky, nulová hypotéza; Kruskalův-Wallisův test – podmínky, nulová hypotéza, interpretace, porovnání s jednoduchým tříděním.
3. Testování rozdělení pravděpodobnosti: test dobré shody při známých / neznámých parametrech – empirické a teoretické četnosti, testovací statistika a její rozdělení pravděpodobnosti; test dobré shody pro diskrétní / spojitou náhodnou veličinu – algoritmus volby kategorií, empirické a teoretické četnosti, podmínka dobré aproximace; empirická distribuční funkce a Kolmogorovův-Smirnovův test – testovací statistika a geometrický význam.
4. Korelační analýza: Pearsonův korelační koeficient – definice, význam; Spearmanův korelační koeficient – definice, použití; Kendallův korelační koeficient – definice, použití, konkordantní a diskordantní páry; model vícenásobné lineární regrese – rovnice modelu s vysvětlením označení, geometrická interpretace; mnohonásobná korelace, (semi)parciální korelace – definice (vysvětlení použitých veličin z vícenásobné lineární regrese), interpretace.
5. Porušení předpokladů lineárního modelu: autokorelace – popis autokorelace, odhad parametrů zobecněnou metodou nejmenších čtverců, Durbinova-Watsonova statistika a test; multikolinearita – popis multikolinearity, VIF, test multikolinearity a testovaná hypotéza.
6. Zobecněný lineární model (GLM): definice GLM – rovnice zobecněného lineárního modelu, lineární prediktor a linkovací funkce, přirozený parametr, příklady linkovacích funkcí; předpoklady GLM a porovnání GLM s lineárním regresním modelem, věrohodnostní funkce, deviance, testování vhodnosti submodelu; logistická regrese – použití, rovnice modelu, podíl šancí, linkovací funkce pro binomická data.
7. Analýza hlavních komponent (PCA): princip PCA, algoritmus nalezení hlavních komponent, volba počtu hlavních komponent.

## MA012 – zadání písemné zkoušky

U zkoušky obdržíte teoretické otázky ze 4 okruhů z výše uvedeného seznamu a 1 početní úkol.

V početním úkolu budete na konkrétně zadaných číselných výběrech malého rozsahu řešit jednu z úloh: výpočet pořadí a testovací statistiky  $U$  ve znaménkovém, nebo jednovýběrovém, anebo dvouvýběrovém Wilcoxonově testu, výpočet pořadí a Spearmanova korelačního koeficientu, nalezení konkordantních a diskordantních párů a výpočet Kendallova korelačního koeficientu. Povolena a doporučena je kalkulačka, níže uvedené vzorce budete mít k dispozici.

Na zadání zkoušky budete mít k dispozici následující vzorce pro výpočet znaménkového, jednovýběrového a dvouvýběrového Wilcoxonova testu a Spearmanova a Kendallova korelačního koeficientu (při neopakujících se pozorováních):

$$S^+ = |\{i : X_i > x_0\}|, \quad U = \frac{2S^+ - n}{\sqrt{n}}.$$

$$S^+ = \sum_{Y_i > 0} R_i^+, \quad S^- = \sum_{Y_i < 0} R_i^+, \quad U = \frac{S^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}.$$

$$T_1 = \sum_{i=1}^m R_i, \quad T_2 = \sum_{j=m+1}^{m+n} R_j, \quad U_1 = mn + \frac{m(m+1)}{2} - T_1, \quad U_2 = mn - U_1, \quad U_{MW} = \frac{2 \min\{U_1, U_2\} - mn}{\sqrt{\frac{mn(m+n+1)}{3}}}.$$

$$r_s = 1 - 6 \frac{\sum_{i=1}^n (R_i - S_i)^2}{n(n^2 - 1)}; \quad \tau = \frac{n_+ - n_-}{n_0} \quad (\text{při neopakujících se pozorováních}).$$