

## Metody regresní analýzy

Předmět SZZ NMS Aplikované matematicko-stochastické metody

### Otázka č. 1:

**Jednoduchá lineární regrese, odhady parametrů a jejich vlastnosti, intervaly spolehlivosti a testy hypotéz o parametrech modelu, intervaly predikce**

- Model jednorozměrné lineární regrese, MLE odhady parametrů a jejich vlastnosti (nestrannost, rozdělení, ...), Gauss-Markov teorém
- Intervaly spolehlivosti a testy hypotéz o parametrech beta, definice  $R^2$  a  $F$  statistiky, rozklad čtverců
- Intervaly spolehlivosti pro střední hodnotu  $EY_0$  a intervaly predikce pro  $Y_0$  v novém pozorování  $x_0$

### Otázka č. 2:

**Vícerozměrná lineární regrese, analytické a numerické řešení normálních rovnic, koeficient determinace, F-test, intervaly predikce**

- Model vícerozměrné lineární regrese, maximálně věrohodné odhady parametrů, soustava normálních rovnic a její řešení
- Vlastnosti odhadů parametrů (nestrannost, rozdělení, nezávislost), Gauss-Markov teorém
- Koeficient determinace  $R^2$  a jeho vlastnosti, obecná lineární hypotéza a  $F$ -test
- Intervaly spolehlivosti pro  $EY_0$  a intervaly predikce pro  $Y_0$  v novém pozorování vysvětlující proměnné

### Otázka č. 3:

**Rezidua a jejich grafická analýza, detekce odlehlých a influenčních pozorování, transformace závislé a nezávislé proměnné**

- Interně a externě studentizovaná rezidua, grafy reziduí, partial residual plots, partial regression plots (added variable plots), PRESS rezidua
- Míry influence, DFBETAS, Cookova vzdálenost, pravidlo pro detekci influenčních pozorování, DF-FITS
- Transformace vysvětlované proměnné  $y$ , Box-Cox transformace, transformace vysvětlujících proměnných  $x$ , Box-Tidwell metoda
- Vážené nejmenší čtverce

### Otázka č. 4:

**Výběr regresního modelu, informační kritéria, kroková regrese a sestupný výběr proměnných**

- Kritéria porovnávání modelů, koeficient determinace, MSE,  $F$ -test pro vnořené modely
- Mallowsova  $C_p$  statistika, AIC, BIC - obecná definice, vlastnosti, použití, PRESS statistika
- Metody výběru modelu, zpětná eliminace, dopředná regrese, postupná regrese, princip marginality

**Otázka č. 5:**

**Model ANOVA, model ANCOVA s jedním faktorem, multikolinearita a její detekce, hřebenová regrese**

- a) Model analýzy rozptylu (ANOVA) a jeho použití, model analýzy kovariance (ANCOVA) s jedním faktorem
- b) Kolinearita, její důsledky a detekce, index podmíněnosti matice, definice VIF
- c) Potlačení kolinearity, hřebenová regrese - princip metody a volba parametru  $\delta$

**Otázka č. 6:**

**Exponenciální rodina distribucí a její vlastnosti, podmínky regularity, definice zobecněného lineárního modelu**

- a) Definice exponenciální rodiny distribucí, tvar hustoty vhodný pro ZLM a její vlastnosti (momentová vytvářející funkce, střední hodnota, rozptyl)
- b) Podmínky regularity, skórový vektor, Fisherova informační matice
- c) Definice zobecněného lineárního modelu, příklady ZLM

**Otázka č. 7:**

**Odhadování parametrů zobecněných lineárních modelů, maximálně věrohodné odhady, numerické metody výpočtu, asymptotické rozdělení odhadů**

- a) Maximálně věrohodné odhady parametrů ZLM, obecný tvar věrohodnostních rovnic, Fisherova informační matice pro ZLM, odhad parametru měřítka  $\phi$
- b) Numerické metody výpočtu MLE, algoritmy Newton-Papson a Fisher-scoring, metoda iterovaných vážených nejmenších čtverců
- c) Asymptotické rozdělení skórového vektoru a maximálně věrohodných odhadů

**Otázka č. 8:**

**Míry adekvátnosti zobecněného lineárního modelu, porovnávání modelů, analýza reziduí a influenčních pozorování**

- a) Míry adekvátnosti ZLM - saturovaný model, deviační statistika a její vlastnosti
- b) Porovnávání vnořených modelů - deviační a Waldova statistika, porovnávání nevnořených modelů - statistiky AIC a BIC
- c) Diagnostika modelu - projekční matice v ZLM, typy reziduí (Pearsonova, Anscombeova, deviační a jejich standardizované verze), grafy reziduí, ověření vhodnosti použité spojovací funkce
- d) Analýza influenčních pozorování - Cookova vzdálenost

**Otázka č. 9:**

**Modely pro binární data, logistický, normální a Gumbelův model, interpretace parametrů modelu, testy a rezidua**

- a) Modely pro binární data - Bernoulliův a binomický model, nejužívanější spojovací funkce
- b) Model logistické regrese, interpretace jeho parametrů, poměr šancí
- c) Míry kvality binomického modelu - deviační statistika, Pearsonova statistika, test Hosmer-Lemeshow, tvar reziduí v binomickém modelu (Pearsonova, deviační) a jejich použití

**Otázka č. 10:**

**Poissonovská regrese, pravděpodobnostní modely pro kontingenční tabulky, log-lineární modely**

- a) Model poissonovské regrese, spojovací funkce a interpretace parametrů modelu, relativní riziko, deviační statistika a rezidua poissonovského modelu
- b) Pravděpodobnostní rozdělení pro kontingenční tabulky, log-lineární modely a odhady jejich parametrů
- c) Test hypotézy nezávislosti a homogenity v kontingenční tabulce a jejich ekvivalent v poissonovském modelu, interpretace interakcí druhého a třetího řádu v log-lineárním modelu