

# Zadania programov

## z predmetu Analýza časových radov

Tomáš Bacigál

### Zásady vyhotovenia programov.

- Technickú správu štrukturujte do kapitol (každému zadaniu vyhradte jednu) a do podkapitol.
- Je odporúčané každý významnejší medzivýsledok *komentovať*.
- Vypracovaný a úplný program – obsahujúci zdrojový súbor (skript .R alebo notebook .Rmd/.qmd, prípadne iné), report (.html alebo pdf) a prípadne dátový súbor – odovzdajte v príslušnom “Mieste odovzdania” v AIS.
- Numerické výsledky sa snažte umiestňovať do tabuliek s popisom, najmä ak slúžia na porovnanie metód/modelov. Všetky grafy by mali byť riadne anotované.
- Zobrazte dôležité výsledky každého štatistického testu (testovacia štatistika, kritická hodnota príp. p-hodnota), napíšte obe hypotézy a zhodnotenie. Zhodnotenie *nesmie* byť automaticky generované programom.
- Prípadné zmeny v predošlom programe indikujte na začiatku aktuálneho zadania.
- Snažte sa o čo najjednoduchšie riešenie, ktorému aj rozumiete. Materiály poskytnuté cvičiacim majú inštruktážny účel, neslúžia ako vzor vypracovania zadanií.

### 1. Výber a popis časového radu. (1b)

1. Nájdite si vhodný (nestacionárny) časový rad, s ktorým budete pracovať v priebehu semestra. Načítajte údaje.
2. Slovné charakterizujte zvolený časový rad (pôvod, merné jednotky, kontext, dĺžka, vzorkovacia frekvencia ...).
3. Vykreslite časový rad s popisom osí (v absolútnej časovej mierke) aj hlavným nadpisom. Ak je k dispozícii časový rad so širším časovým rozpätím (teda váš ČR je jeho podmnožinou), zobrazte aj ten.
4. Vypočítajte najmenšiu a najväčšiu hodnotu, medián, strednú hodnotu, smerodajnú odchýlku a zobrazte ich v prehľadnej tabuľke.

5. Rozdeľte časový rad na tréningovú a vyhodnocovaciu vzorku (orientačne v pomere 80%-20% so zohľadnením kalendárnych jednotiek a periódy sezónnej zložky).

Veľký výber dát poskytuje napr. [datamarket.com](https://datamarket.com), zoznam zdrojov na [CRAN views](https://cran.r-project.org/web/packages/views/index.html), portál [Kaggle](https://www.kaggle.com), alebo databáza [Rdatasets](https://www.r-datasets.org).

## 2. Dekompozícia časového radu. (10b)

1. Zhodnoťte povahu časového radu – prítomnosť trendu a sezónnosti, zvážte potrebu stabilizácie rozptylu.
2. Odhadnite aspoň dva regresné modely pre trend, a vyberte ten, ktorý je pre skúmaný časový rad najvhodnejší (na základe popisných schopností i predpokladaného vývoja). Ak obsahuje skokovité zmeny, zahrňte do modelu aj zodpovedajúce indikačné premenné.
3. Ku rezíduám po trende vykreslite korelogram, a posúďte, či indikuje prítomnosť sezónnej zložky. Regresiou odhadnite parametre modelu sezónnej zložky
  - s indikačnými funkciami (ak nie je perióda príliš veľká vzhľadom ku dĺžke radu),
  - s goniometrickými funkciami, aj keď nie je prítomná (cvične), no ďalej s ňou už nepočítajte.

Zobrazte model v období jedinej sezóny (rok, mesiac, alebo deň) s vhodným značením osí tak, aby bol model ľahko interpretovateľný.

Sezónnu zložku vykreslite spolu s trendom na pozadí s tréningovou vzorkou.

4. Pre reziduá po odstránení sezónnej zložky vypočítajte a zobrazte spektrum. Navýznamnejšie zložky odhadnite a v tabuľke zobrazte ich amplitúdy. Rozhodnite, či budete tieto systematické zvyšky v časovom rade modelovať ako súčasť trendu (flexibilnejšou funkciou), sezonality (pridáte ďalšie násobky základnej frekvencie) alebo cyklickej zložky.
5. Odhadnite *všetky systematické zložky* v spoločnom regresnom modeli a zobrazte ho. Vyšetrite náhodnosť rezíduí (pomocou štatistických testov) a posúďte, či je ďalej potrebné modelovať ich autokorelačnú štruktúru pomocou lineárnych modelov stacionárnych časových radov.
6. Vyjadrite intenzitu každej systematickej zložky.
7. V časovom rade reziduálnej zložky pomocou neparametrického i parametrického prístupu identifikujte aspoň dva lineárne modely, odhadnite ich parametre, porovnajte ich reziduálny rozptyl a overte stacionaritu (koreňmi AR polynómu alebo simuláciou). Jeden model reziduálnej zložky zobrazte spolu s modelom systematických zložiek na pozadí s tréningovou vzorkou.

### 3. Diagnostika spoločného modelu a predpovede. (5b)

1. Odhadnite spoločný model *systematických a reziduálnej* zložky.
2. Na základe diagnostiky zhodnoťte vhodnosť všetkých alternatívnych modelov.
3. Vypočítajte 1-krokové a viackrokové bodové predpovede a každé zvlášť zobrazte na pozadí s validačnou vzorkou a naivnými predpoveďami. Vykreslite aj intervalové predpovede.
4. Vypočítajte stredné chyby predpovedí.
5. Zobrazte tréningovú, validačnú vzorku, bodové a intervalové viackrokové predpovede najlepšie modelu v jednom grafe. Zhodnoťte, či je schopnosť modelu predpovedať porovnateľná s popisnou schopnosťou a kde by mohol byť priestor na zlepšenie.

### 4. Modelovanie integrovaných procesov. (5b)

1. Predbežnou diagnostikou a štatistickými testami posúďte, či sa v pôvodnom časovom rade (okrem prípadného deterministického trendu a sezónnej zložky) nachádza stochastický trend.
2. Na pôvodnom časovom rade identifikujte niekoľko modelov ARIMA, SARIMA či ARFIMA (pomocou korelogramu aj informačných kritérií), odhadnite ich parametre a porovnajte z popisného a predpovedného hľadiska (jedno aj viac-krokové predpovede). Posúďte, či sú predpovede oproti naivnej predpovedi (prípadne s lin. trendom) presnejšie.
3. Zostavte tabuľku kvantitatívnych ukazovateľov popisných a predpovedných schopností a vykreslite modely časového radu (na tréningovej aj validačnej časti, v pôvodnej mierke). Vyberte jeden, (podľa Vás) najlepší model.