



SEMINÁŘ VÝPOČETNÍ STATISTIKY

P14

2008-05-19

ANALÝZA ČASOVÝCH ŘAD:

Časové řady v SASu:

- ✓ Vývoj nezaměstnanosti v EU15 v letech 1996 – 2007.
- ✓ Tři časové okruhy, na začátku datové okruhy nastaveny stejně:
 - Data Range
 - Fit Range
 - Evaluation Range
- ✓ **Useknutý úsek** by měl být tvořit cca 1/4 délky časové řady. Forecast Horizont je horizont **předpovědi**, kam až bychom měli vycházet, je to max. 1/3 délky referenčního období.
- ✓ **Testy** s nabídkou yes, no a maybe, varianty, co všechno se má otestovat. Tečkou se označí to, co SAS zjistil a provedl úpravu.
- ✓ **Složky:**
 - **Trend** – časová řada bude vykazovat systematický nárůst nebo pokles či střídání, pokud se časová řada nemění, řada stagnuje, trend je konstantní, tedy žádný není
 - **Periodická složka** – pravidelně se opakující výkyvy, nemusí v časové řadě být
 - **Reziduální = náhodná** – ostatní zbývající vlivy, které se nepodařilo postihnout trendem a periodickou složkou
- ✓ **Testy** – charakteristiky: logaritmická transformace, trend, sezónnost
- ✓ **Odmocnina ze střední kvadratické chyby** (směrodatná odchylka) – přednastavené kritérium, které využíváme, pokud pracujeme pouze s jednou řadou, pokud by byly ve více řadách nestejné jednotky, nebylo by možné porovnávání. RMSE je odmocnina ze střední kvadratické chyby.
- ✓ Mean Absolute Percent Error – potřeba nastavit
- ✓ První ikonka na panelu vlevo umožní automatické vygenerování, nebo můžeme kliknout do prázdného bílého pole a vybereme nabídku na automatické generování, kterou potvrdíme. Objeví se informace, kolik vhodných modelů lze vygenerovat, lze si nechat zobrazit všechny.
- ✓ Vpravo je uvedena střední absolutní procentuální chyba, jsou seřazeny od nejmenší. Ne všechny modely, které splnily diagnostická kritéria, jsou vhodné.
- ✓ **Trendy:**
 - Log Linear Trend – log znamená, že řada není stacionární a SAS provedl potřebné úpravy, tedy hodnoty zlogaritoval a pak teprve provedl výpočet.
 - Denge Trend Exponential ???hing
 - „Model náhodné procházky“ – tzv. naivní model, konstruuje předpovědi jednoduše
- ✓ Tlačítko pro zobrazení grafu. V jeho pravé části jsou tlačítka. Tlačítko **beta** (β) skrývá další informace o modelu.
- ✓ **Další informace** (Parametr Estimates) – obsahuje vyrovnávací konstantu, hodnota je 0,99 – vzdálenější hodnoty jsou zařazovány s menší vahou, p-hodnota uvádí, zda je konstanta statisticky významná, pokud ano, můžeme model považovat za obecně platný. Blíží-li se hodnota 1, vstupují do konstrukce s větší vahou novější data.
- ✓ **Sigma kvadrát** – další kritéria, které by bylo možné využít, např. Mean Square Error, je zde také R-Square, tedy koeficient determinace, hodnota 0,45 – ze 45% model vystihnul kolísání časové řady.
- ✓ **Overfitting** – neparametrizování, může vyjít záporný, model, který je zbytečně složitý, za každou cenu se snaží procházet body časové řady, výstup je pak značně chaotický.
- ✓ Pomocí Set Ranges se vrátíme zpět a odstraníme zkrácení časové řady. Chceme-li využít pro předpověď konkrétní model, označíme ho, klikneme do pole a vybereme Refit Model, což vede ke zkonstruování **pro celou řadu** (ne pouze pro zkrácenou). Předpovědi budou vycházet z celé řady, mohou být grafické nebo numerické.
 - **Grafické výstupy** – využijeme ikonku s obrázkem časové řady proložené příslušným modelem a „vějířem“. Znovu se zobrazí původní řada, která za posledním rokem pokračuje předpovědí pro další roky. Jedná se o tzv. **bodovou předpověď**. Klikneme-li na skutečný bod, objeví se skutečný údaj – hodnota, rok. Klikneme-li na čáru modelu, zobrazí se Forecasted Value, tedy hodnota získaná příslušným modelem.



Kolem bodové předpovědi je tzv. konfidenční pás (pás spolehlivosti), což je 95% interval spolehlivosti, ve kterém by se hodnoty měly pohybovat.

- **Tabulková podoba** – vedle grafu ikonka s obrázkem tabulky. Tabulka má několik sloupečků – skutečné údaje (actual), predikované hodnoty (predict), vyžádané hodnoty konfidenčního pásu označeny U95 a U99. Sloupec Error představuje rozdíly skutečných a vyrovnaných hodnot, tedy rezidua.
- ✓ Nabídnutých modelů bylo víc, ale my jsme vybrali pouze jeden, je však možné v jednom vybraném využít informaci z ostatních modelů, lze vytvořit **agregovaný model kombinované předpovědi**. Po kliknutí do volného pole je v nabídce Combine Forecasts. Dále máme možnost vytvořit kombinaci z několika nejlepších modelů, jejichž MAPE jsou nejmenší. Vytvoří se průměr, lze dvojím způsobem, buď pomocí Fit Regression Weights (a potvrdit OK), a nebo se zvolí prostý aritmetický průměr, naklikáme tedy jména modelů, které chceme použít a vlevo od nich se ve sloupci Weight zobrazí váhy (u dvou 1/2, u tří 1/3 atd.). Často se stává, že tento model vyjde ze všech zcela nejlépe, ale i když ne, je konfidenční pás užší.
- ✓ Využívali jsme automatické generování, klasické modely se tam však neobjevovaly. Pokud nám však data připomínají nějaký model, například parabolu, lze ji získat. Klikneme v nabídce na Fit Arima Model a v následující tabulce je možné vybrat pomocí Add trendovou křivku, kterou jsme si zvolili.