Trendová analýza - přímka a parabola

Zadání příkladu:

Odhadněte parametry lineárního a kvadratického trendu u časové řady středního počtu obyvatelstva České republiky (ozn. $P(t)$) od roku 1989 do roku 1998 a rozhodněte, který je vhodnější. Dále odhadněte pomocí vhodnějšího modelu počet obyvatel v roce 1999 a 2000. Hodnoty $P(t)$ jsou uvedeny v tabulce 1.

Tab. 1 - Počet obyvatel ČR letech 1989-1998 v tisících - ozn. P_t

Rok	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
P_t	10362	10363	10309	10318	10331	10336	10331	10315	10304	10295

Zdroj dat: Český statistický úřad (ČSÚ)

Vypracování příkladu

V trendové analýze je výstupem matematický model ČŘ, kde je sledovaná veličina y_t funkcí času t a náhodné složky ϵ_t. To lze zapsat jako $y_t = f(t, \epsilon_t)$. Reálnou hodnotu y_t můžeme vyjádřit jako $y_t = Y_t + \epsilon_t$, kde $Y_t = f(t)$ je tzv. *teoretická (modelová) hodnota* ukazatele. Klasický model dále rozkládá ČŘ na čtyři složky časového pohybu:

- Trendová složka T_t - dlouhodobá tendence ve vývoji ČŘ
- Sezónní složka S_t - pravidelně se opakující odchylka od trendu, periodicita max. 1 rok
- Cyklická složka C_t - odchylka od trendu s periodicitou větší než 1 rok
- Náhodná složka ϵ_t - náhodné odchylky, které nemají systematický charakter.

Podle způsobu rozkladu existují dva modely:

- **Aditivní** $\quad y_t = T_t + S_t + C_t + \epsilon_t$
- **Multiplikativní** $\quad y_t = T_t.S_t.C_t.\epsilon_t$

V dalším postupu bude použit model aditivní.

Přímka - lineární trend

Jedná se o jednoduchý, běžně používaný model ve tvaru:

$$T_t = \alpha_0 + \alpha_1 t \quad t = 1, \ldots, n$$

Odhady parametrů α_0 a α_1 označíme a_0 a a_1. Výpočet provedeme pomocí **metody nejmenších čtverců** známé z regresní analýzy. Odhad $\hat{T}_t$ trendové složky T_t vyjádříme jako

$$\hat{T}_t = a_0 + a_1 t$$

Odhady spočteme podle vzorců analogických vzorcům z regrese, a to:

$$a_1 = \frac{\overline{ty_t} - \bar{t}.\overline{y_t}}{\overline{t^2} - (\bar{t})^2}$$

$$a_0 = \overline{y_t} - a_1.\bar{t}$$

Připomeňme pouze, že parametr a_1 se česky nazývá **směrnice**, anglicky **slope**, parametr a_0 je potom **kvocient**, anglicky **intercept**.

Parabola - kvadratický trend

Model trendu má tvar

$$T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \qquad t = 1, \ldots, n$$

Vztahy pro odhady a_0, a_1, a_2 lze opět získat pomocí *soustavy normálních rovnic*, uvádět je zde však nebudeme.

Program STATGRAPHICS Centurion XVII

V programu STATGRAPHICS je pro analýzu trendu ČŘ vytvořena funkcionalita v menu *Forecast/User-Specified Model*. Do povinného pole **Data** je třeba zadat název proměnné, kde jsou hodnoty sledované veličiny y_t. Do nepovinného pole **(Time Indices:)** je možné zadat odpovídající hodnoty času. Po kliknutí na tlačítku OK se objeví další formulář *Model Specification Options*. Zde v rámu *Type* zvolíme **Linear Trend** a klikneme na tlačítko OK. V dalším formuláři *Tables and Graphs* přidáme tabulku *Forecast Table* a potvrdíme tlačítkem OK.
Pokud je třeba odhadnout parametry kvadratického trendu, je postup stejný, pouze zvolíme typ trendu **Quadratic Trend**.
Pro porovnání vhodnosti modelů slouží různé charakteristiky. Například MSE (Mean Squarred Error), RMSE (Root Mean Squarred Error) nebo MAE (Mean Absolute Error) nebo tzv. F-statistiku. MSE lze vypočítat podle vzorce

$$MSE = \frac{\sum_{t=1}^{n}(y_t - \hat{T}_t)^2}{n}$$

Statistiku F potom podle

$$F = \frac{\frac{\sum_{t=1}^{n}(\bar{y} - \hat{T}_t)^2}{p-1}}{\frac{\sum_{t=1}^{n}(y_t - \hat{T}_t)^2}{n-p}}$$

kde p je počet odhadovaných parametrů modelu. Platí, že čím je MSE nižší, tím je model vhodnější. U F-statistiky je tomu naopak - čím je její hodnota vyšší, tím je model vhodnější. Pokud **modely mají různý počet parametrů**, je nutné použít tzv. **upravený index determinace** I^2_{adj}, který nestejné počty parametrů zohledňuje - viz kapitola o regresní analýze.

Vypočítají se podle těchto vztahů

$$I^2 = \frac{\sum(\hat{T}_t - \overline{y})^2}{\sum(y_t - \overline{y})^2}$$

$$I^2_{adj} = 1 - (1 - I^2)\frac{n-1}{n-p}$$

Výsledky:
Nejprve vytvoříme graf (obr. 1) - postup viz téma 9, příklad 1.

Lineární trend
Odhady parametrů: a_0 = 10357,8, a_1 = 5,71. Model má proto tvar T_t = 10357,8 - 5,7091t.

Nyní je nutné zkontrolovat, zdali jsou parametry modelu statisticky významné. V rámu *Analysis Summary* jsou v tabulce **Trend Model Summary** uvedeny odhady všech parametrů a příslušné hodnoty **P-value**. Pokud je hodnota P-value pro daný parametr nižší než zvolená hladina významnosti α (ta bude vždy 5 %, pokud nebude stanoveno jinak), je daný parametr statisticky významný (nelze přijmout hypotézu, že je roven nule). Zde je P-value(a_0) = 0,00, P-value(a_1) = 0,0118. Oba parametry jsou proto statisticky významné.

Výpočet předpovědí počtu obyvatel lze provést tak, že do vypočteného lineárního modelu dosadíme příslušnou hodnotu t a vypočteme $\hat{Y}_t$. Pro rok 1999 je $t = 11$, pro rok 2000 je $t = 12$, takže $\hat{Y}_{11}$ = 10295 a $\hat{Y}_{12}$ = 10289. Ve STATGRAPHICS existuje ve výše zmíněné funkcionalitě *Forecast/User-Specified Model* tabulka **Forecast Table**, kde program vypočte předpověď pro libovolný časový okamžik. Požadovaný počet předpovědí lze nastavit ve formuláři Forecasting při zadávání proměnných pro analýzu (standardně 12).

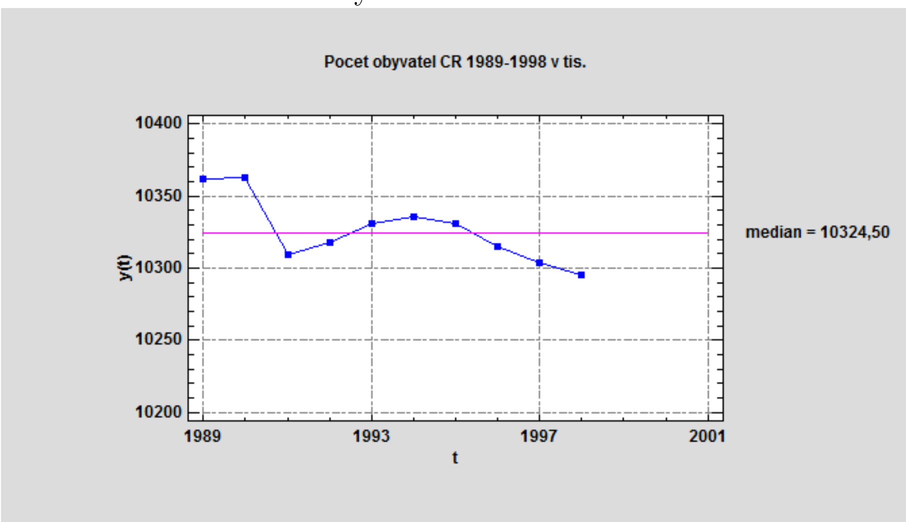
Kvadratický trend
Odhady parametrů: a_0 = 10360,9, a_1 = -7,25, a_2 = 0,14.
Model má tvar T_t = 10360,9 - 7,25t + 0,14t^2.

Hodnoty P-value pro odhadované parametry jsou následující:
P-value(a_0) = 0,00, P-value(a_1) = 0,42 a P-value(a_2) = 0,86. Z toho plyne, že je statisticky významný pouze kvocient, takže o kvadratickém trendu nemá smysl uvažovat.

Interpretace výsledků

Výsledkem jsou dva modely. Lineární model T_t = 10357,8 - 5,7091t a kvadratický model T_t = 10360,9 - 7,25t + 0,14t^2. Testy významnosti regresních parametrů ukázaly, že u kvadratického modelu je významný pouze kvocient a_0, takže se fakticky nejedná o kvadratický trend. U lineárního modelu jsou naopak významné oba odhadované parametry. Pokud bychom měli interpretovat hodnotu a_1, je možné říci, že za každý rok počet obyvatel klesne přibližně o 5709. Předpovědi počtu obyvatel pro rok 1999 a 2000 jsou: 10295 a 10289. Pro srovnání, skutečný počet obyvatel ČR v uvedených letech činil 10283 a 10272 tis. obyvatel.

Obrázek 1: Počet obyvatel ČR v letech 1989-1988 v tis.



Zdroj: Údaje ČSÚ, vlastní zpracování