HANČLOVÁ, J., TVRDÝ, L. (2003). Úvod do analýzy časových řad. Ostrava: VŠB-TU

1 Úvod

Cílem analýzy časových řad je většinou konstrukce vhodného modelu. Sestrojení „dobrého" modelu nám zpravidla umožní porozumět mechanismu, na jehož základě vznikají hodnoty časové řady, a pochopit podmínky a vazby, které působí na vznik těchto hodnot. Na základě změn těchto podmínek či vazeb lze simulovat jejich vliv působící změny ve vývoji časové řady.

Dalším cílem je využití těchto získaných poznatků při předpovědi budoucího chování. Používané postupy jsou založeny na principu, že "historie se opakuje". Tento předpoklad bývá v praxi splněn s různou přesností, a proto je vhodné u vyhlazování a předpovědí v časových řada uvádět i spolehlivost získaných výsledků a hodnotit úspěšnost predikce.

2 Teoretické základy pro analýzu časových řad

2.1 Základní pojmy

Časovou řadou rozumíme posloupnost hodnot ukazatelů, měřených v určitých časových intervalech. Tyto intervaly jsou zpravidla rovnoměrné (ekvidistantní), a proto je můžeme zapsat následujícím způsobem:

$y_1, y_2, \ldots, y_n$ neboli $y_t,\ t = 1, \ldots, n,$

kde y značí analyzovaný ukazatel, t je časová proměnná s celkovým počtem pozorování n.

2.1.1 Druhy časových řad

Časové řady členíme *podle charakteru ukazatele*:

- *okamžikové* - hodnota ukazatele k určitému okamžiku t (např. počet evidovaných uchazečů),
- *intervalové* - velikost sledovaného ukazatele závisí na délce intervalu, za který je sledován (např. měsíční náklady na rekvalifikace).

Podle *druhu ukazatelů* rozlišujeme časové řady obsahující:

- *absolutní* ukazatele (očištěné),
- *odvozené* ukazatele (součtové, poměrové).

2.1.2 Grafická analýza

Analýza časových řad se v současnosti provádí výhradně na počítačích pomocí vhodného softwaru. Velká většina statistických a ekonometrických softwarů má algoritmy těchto analýz zabudované ve svých standardních nabídkách. Bohužel program EXCEL mezi ně nepatří, proto se budeme muset věnovat relativně jednoduchým algoritmům, které lze vysvětlit. Pro pokročilejší analýzy časových řad doporučujeme statistické softwary: SPSS, STATISTICA, S +.

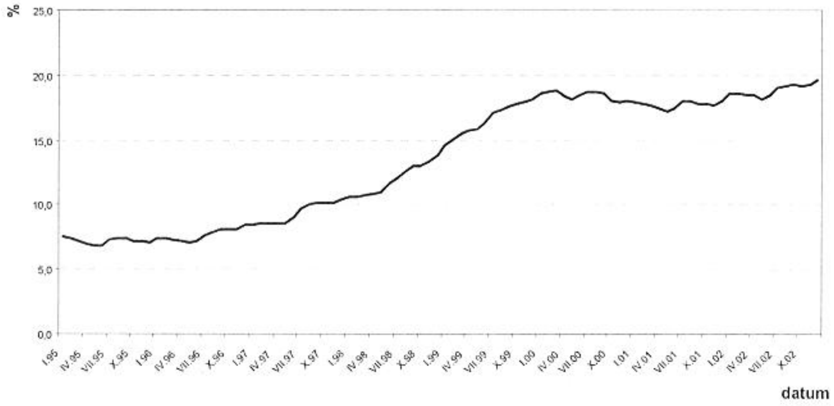
V programu EXCEL je nejvhodnější datovou strukturou pro časové řady standardní **datová matice** ve které je první řádek tvořen krátkým názvem proměnné a potom následují naměřené hodnoty. Jeden řádek datové matice obsahuje pozorování v jednom časovém okamžiku. Hodnoty jsou seřazeny podle času, vzestupně. Ukázka datové matice v EXCELu uvádí tabulka 1, která zahrnuje vývoj měsíční míry nezaměstnanosti v Karviné (%) – *u_KI* za období leden 1995 – březen 1996.

Tab. 1: Datová matice vývoje měsíční míry nezaměstnanosti v Karviné (*u_KI* v %)

Datum	*t*	*Rok*	*Měsíc*	*u_KI*
I.95	1	1995	1	7,53
II.95	2	1995	2	7,38
III.95	3	1995	3	7,18
IV.95	4	1995	4	7,00
V.95	5	1995	5	6,84
VI.95	6	1995	6	6,91
VII.95	7	1995	7	7,30
VIII.95	8	1995	8	7,37
IX.95	9	1995	9	7,42
X.95	10	1995	10	7,18
XI.95	11	1995	11	7,19
XII.95	12	1995	12	7,10
I.96	13	1996	1	7,40
II.96	14	1996	2	7,37
III.96	15	1996	3	7,29

Kromě proměnné *t* výše definované se obvykle používají další **časové proměnné** dle typu časových řad. Pokud pracujeme s ročními údaji je vhodné zavést další proměnou *rok*. U čtvrtletních dat kromě proměnné *r* i proměnou *q*, jenž nabývá hodnot 1 až 4 podle čtvrtletí. A analogicky postupujeme i u měsíčních údajů. Vedle těchto numerických proměnných se používá v programu EXCEL i proměnná ve formátu datum např. ve tvaru *I.99* pro grafické znázornění časových řad.

Pro **zobrazení časových řad** a jejich prvotní analýzu slouží *spojnicové grafy*. Vodorovná osa u těchto grafů zaznamenává časovou proměnnou a na svislé ose se zobrazují hodnoty ukazatele časové řady y_t. Příkladem spojnicového grafu vývoje míry nezaměstnanosti v okrese Karviná v letech 1995 až 2002 je obr. 1.



Obr. 1: Vývoj míry nezaměstnanosti v okrese Karviná

Spojnicový graf může zahrnovat i více časových řad, avšak měřítko na svislé ose je stejné.

Dalším důležitým grafem v EXCELu je *graf XY bodový*, který sleduje vývoj časové řady y_t na vývoji hodnot časové řady x_t tzn., že znázorní bod se souřadnicemi $[x_t, y_t]$ pro každý časový okamžik t. Tento typ grafu je vhodný u regresní analýzy.

2.1.3 Popisné charakteristiky

Charakteristiky polohy (průměry)

Při práci s časovými řadami je někdy důležité zjistit jejich průměrné hodnoty:

- *prostý aritmetický průměr* $\overline{y} = \dfrac{\sum_{t=1}^{n} y_t}{n}$;
- *vážený aritmetický průměr* $\overline{y} = \dfrac{\sum_{t=1}^{n} v_t y_t}{\sum_{t=1}^{n} v_t}$, kde v_t je váha ukazatele y_t v čase t;
- *vážený chronologický průměr* $\overline{y}_{ch} = \dfrac{\frac{y_1 + y_2}{2} d_2 + \frac{y_2 + y_3}{2} d_3 + \cdots + \frac{y_{n-1} + y_n}{2} d_n}{d_2 + d_2 + \cdots d_n}$, kde d_t je délka jednotlivých časových intervalů.

Charakteristiky variability

Nejdůležitější míry variability ve statistice patří rozptyl a směrodatná odchylka:

- *rozptyl* je aritmetickým průměrem kvadrátů odchylek od aritmetického průměru: $s_y^2 = \frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2$;

- *směrodatná odchylka* je odmocninou z rozptylu $s_y = \sqrt{s_y^2} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(y_t - \bar{y})^2}$.

Míry dynamiky

Jednoduché míry dynamiky časových řad umožňují charakterizovat jejich základní rysy chování. Mezi základní míry dynamiky časové řady y_t patří:

- *absolutní přírůstek (první diference)* $\Delta y_t = y_t - y_{t-1}$ a *průměrný absolutní přírůstek* $\bar{\Delta} = \frac{\sum_{t=2}^{n}\Delta y_t}{n-1} = \frac{y_n - y_1}{n-1}$;

- *koeficient (tempo) růstu* $k_t = \frac{y_t}{y_{t-1}}$, kde $t = 2, \ldots, n$, a *průměrný koeficient růstu* $\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdots k_n} = \sqrt[n-1]{y_n / y_1}$;

- *meziroční koeficient růstu* např. v případě čtvrtletní časové $k_{(4),t} = \frac{y_t}{y_{t-4}}$, kde $t = 5, 6, \ldots, n$;

- *relativní přírůstek* $\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_t} = \frac{y_t}{y_{t-1}} - 1$ a průměrný relativní přírůstek $\bar{\delta} = \bar{k} - 1$.

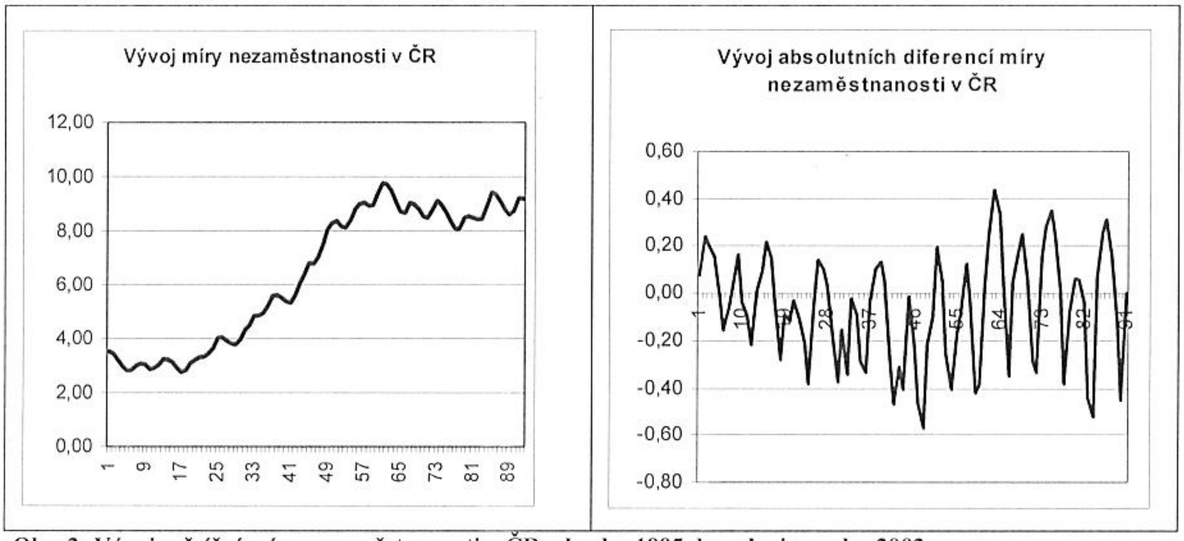
Korelace

Korelace vyjadřuje relativní míru závislosti ve vzájemném vývoji dvou časových řad např. y_t a x_t a je dána vztahem $s_{xy} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})\cdot(y_t - \bar{y})}{s_x \cdot s_y} \in \langle -1;\ 1 \rangle$. Hodnoty korelace blížící se ke hraniční hodnotě -1 vyjadřují, že obě sledované časové řady mají zcela opačný směry v jejich časovém vývoji. Hodnoty s_{xy} blížící se k 1 prozrazují, že časové řady x a y se vyvíjí téměř shodně s hlediska stejných směrů pohybů a vykazují stejnou relativní míru ve vzájemném vývoji.

Stacionární a nestacionární časová řada

Chování časové řady může ze *statistického hlediska* buď podléhat změnám v průměru či variabilitě (řada *nestacionární*), nebo být stále stejná (řada *stacionární*). Zhruba řečeno to znamená, že u stacionární řady nejsme schopni na základě zjištěných statistických parametrů, jako jsou aritmetický průměr hodnot nebo jejich rozptyl, schopni odlišit jeden úsek řady od druhého. Nestacionární řada naopak vykazuje změny v chování: například aritmetický průměr hodnot ze začátku řady je signifikantně jiný než průměr členů na konci (o takové řadě říkáme, že vykazuje *trend*).

Stacionární chování je podstatným předpokladem některých typů analýz. Je pak třeba stacionaritu testovat a řadu případně vhodným způsobem transformovat s cílem odstranění nestacionarity.



Obr. 2: Vývoj měsíční míry nezaměstnanosti v ČR od roku 1995 do poloviny roku 2002

V grafu je vyobrazen průběh typické nestacionární časové řady, vykazující rostoucí trend, sezónní vlivy v průběhu každého roku a s časem rostoucí rozptyl (sezónní odchylky od průměru se stále zvětšují. Taková řada nevykazuje žádnou časovou změnu parametrů, protože její obecný člen nezávisí ani na čase, ani na předchozích členech řady.

V literatuře se i-tý člen časové řady s charakterem nezávislých realizací normálně rozložené náhodné veličiny se střední hodnotou $\mu = 0$ a konstantním rozptylem označuje jako *bílý šum*. Taková řada je svým způsobem „nejnáhodnější" ze všech „rozumných" časových řad, protože o jejím příštím členu v podstatě nevíme na základě předchozího průběhu víc, než že půjde o „nějaké číslo kolem nuly". Název bílý šum vznikl z toho, že tato časový řada obsahuje rovnoměrný podíl frekvenčních složek všech vlnových délek podobně jako bílé světlo obsahuje složky všech barev spektra.