KGG/STG Statistika pro geografy

11. Analýza časových řad

Mgr. David Fiedor
4. května 2015

Motivace

- chceme získat představu o charakteru procesu, která časová řada reprezentuje

Jaké jevy lze znázornit a zanalyzovat časovou řadou?

- vývoj cen akcií
- objem obchodování na burze
- vývoj počtu obyvatelstva určité lokality
- maximální denní srážkové úhrny na dané stanici, průměrný roční odtok vody z povodí
- průměrné měsíční teploty vzduchu na dané stanici

Základní pojmy

Časová řada

Posloupnost věcně a prostorově srovnatelných pozorování, která jsou jednoznačně uspořádána z hlediska času ve směru minulost-přítomnost.

Analýza časových řad

Soubor metod, které slouží k popisu těchto dynamických systémů (a případně k předvídání jejich budoucího chování).

Členění časových řad

- jde o vyjádření rozdílnosti ve věcném vymezení sledovaných ukazatelů, které je mnohdy provázeno i specifickými statistickými vlastnostmi
- *podle rozhodného časového hlediska*
  - okamžikové časové řady (počet obyvatel k datu sčítání)
  - intervalové časové řady (denní úhrn srážek)
- *podle periodicity, s jakou jsou údaje v řadách sledovány*
  - dlouhodobé časové řady
  - krátkodobé časové řady
- *podle druhu sledovaných ukazatelů*
  - časové řady absolutních ukazatelů
  - časové řady odvozených charakteristik (relativní veličiny, průměry, . . .)

Problémy při sestavování časových řad

- problém volby časových bodů pozorování
- problémy s délkou časové řady
- problémy s kalendářem
- problémy s nesrovnalostí jednotlivých měření

$\Rightarrow$ narušení homogenity časové řady

Okamžikové časové řady

- sestavovány z ukazatelů, které se vztahují k určitému okamžiku – záleží na okamžiku šetření
- součet několika po sobě jdoucích hodnot nedává smysl, a proto se ani neprovádí
- typickou charakteristikou je však pro ně počítání průměrů v čase (**chronologický průměr**)

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \cdots + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{\frac{1}{2}y_1 + y_2 + \cdots + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

Okamžikové časové řady

- uvedený vzorec vyjadřuje situaci, kdy je délka mezi jednotlivými časovými okamžiky stejná
- pokud je délka různá, musíme použít vzorec pro *vážený chronologický průměr*

$$\bar{y} = \frac{\frac{y_1 + y_2}{2}(t_2 - t_1) + \cdots + \frac{y_{n-1} + y_n}{2}(t_n - t_{n-1})}{t_n - t_1}$$

Intervalové časové řady

- velikost ukazatele závisí na délce intervalu, za který je sledován
- je možno tvořit součty – dávají v tomto případě smysl
- součtová (kumulativní) řada vznikne postupným sčítáním hodnot za sebou jdoucích časových intervalů (podle průběhu součtové čáry lze posoudit rovnoměrnost vývoje hodnot znaku)
- určitým požadavkem je vztahovat údaje k časovým intervalům stejné délky
  - pokud nejsou intervaly stejně široké, je nutno pro zajištění srovnatelnosti řady přepočítat jednotlivé hodnoty
  - pro kalendářní nesrovnalosti lze například normalizovat hodnoty měsíců pro měsíc s 30 dny (všechny měsíce nemající 30 dnů vydělíme počtem dnů a vynásobíme 30)

Průměr intervalové časové řady

- spočteme jej jako prostý aritmetický průměr jednotlivých hodnot v případě, že všechny intervaly mají stejnou šířku
- vážený aritmetický průměr, pokud intervaly mají různou šířku (šířka intervalů bude fungovat jako „váha“ hodnot)

Elementární charakteristiky časové řady

- prvním úkolem při analýze časové řady je získat rychlou a orientační představu o charakteru procesu
- vizuální analýza chování ukazatele - grafy doplněné základními statistickými charakteristikami
- díky grafické reprezentaci lze rozpoznat dlouhodobou tendenci v průběhu řady či periodicky se opakující vývojové změny
- tato základní analýza není postačující k poznání hlubších souvislostí a mechanismů studovaného procesu
- neumožňuje přehledně a koncentrovaně popsat všechny vlastnosti studovaného procesu (jevu)

Odvozené časové řady

- řady sestavené z průměrů či z relativních (poměrných) hodnot
- nejedná se u nich o závislost na délce intervalu, ale na hodnotách znaků v daném intervalu
- například průměrné počty zaměstnanců místo okamžikových údajů, ...

- často nepracujeme s původní časovou řadou, nýbrž s nějakou řadou z ní odvozenou (pomocí nějaké transformace)
- nejjednoduššími charakteristikami časových řad jsou absolutní a relativní přírůstky

Odvozené ukazatele časové řady

Absolutní přírůstek (první diference) časové řady určíme jako rozdíl dvou po sobě jdoucích členů řady:

$$\Delta y_t = y_t - y_{t-1}$$

Relativním přírůstkem časové řady, který nás informuje o rychlosti (tempu) růstu, rozumíme podíl absolutního přírůstku a hodnoty jemu předcházející, tedy:

$$\delta_i = \frac{\Delta y_i}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1$$

Řetězový index

- vyjadřuje, o kolik procent vzrostla hodnota časové řady v okamžiku t_i ve srovnání s hodnotou řady v čase t_{i-1}
$$k_i = \delta_i + 1 = \frac{y_i}{y_{i-1}} \cdot 100$$
- vyjadřujeme jej v procentech

Průměrný koeficient (index) růstu

- s touto charakteristikou jsme se setkali již u základních charakteristik polohy (úrovně) v popisné statistice
- určujeme jej pro celou časovou řadu (výsledkem je jedna hodnota pro celou řadu)
- smysl má pouze v případě stálého růstu či stálého poklesu

Průměrný koeficient (index) růstu

- je nutné, aby nárůst (pokles) byl za celé zkoumané období přibližně stejné - stálý (jinak je lepší řadu rozdělit na více částí)
- průměrným koeficientem růstu myslíme geometrický průměr jednotlivých hodnot koeficientů růstu, tedy:

$$\bar{k} = \sqrt[n-1]{k_1 \cdot k_2 \cdots k_{n-1}} = \sqrt[n-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Bazický index

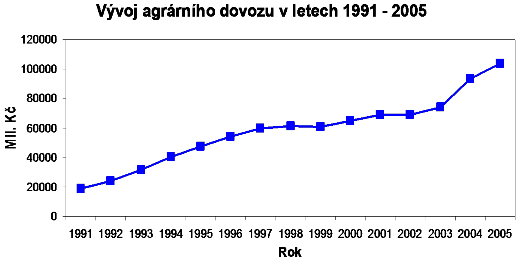
- slouží pro účely srovnání různých časových řad
- index se stejným základem, vzhledem ke kterému vyjadřujeme změny

$$k_i^{'} = \frac{y_i}{yz} \cdot 100$$

- hodnota y_z je obvykle prvním nebo posledním členem časové řady (základ)
- hodnoty bazických indexů vyjadřujeme často v procentech

Příklad

Máme k dispozici údaje o agrárním dovozu za období 1991–2005. Je potřeba charakterizovat tuto časovou řadu pomocí elementárních charakteristik.



Řešení

rok	agrární dovoz	první absolutní diference	tempo přírůstku (%)	řetězový index (%)	bazické indexy (%)
1991	18701				100,00
1992	24116	5415	28,96	128,96	128,96
1993	31497	7381	30,61	130,61	168,42
1994	40195	8698	27,62	127,62	214,94
1995	47636	7441	18,51	118,51	254,72
1996	54353	6717	14,10	114,10	290,64
1997	59677	5324	9,80	109,80	319,11
1998	61044	1367	2,29	102,29	326,42
1999	60582	-462	-0,76	99,24	323,95
2000	65012	4430	7,31	107,31	347,64
2001	69081	4069	6,26	106,26	369,40
2002	68850	-231	-0,33	99,67	368,16
2003	74029	5179	7,52	107,52	395,86
2004	93544	19515	26,36	126,36	500,21
2005	103756	10212	10,92	110,92	554,82

Transformace časových řad

Úpravy časové řady tak, aby

- splňovaly podmínky pro následnou analýzu (například stacionarita - viz. dále)
- zvýrazňovala dále analyzovanou složku

Běžné druhy transformací:

- přidání konstanty $y = y + c$
- linearizace řady $y = \ln(y)$
- odečtení průměru $y = y - \bar{y}$
- standardizace $y = \dfrac{y - \bar{y}}{s_d}$

Stacionární řada

- stacionarita řady je jedna z nutných podmínek pro mnoho metod analýzy časových řad
- časovou řadu považujeme za stacionární, jestliže splňuje následující podmínky:
  - má konstantní průměr
  - má konstantní variabilitu
  - korelace dvou časově posunutých pozorování (**autokorelace**) závisí na délce posunu
- stacionarity časové řady lze docílit transformací na řadu diferencí či odečtením trendu

Základy analýzy časových řad

Hlavní cíle analýzy časových řad:

- odhalení zákonitostí a příčin dosavadního vývoje
- prognóza chování časových řad

Každá řada může obsahovat čtyři základní složky

- trend (T_t)
- periodická (sezónní) složka (S_t)
- cyklická složka (C_t)
- náhodná složka (ϵ_t)

Trendová složka časové řady

- trend je hlavní tendence dlouhodobého vývoje hodnot analyzovaného ukazatele v čase
- může být lineární či nelineární
- rostoucí, klesající nebo může existovat řada bez trendu (s nulovým trendem) = stacionární řada

Periodická (sezónní) složka časové řady

- pravidelně se opakující odchylka od trendové složky s pevnou délkou periody T
- vyskytuje se u časových řad údajů s periodicitou kratší než jeden rok nebo rovnou právě jednomu roku
- typickým případem jsou sezónní kolísání a nebo řady denních, měsíčních, čtvrtletních ukazatelů
- příčiny sezónnosti jsou různé, většinou však dobře definovatelné
- sezónnost je typická pro časové řady ekonomických ukazatelů

Cyklická složka

- udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje
- vyskytuje se u časových řad s délkou vlny delší než jeden rok
- cyklická složka může vykazovat změny v délce a amplitudě cyklu
- délka cyklu většinou neznámá (např. demografický trend, kolísání teploty vzduchu)
- bývá typickou součástí časových řad meteorologických prvků (např. problém globálního oteplování) či hydrologických jevů

Náhodná složka časové řady

- náhodná = stochastická
- nedá se popsat žádnou funkcí času
- „zbývá“ po vyloučení trendu, sezónní a cyklické složky
- v ideálním případě lze počítat s tím, že jejím zdrojem jsou drobné, vzájemně nezávislé a v jednotlivostech nepostižitelné příčiny
- lze ji popsat pravděpodobnostně

Modely analýzy časových řad

- časová řada je funkcí času a náhodné složky

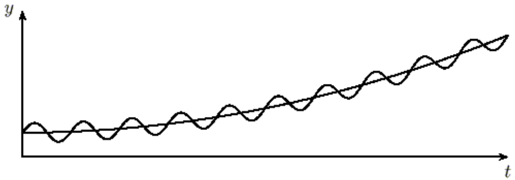
$$y_t = f(t, \epsilon_t)$$

- k analýze a popisu časových řad se používá několik základních modelů:
  1. klasický (formální) model
  2. Box-Jenkinsova metodologie
  3. Lineární dynamické a regresní modely
  4. spektrální analýza

Klasický (formální) trend

- klasický model je pouze popisem jednotlivých složek časové řady jako forem pohybu, nikoliv poznání příčin
- v podstatě se jedná o dekompozici na jednotlivé složky a jejich formální popis například aditivním modelem základem je popis systematické složky (tzn. trendu, cyklických a periodických kolísání)

$$y_t = Y_t + \epsilon_t = T_t + S_t + C_t + \epsilon_t$$



Analýza trendu

A) klasický přístup - založený na matematicko–statistickém modelování

- modelované parametry jsou konstantní v čase
- neadaptivní metody - například regresní modely, umožňují snadnou předpověď

B) adaptivní přístup - parametry se v čase vyvíjejí

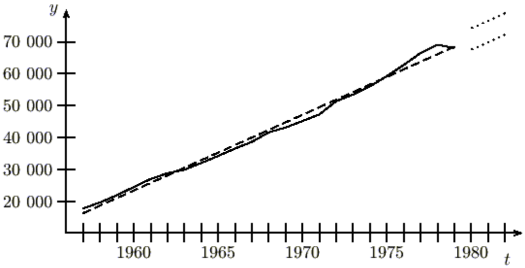
- například charakter lineárního trendu se mění (mění se směrnice trendu)
- za jednoduchou adaptivní metodu lze považovat i metodu klouzavých průměrů

Analytické vyrovnávání trendu matematickou křivkou

- patří mezi neadaptivní metody
- vychází z předpokladu, že se trend po celou sledovanou dobu nemění a že je možné ho popsat některým typem matematické křivky
- identifikace trendu se redukuje na výběr správného typu matematické křivky a odhad jejích parametrů
- na problém analýzy trendu lze pohlížet jako na speciální případ regresní závislosti, kdy nezávisle proměnnou je čas
- časovou řadu vyrovnáváme křivkou, která nejlépe vystihuje její vývojový trend a výpočet parametrů této křivky lze provést medotou nejmenších čtverců

Lineární trend

- $y_t = b_0 + b_1 t$
- parametr b_1 představuje přírůstek hodnoty y připadající na jednotkovou změnu časové proměnné
- časová řada se vyznačuje konstantními absolutními přírůstky



Lineární trend

$$y = b_0 + b_1 x$$

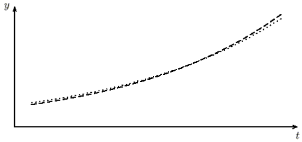
- hodnoty parametrů b_0 a b_1 získáme metodou nejmenších čtverců obdobně jako u lineární regrese
- předpověď budoucí hodnoty (*bodová předpověď*) má tvar $\hat{y}_r = b_0 + b_1 T$

Exponenciální trend

$$y_t = b_0 \cdot b_1^t$$

- parametr b_1 představuje průměrný přírůstek hodnot y_t, které se chovají jako členy geometrické posloupnosti
- již se nejedná o lineární funkci, proto lze metodu nejmenších čtverců využít pro určení parametrů až po její logaritmickou transformaci

$$\log y_t = \log b_0 + t \cdot \log b_1$$



Polynomický trend

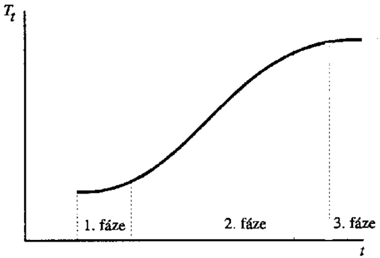
$$y_t = b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k$$

- při volbě stupně polynomu je potřeba postupovat opatrně
- vyšší stupeň zajišťuje těsnější proložení empirických hodnot křivkou, vede ale také k nestabilitě trendu
- vyšší polynomy se většinou nehodí k extrapolacím
- k odhadu parametrů lze využít metody nejmenších čtverců

Logistická křivka

$$y_t = \frac{1}{k + b_0 \cdot b_1^t}$$

- křivka se vyznačuje třemi úseky
- první úsek označuje pozvolný vzestup, druhá v okolí inflexního bodu prudkým růstem a třetí určitou vrcholovou stagnací



Děkuji za pozornost...