

Časové řady a jejich periodicitata – úvod

Jana Klicnarová

Katedra aplikované matematiky a informatiky
Jihočeská Univerzita v Českých Budějovicích, Ekonomická fakulta

2010



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Časové řady

Data, která získáváme v mnoha oblastech lidského konání a která chceme následně statisticky hodnotit, jsou mnohdy ve formě tzv. časových řad.

Časová řada

Časovou řadou rozumíme chronologicky uspořádaná pozorování hodnot nějaké náhodné veličiny. (V řeči teorie náhodných procesů je časová řada konkrétní realizací odpovídajícího náhodného procesu.)

Časové řady

Data ve tvaru časových řad získáváme v mnoha oblastech – ekonomii, biologii, lékařství, meteorologii, technice.

Příklady časových řad

Časovou řadou jsou ceny dané akcie na burze, maximální denní teploty, údaje o denní návštěvnosti nějaké historické památky, roční migrační saldo dané obce, apod.

V rámci hodnocení regionálního rozvoje se můžeme setkat s časovými řadami udávajícími např. objem produkce nějaké komodity, přirozené přírůstky obyvatelstva, procentuální nezaměstnanost, počty nehod na silnicích v daném regionu, počty absolventů, atd.

Cíl analýzy časových řad

Cílem analýzy časových řad je většinou konstrukce modelu, který nejlépe vystihuje chování zkoumané časové řady.

Model časové řady

Získaný model nám následně může umožnit pochopit mechanismus časové řady. Na základě konstrukce modelu, můžeme objevit, že řada vykazuje pravidelné cykly, trend apod.

Cíl analýzy časových řad

Využití modelu ČR

Díky takové konstrukci, můžeme odhalit, co nám ovlivňuje hodnoty časové řady a na čem naopak pravděpodobně nezávisí. Což nám může také umožnit částečně řídit (optimalizovat) chování časové řady (volbou vhodných parametrů). Můžeme také podrobněji studovat sezónní a cyklické složky nebo naopak časovou řadu od těchto složek očistit a studovat očištěnou řadu, její vývoj, její trend. Získaný model časové řady můžeme také využít k predikci budoucích hodnot této řady. Nezapomeňme, že hodnoty časových řad jsou pozorováním realizace náhodných veličin, a tedy ani sebelepší model nám není schopen určit budoucí hodnotu, je pouze schopen dát nám odhad budoucí hodnoty (ať už bodový či intervalový).

Základní značení

Jak je zvykem, budeme náhodné veličiny značit velkými písmeny, v našem případě

$$Y_t,$$

kde t bude vyjadřovat časový parametr.

Pozorované hodnoty těchto náhodných veličin (tedy jejich realizace) budeme, též podle zvyku, značit malými písmeny

$$y_t.$$

A na základě našeho modelu odhadované hodnoty (tedy bodové odhady realizací jednotlivých náhodných veličin) budeme také standardně značit

$$\hat{y}_t.$$

Základní značení

Příklad

Uvažujeme-li například časovou řadu udávající počet nastupujících prvňáčků v jednotlivých letech do dané základní školy, potom Y_{2010} vyjadřuje náhodnou veličinu – počet nastupujících prvňáčků v roce 2010. Tato náhodná veličina má nějaké rozdělení s nějakou střední hodnotou a nějakým rozptylem. My, poněvadž chceme predikovat, kolik prvňáčků do dané školy nastoupí, se snažíme toto rozdělení odhadnout, popř. odhadnout střední hodnotu této veličiny nebo rozptyl. Na základě těchto odhadů stanovíme bodový odhad počtu nastupujících prvňáčků pro rok 2010 – \hat{y}_{2010} . Po září 2010 již víme, kolik prvňáčků skutečně nastoupilo, tato hodnota se označuje – y_{2010} .



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Časový parametr

Parametr t někdy udává přímo časový úsek, ke kterému se vztahuje ($t = 2010$), někdy udává pořadí časového úseku – zajímáme-li se o počty nastupujících prvňáčků od roku 2000 dále, potom hodnoty patřící k roku 2000 mají index $t = 1$ (první v analyzované řadě), a tedy hodnoty odpovídající roku 2010 mají index $t = 11$.

Diskrétní a spojité časové řady

Časové řady rozlišujeme dvou základních typů, a to spojité a diskrétní časové řady.

Spojité časové řady

Spojité časové řady jsou řady, kde hodnoty časové řady známe a můžeme měřit v každém okamžiku (např. odběr elektřiny nějakého odběrového místa). Pro spojité časové řady nabývá parametr t hodnot z nějakého intervalu.

Diskrétní a spojité časové řady

Diskrétní časové řady

Naproti tomu **diskrétní časové řady** jsou řady, jejichž hodnoty známe jen v určitých nespojitých časových bodech (teploty o desáté hodině, přirozený přírůstek za daný rok, úhrn produkce konkrétní zemědělské plodiny za daný rok, ...). V diskrétních časových řadách nabývá parametr t diskrétních hodnot, typicky $1, 2, 3, \dots$, kde $t = 1$ je pro první pozorování, atd.

Parametr t v diskrétních časových řadách

Někdy parametr t nabývá (z důvodu lepší přehlednosti) hodnot např. $1990, 1991, \dots$, to v případě, že vyšetřujeme časovou řadu ročních pozorování od roku 1990.

Diskrétní časové řady

Parametr t v diskrétních časových řadách

V případě, že máme pozorování v nějakých kratších časových jednotkách, které se periodicky opakují v jiných delších časových jednotkách, se může používat dvojrozměrný časový parametr. Např. pokud máme čtvrtletní pozorování za posledních deset let. Pro některé analýzy je očíslováme klasicky, tak jak jdou po sobě. V některých případech, bude pro přehlednost snažší očíslovat je dvou parametry, kde první parametr bude udávat rok a druhý čtvrtletí. Tedy hodnota $y_{6,4}$ bude udávat naměřenou hodnotu v šestém roce pozorování, ve čtvrtém čtvrtletí.

Diskrétní časové řady

V dalším textu se budeme zabývat výhradně diskretním časovými řadami. Jak už jsme uvedli výše, znamená to, že pozorování vyšetřovaných časových dat budeme mít k dispozici pouze v určitých diskretních časových okamžicích. Typicky se bude jednat o roční, čtvrtletní, měsíční, týdenní či denní data.

Diskrétní časové řady

Trojí typ

Diskrétní časové řady mohou vznikat trojím způsobem. Data mohou být diskrétní svou povahou – denní produkce mléka ekofarem v regionu. V jiném případě se může jednat o diskrétní pozorování spojité časové řady – cena určitého zboží na daném trhu v danou hodinu. Další možností mohou být data, jež jsou agregací hodnot za dané časové období – počet nehod v daném regionu za dané období.

Problematika volby časových bodů pozorování

Je zřejmé, že v některých případech nemáme možnost volby časových bodů pozorování – data nám již někdo dodal, není možné (nebo by bylo nákladné) odečítat data v jiných časových bodech, atd. Naopak někdy tuto možnost máme. Otázkou tedy je, jaké časové body v takovém případě zvolit. Je to třeba pečlivě uvážit. Někdy nemá smysl volit pozorování příliš často, s příliš malým časovým odstupem – nedozvíme se téměř nic nového a jen zvýšíme náročnost výpočtů, popř. nákladnost celé analýzy. Na druhou stranu, nesmíme volit časové body příliš vzdálené, abychom neztratili, neminuli, nějakou podstatnou periodu. V každém případě se snažíme volit pozorování stále se stejným intervalem.

Práce s agregovanými daty

Jak už jsme uvedli výše, diskrétní časové řady jsou někdy tvořeny agregací hodnot za nějaké časové období – typicky pracovní týden, měsíc (nebo za pracovní dny v měsíci) apod. Zde se setkáváme s problémy s kalendářem – měsíce jsou různě dlouhé, s různým počtem pracovních dní, navíc musíme brát v úvahu pohyblivé svátky, atp.

V případě, že chceme s takovou řadou pracovat, je velmi vhodné získané hodnoty nejprve standardizovat.

Standardizace

Máme-li například počet nehod na silnicích za jednotlivé měsíce, potom je vhodné před aplikací metod časových řad, tato data standardizovat – například na měsíc o délce 30 dnů, popř. na měsíc o délce $1/12$ roku, tj. 30,4167 dní.

Z principu problému je zřejmé, jak se standardizace provádí:

$$y'_i = \frac{y_i}{k_i} \bar{k}_i,$$

kde y'_i je standardizovaná hodnota za i . období, y_i je naměřená hodnota za toto období, k_i je počet menších časových jednotek (např. dnů) v i . časovém období (např. měsíci) a \bar{k}_i je počet menších časových jednotek (např. dní) ve standardizovaném časovém období (např. měsíci).

Příklad

Porovnejte nehodovost v Královéhradeckém kraji v měsících leden a únor, máte-li následující data. (Viz. statistika Policie ČR.)

Dopravních nehod celkem	2009	2010	2011
leden	66	75	67
únor	59	82	55

Příklad

Podíváme-li se pouze na prostý součet nehod v jednotlivých měsících za poslední tři roky, popř. za průměrný počet nehod v lednu a v únoru, potom zjistíme, že více nehod se se stalo v lednu (součet nehod v lednu je 208, v únoru 196, průměrný počet nehod v lednu je 69,3 a v únoru 65,3 (pomiňme, že bychom ještě měli pro přesné závěry zjistit statistickou průkaznost)). Uvědomme si, že leden má 31 dní a únor 28. Pro porovnání nehodovosti v lednu a únoru je tedy smysluplné přepočítat průměrnou nehodovost na den. V takové případě dostaneme, že v lednu je průměrná denní nehodovost 2,237 a v únoru 2,333. A tedy najednou zjišťujeme, že bylo bezpečnější se po silnicích pohybovat v lednu než v únoru.

Dekompozice časových řad

Složky dekompozice

Princip dekompozice časových řad je velmi jednoduchý. Základní myšlenkou je představa, že časová řada obsahuje čtyři složky – **trend** (T), **sezónní složku** (S), **cyklickou složku** (C) a náhodnou složku (ε) (někdy též zvaná **reziduální** či **rezidium**).

Dekompozice časových řad – jednotlivé složky

Trend

Trend – tato složka vyjadřuje dlouhodobé změny v chování řady, typ (tvar, funkci) dlouhodobého růstu či poklesu. V případě, že z dlouhodobého hlediska nedochází k žádné změně, mluvíme o **konstantním trendu**. Funkci trendu můžeme získat například s využitím regresní analýzy.

Sezónní složka

Sezónní složka vyjadřuje periodické změny, které se odehrávají v průběhu nějakého časového období (typicky roku, popř. týdne apod.) Typicky tyto změny souvisejí se změnou ročního období.



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Cyklická složka

Cyklická složka vyjadřuje nějakou fluktuaci okolo trendu, tedy nějaké pravidelné fáze růstu a poklesu okolo složky (typickým příkladem cyklu je střídání dob ledových a meziledových). Délky cyklů bývají různé, cyklická složka se také může skládat z několika pod složek, z nichž každá má jinou délku a jinou intenzitu. Zatímco periodické složky lze očekávat (známe i její příčinu – například střídání ročního období), co se týče délky periody i vlivu na hodnoty časové řady, cyklické složky většinou objevíme až na základě analýzy časové řady a teprve zpětně se snažíme zjistit, co je její příčinou. Někdy se nám ani nepodaří příčinu nalézt.

Náhodná (reziduální) složka

Náhodná (reziduální) složka představuje náhodné fluktuace, které již nemají žádný systematický charakter. Mohou být způsobené pouze skutečností, že se jedná o pozorování náhodné veličiny nebo v sobě mohou zahrnovat chybu měření, zaokrouhlovací chybu apod.