Kapitola 10.: Úvod do analýzy časových řad

Cíl kapitoly

Po prostudování této kapitoly budete umět
- očistit časovou řadu od důsledků kalendářních variací
- graficky znázornit okamžikovou i intervalovou časovou řadu
- vypočítat popisné a dynamické charakteristiky časové řady
- odhadnout trend časové řady metodami regresní analýzy a pomocí klouzavých průměrů

Časová zátěž

Na prostudování této kapitoly a splnění úkolů s ní spojených budete potřebovat asi 8 hodin studia.

10.1. Motivace

Při analýze časových řad chceme získat představu o charakteru procesu, který tato řada reprezentuje. Průběh časové řady graficky znázorňujeme pomocí spojnicového resp. sloupkového diagramu. K jejímu popisu používáme různé charakteristiky, a to jak statické tak dynamické.

K modelování časových řad slouží celá řada metod, např. dekompoziční metoda, Boxova – Jenkinsonova metodologie, lineární dynamické modely, spektrální analýza časových řad. Zde se omezíme na speciální případ dekompoziční metody, kdy pomocí regresní analýzy a pomocí klouzavých průměrů odhadneme trend časové řady.

10.2. Základní pojmy

10.2.1. Pojem časové řady

Časovou řadou rozumíme řadu hodnot $y_{t_1}, \ldots, y_{t_n}$ určitého ukazatele uspořádanou podle přirozené časové posloupnosti $t_1 < \ldots < t_n$. Jsou-li časové intervaly $(t_1, t_2), \ldots, (t_{n-1}, t_n)$ stejně dlouhé (ekvidistantní), zjednodušeně zapisujeme časovou řadu jako $y_1, \ldots, y_n$. Přitom ukazatel je veličina, která charakterizuje nějaký sociálně ekonomický jev v určitém prostoru a v určitém čase (okamžiku či intervalu).

10.2.2. Druhy časových řad

a) Časová řada okamžiková: příslušný ukazatel udává, kolik jevů existuje v daném časovém okamžiku (např. počet obyvatelstva k určitému dnu).
b) Časová řada intervalová: příslušný ukazatel udává, kolik jevů vzniklo či zaniklo v určitém časovém intervalu (např. počet sňatků během roku). Nejsou-li jednotlivé časové intervaly ekvidistantní, musíme provést očištění časové řady od důsledků kalendářních variací.

10.2.3. Příklad: Máme k dispozici údaje o tržbě obchodní organizace (v tis. Kč) v jednotlivých měsících roku 1995: 2400, 2134, 2407, 2445, 2894, 3354, 3515, 3515, 3225, 3063, 2694, 2600. Vypočtěte očištěné údaje.

Řešení: Průměrná délka měsíce je 365/12 dne. Očištěná hodnota pro leden je tedy $y_1^{(o)} = 2400 \cdot \frac{365}{12 \cdot 31} = 2354{,}84$, pro únor $y_2^{(o)} = 2134 \cdot \frac{365}{12 \cdot 28} = 2318{,}18$. Pro ostatní měsíce analogicky dostaneme 2361,71; 2478,96; 2839,54; 3400,58, 3448,86; 3448,86; 3269,79; 3005,36; 2731,42; 2551,08.

Výpočet pomocí systému STATISTICA:
Vytvoříme nový datový soubor o třech proměnných: trzba, dm (délky jednotlivých měsíců) a ot (očištěná tržba) a 12 případech. Do proměnné trzba zapíšeme zjištěné hodnoty. Do proměnné dm vložíme délky jednotlivých měsíců, tj. 31, 28, 30, …, 31. Do Dlouhého jména proměnné ot napíšeme =trzba*365/(12*dm).

	1 trzba	2 dm	3 ot
1	2400	31	2354,839
2	2134	28	2318,185
3	2407	31	2361,707
4	2445	30	2478,958
5	2894	31	2839,543
6	3354	30	3400,583
7	3515	31	3448,858
8	3515	31	3448,858
9	3225	30	3269,792
10	3063	31	3005,363
11	2694	30	2731,417
12	2600	31	2551,075

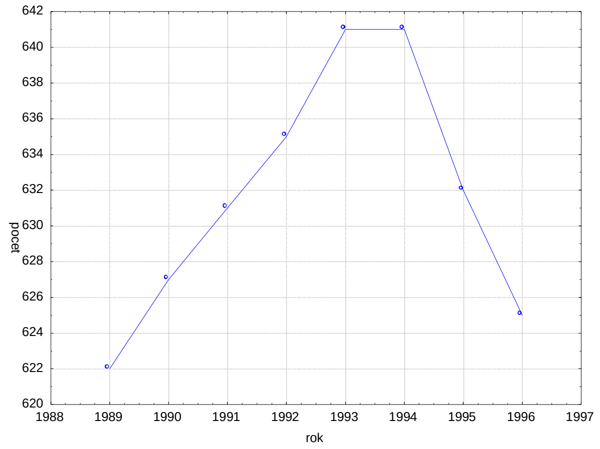
10.2.4. Grafické znázornění časové řady
a) Okamžikovou časovou řadu graficky znázorňujeme pomocí spojnicového diagramu. Na vodorovnou osu vynášíme časové okamžiky t_1, ..., t_n, na svislou osu odpovídající hodnoty y_1, ..., y_n. Dvojice bodů (t_i, y_i), i = 1, ..., n spojíme úsečkami.

10.2.5. Příklad
Časová řada obsahuje údaje o počtu zaměstnanců určité akciové společnosti v letech 1999 – 2006 vždy k 31.12.

1999	2000	2001	2002	2003	2004	2005	2006
622	627	631	635	641	641	632	625

Znázorněte tuto časovou řadu graficky.

Řešení pomocí systému STATISTICA:
Vytvoříme datový soubor o dvou proměnných nazvaných rok a pocet a 8 případech.
Grafy – Bodové grafy – odškrtneme Lineární proložení – Proměnné X – rok, Y – počet – OK – OK. 2x klikneme na pozadí grafu – vybereme Graf: obecné – zaškrtneme Spojnice – OK.



b) Intervalovou časovou řadu nejčastěji znázorňujeme sloupkovým diagramem. Je to soustava obdélníků, kde šířka obdélníku je rovna délce intervalu a výška odpovídá hodnotě ukazatele v daném intervalu. Ke znázornění intervalové časové řady lze použít i spojnicový diagram, přičemž na vodorovnou osu vynášíme středy příslušných intervalů.

10.2.6. Příklad

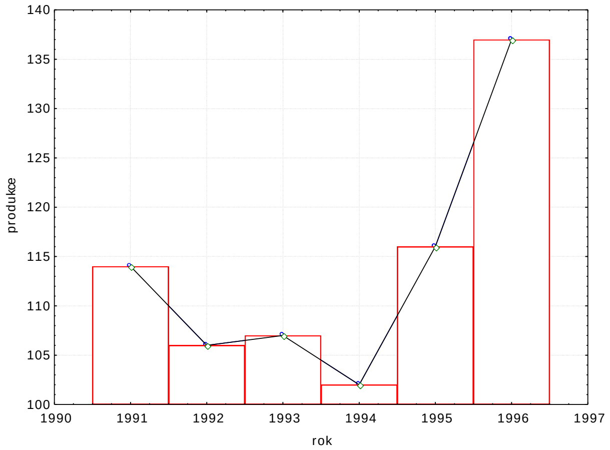
Máme k dispozici údaje o produkci určitého podniku (v tisících výrobků) v letech 1991-1996.

1991	1992	1993	1994	1995	1996
114	106	107	102	116	137

Znázorněte tuto časovou řadu graficky.

Řešení pomocí systému STATISTICA:

Vytvoříme datový soubor o dvou proměnných nazvaných rok a produkce a 6 případech. Grafy – Bodové grafy – odškrtneme Lineární proložení – Proměnné X – rok, Y – produkce – OK – OK. 2x klikneme na pozadí grafu – vybereme Graf: obecné – zaškrtneme Spojnice – Přidat nový graf – typ Sloupcový graf – OK. Do sloupců označených jako Nový1, Nový2 okopírujeme hodnoty proměnných rok a produkce. Ve Všech možnostech: Sloupce upravíme šířku sloupce na 1.



10.3. Popisné charakteristiky časových řad

10.3.1. Průměr okamžikové časové řady

Nejprve vypočteme průměry pro jednotlivé dílčí intervaly (t_1, t_2), (t_2, t_3), ..., (t_{n-1}, t_n):

$\frac{y_1 + y_2}{2}, \frac{y_2 + y_3}{2}, \ldots, \frac{y_{n-1} + y_n}{2}$. Jsou-li všechny tyto intervaly stejně dlouhé, vypočteme prostý chronologický průměr okamžikové časové řady:

$$\bar{y} = \frac{1}{n-1}\sum_{i=2}^{n}\frac{y_{i-1} + y_i}{2} = \frac{1}{n-1}\left(\frac{y_1}{2} + \sum_{i=2}^{n-1} y_i + \frac{y_n}{2}\right).$$

Nemají-li intervaly stejnou délku, vypočteme $d_i = t_i - t_{i-1}$, $i = 2, ..., n$ a použijeme vážený chronologický průměr okamžikové časové řady:

$$\bar{y} = \frac{1}{\sum_{i=2}^{n} d_i}\sum_{i=2}^{n}\frac{y_{i-1} + y_i}{2}\cdot d_i .$$

10.3.2. Příklad

Časová řada vyjadřuje počet obyvatelstva ČR (v tisících) v letech 1989 až 2008 vždy ke dni 31.12.

rok	počet obyvatel	rok	počet obyvatel
1989	10362,102	1999	10278,098
1990	10364,124	2000	10266,546
1991	10312,548	2001	10206,436
1992	10325,697	2002	10203,269
1993	10334,013	2003	10211,455
1994	10333,161	2004	10220,577
1995	10321,344	2005	10251,079
1996	10309,137	2006	10287,189
1997	10299,125	2007	10381,13
1998	10289,621	2008	10467,542

Charakterizujte tuto časovou řadu chronologickým průměrem.

Řešení: $\bar{y} = \frac{1}{19}\left(\frac{10362{,}102}{2} + 10364{,}124 + \ldots + 10381{,}13 + \frac{10467{,}542}{2}\right) = 10295{,}23.$

Výpočet pomocí systému STATISTICA:

Vytvoříme datový soubor o 21 proměnných a jednom případu. Do prvních 20 proměnných vložíme zjištěné hodnoty, do Dlouhého jména poslední proměnné napíšeme
=(v1/2+sum(v2:v19)+v20/2)/19
Dostaneme výsledek 10 295,23.

10.3.3. Průměr intervalové časové řady

Průměr intervalové časové řady počítáme podle vzorce $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

10.3.4. Příklad

Vypočtěte průměrnou hodnotu roční časové řady HDP ČR (v miliardách Kč) v letech 1995 až 2008.

1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1466,5	1683,3	1811,1	1996,5	2080,8	2189,2	2352,2	2464,4	2577,1	2814,8	2983,9	3222,4	3535,5	3689

Řešení: $\bar{y} = \frac{1}{14}(1466{,}5 + \ldots + 3689) = 2490{,}5$.

Výpočet pomocí systému STATISTICA:

Použijeme Popisné statistiky z nabídky Základní statistiky/tabulky.

10.4. Dynamické charakteristiky časových řad

10.4.1. Absolutní přírůstky

1. diference: $\Delta y_i = y_i - y_{i-1}, i = 2, \ldots, n$

2. diference: $\Delta^{(2)} y_i = \Delta y_i - \Delta y_{i-1} = y_i - 2y_{i-1} + y_{i-2}, i = 3, \ldots, n$

atd.

(Diferencování má velký význam při odhadu trendu časové řady regresními metodami.)

Průměrný absolutní přírůstek: $\bar{\Delta} = \frac{\sum_{i=2}^{n} \Delta y_i}{n-1} = \frac{y_n - y_1}{n-1}$

10.4.2. Relativní přírůstek

$\delta_i = \frac{\Delta y_i}{y_{i-1}}, i = 2, \ldots, n$

(Relativní přírůstek po vynásobení 100 udává, o kolik procent se změnila hodnota v čase t_i oproti času t_{i-1}.)

10.4.3. Koeficient růstu (tempo růstu)

$k_i = \frac{y_i}{y_{i-1}}, i = 2, \ldots, n$

(Koeficient růstu po vynásobení 100 udává, na kolik procent hodnoty v čase t_{i-1} vzrostla či poklesla hodnota v čase t_i.)

10.4.4. Průměrný koeficient růstu

$\bar{k} = \sqrt[n-1]{k_2 \cdot k_3 \cdot \ldots \cdot k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$

10.4.5. Průměrný relativní přírůstek

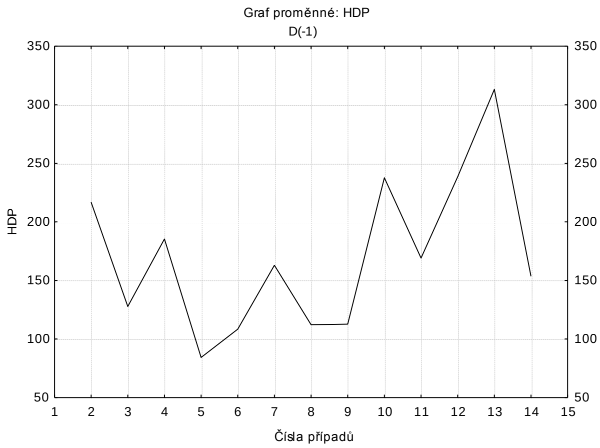
$\bar{\delta} = \bar{k} - 1$

10.4.6. Příklad

Pro časovou řadu HDP ČR v letech 1995 až 2008 (v miliardách Kč) vypočtěte základní charakteristiky dynamiky a graficky znázorněte relativní přírůstky a koeficienty růstu.

rok	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
HDP	1466,5	1683,3	1811,1	1996,5	2080,8	2189,2	2352,2	2464,4	2577,1	2814,8	2983,9	3222,4	3535,5	3689

Výpočet pomocí systému STATISTICA:
Vytvoříme nový datový soubor o 2 proměnných a 14 případech. První proměnnou nazveme ROK, druhou HDP.
Výpočet 1. diferencí: $\Delta y_i = y_i - y_{i-1}$ pro i = 2,...,n
Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné HDP – OK – OK (transformace, autokorelace, křiž. korelace, grafy) – Oddělit-sloučit - OK (transformovat vybrané řady) – vykreslí se graf.



Vrátíme se do Transformace proměnných – Uložit proměnné. Otevře se nové datové okno, kde v proměnné HDP_1 jsou uloženy 1. diference.

	HDP	HDP_1
1	1466,5	
2	1683,3	216,8
3	1811,1	127,8
4	1996,5	185,4
5	2080,8	84,3
6	2189,2	108,4
7	2352,2	163,0
8	2464,4	112,2
9	2577,1	112,7
10	2814,8	237,7
11	2983,9	169,1
12	3222,4	238,5
13	3535,5	313,1
14	3689,0	153,5

Výpočet relativních přírůstků: $\delta_i = \dfrac{\Delta y_i}{y_{i-1}}$ pro i = 2,...,n

Vrátíme se do Transformace proměnných – označíme proměnnou, kterou chceme transformovat (HDP) – vybereme Posun – OK, (Transformovat vybrané řady) – vykreslí se graf.
Vrátíme se do Transformace proměnných – Uložit proměnné. Tato transformovaná veličina se uloží do tabulky pod názvem HDP_1 (proměnná s 1. diferencemi se přejmenuje na HDP_2).
Přidáme novou proměnnou RP a do jejího Dlouhého jména napíšeme vzorec =HDP_2/HDP_1.

Výpočet koeficientů růstu: $k_i = \frac{y_i}{y_{i-1}}$ pro i = 2,...,n

Do tabulky přidáme proměnnou KR a do jejího Dlouhého jména napíšeme vzorec =HDP/HDP_1.
Získáme tabulku

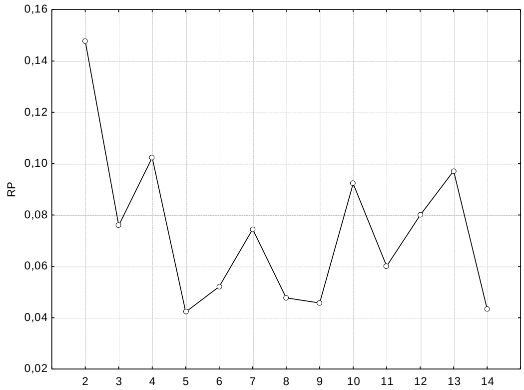
	HDP	HDP_2	HDP_1	RP	KR
1	1466,5				
2	1683,3	216,8	1466,5	0,1478	1,1478
3	1811,1	127,8	1683,3	0,0759	1,0759
4	1996,5	185,4	1811,1	0,1024	1,1024
5	2080,8	84,3	1996,5	0,0422	1,0422
6	2189,2	108,4	2080,8	0,0521	1,0521
7	2352,2	163,0	2189,2	0,0745	1,0745
8	2464,4	112,2	2352,2	0,0477	1,0477
9	2577,1	112,7	2464,4	0,0457	1,0457
10	2814,8	237,7	2577,1	0,0922	1,0922
11	2983,9	169,1	2814,8	0,0601	1,0601
12	3222,4	238,5	2983,9	0,0799	1,0799
13	3535,5	313,1	3222,4	0,0972	1,0972
14	3689,0	153,5	3535,5	0,0434	1,0434
15			3689,0		

Průměrný absolutní přírůstek: $\overline{\Delta} = \frac{3689 - 1466,5}{13} = 170,96$, tzn., že v období 1995–2008 rostl HDP průměrně o 170,96 miliard Kč ročně.

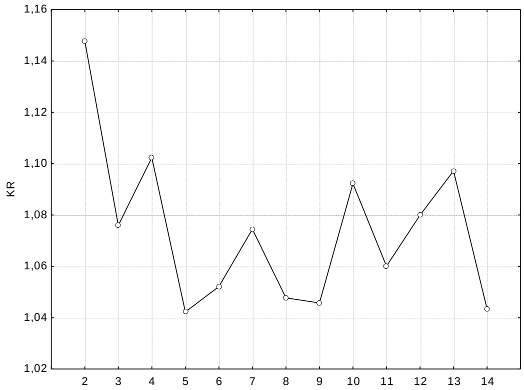
Průměrný koeficient růstu: $\bar{k} = \sqrt[13]{\frac{3689}{1466,5}} = 1,0735$, tzn., že v období 1995–2008 rostl HDP průměrně o 7,35 % ročně.

Pomocí Grafy - 2D Grafy – Spojnicové grafy (Proměnné) vykreslíme průběh relativních přírůstků a koeficientů růstu.

Graf relativních přírůstků



Graf koeficientů růstu



10.5. Aditivní model časové řady

10.5.1. Popis modelu

Předpokládejme, že pro časovou řadu $y_1, ..., y_n$ platí model

$y_t = f(t) + \varepsilon_t$, $t = 1, ..., n$, kde

f(t) je neznámá trendová funkce (trend), kterou považujeme za systematickou (deterministickou) složku časové řady (popisuje hlavní tendenci dlouhodobého vývoje časové řady),
ε_t je náhodná složka časové řady zahrnující odchylky od trendu. Náhodná složka splňuje předpoklady
$E(\varepsilon_t) = 0$,
$D(\varepsilon_t) = \sigma^2$,
$C(\varepsilon_t, \varepsilon_{t+h}) = 0$,
$\varepsilon_t \sim N(0, \sigma^2)$ (říkáme, že ε_t je bílý šum).

10.5.2. Cíl regresní analýzy trendu

Regresní analýza trendu má objasnit vztah mezi závisle proměnnou veličinou Y a časem t. Předpokládáme, že trend f(t) závisí (lineárně či nelineárně) na neznámých parametrech $\beta_0, \beta_1, ..., \beta_k$ a známých funkcích $\varphi_0(t), \varphi_1(t), ..., \varphi_k(t)$, které již neobsahují žádné neznámé parametry, tj. $f(t) = g(t;\beta_0, \beta_1, ..., \beta_k)$. Odhady $b_0, b_1, ..., b_k$ neznámých parametrů $\beta_0, \beta_1, ..., \beta_k$ lze získat např. metodou nejmenších čtverců a pak vyjádřit odhad $\hat{f}(t)$ neznámého trendu v bodě t pomocí odhadů $b_0, b_1, ..., b_k$ a funkcí $\varphi_0(t), \varphi_1(t), ..., \varphi_k(t)$, tj.
$\hat{f}(t) = g(t;b_0, b_1, ..., b_k)$.

10.5.3. Nejdůležitější typy trendových funkcí

Volba typu trendové funkce se provádí
- na základě teoretických znalostí a zkušeností se zkoumanou veličinou Y_t
- pomocí grafu časové řady
- pomocí informativních testů založených na jednoduchých charakteristikách časové řady

a) Lineární trend
   Analytické vyjádření: $f(t) = \beta_0 + \beta_1 t$
   Informativní test: 1. diference jsou přibližně konstantní.

b) Kvadratický trend

Analytické vyjádření: $f(t) = \beta_0 + \beta_1 t + \beta_2 t^2$

Informativní test: 1. diference mají přibližně lineární trend, 2. diference jsou přibližně konstantní.

c) Exponenciální trend

Analytické vyjádření: $f(t) = \beta_0 \beta_1{}^t$.

Model lze linearizovat logaritmickou transformací: $\ln f(t) = \ln \beta_0 + t \ln \beta_1$

Informativní test: koeficienty růstu jsou přibližně konstantní.

d) Modifikovaný exponenciální trend

Analytické vyjádření: $f(t) = \alpha + \beta_0 \beta_1{}^t$.

Informativní test: řada podílů sousedních 1. diferencí je přibližně konstantní.

e) Logistický trend

Analytické vyjádření: $f(t) = \dfrac{\alpha}{1 + \beta_0 \beta_1{}^t}$

Informativní test: průběh 1. diferencí je podobný Gaussově křivce a podíly $\dfrac{1/y_{t+2} - 1/y_{t+1}}{1/y_{t+1} - 1/y_t}$ jsou přibližně konstantní.

f) Gompertzova křivka

Analytické vyjádření: $f(t) = \alpha \beta_0{}^{\beta_1{}^t}$

Informativní test: podíly $\dfrac{\ln y_{t+2} - \ln y_{t+1}}{\ln y_{t+1} - \ln y_t}$ jsou přibližně konstantní.

Modely (a), (b), (c) jsou lineární nebo se dají linearizovat a odhady parametrů získáme metodou nejmenších čtverců. Modely (d), (e), (f) jsou nelineární a odhady parametrů se získávají speciálními numerickými metodami.

10.5.4. Orientační ověřování kvality modelu

- Index determinace (tj. podíl vysvětlené a celkové variability závisle proměnné veličiny) by měl být blízký 1.
- Body grafu $\left(y_t, \hat{f}(t)\right)$, t = 1, 2, ..., n by se měly řadit do přímky se směrnicí 1.
- Při srovnání několika modelů se stejným počtem parametrů volíme ten model, pro který je střední kvadratická chyba odhadu ($MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{f}(t)\right)^2$) nejnižší.

10.5.5. Příklad

Uvažme časovou řadu HDP ČR v letech 1995 až 2008 (v miliardách Kč) – viz př. 10.4.6.

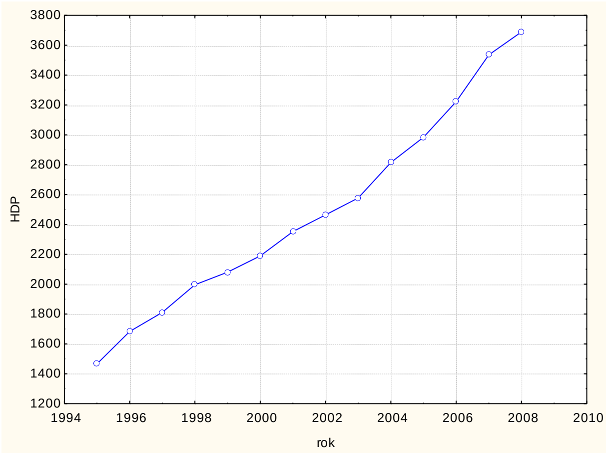
a) Graficky znázorněte průběh této časové řady.

b) Z grafu časové řady lze usoudit, že časová řada má lineární trend $f(t) = \beta_0 + \beta_1 t$. Odhadněte jeho parametry a nakreslete průběh trendu do grafu časové řady.

c) Zjistěte odhad HDP v roce 2009.

d) Vypočtěte index determinace a sestrojte graf $\left(y_t, \hat{f}(t)\right)$, t = 1, ..., 14.

Řešení pomocí systému STATISTICA:
Vytvoříme datový soubor se třemi proměnnými rok, t, HDP a 14 případy. Do proměnné t uložíme hodnoty 1, …, 14.
ad a) Graficky znázorníme průběh této časové řady:
Grafy – Bodové grafy – Proměnné rok, HDP – OK – vypneme proložení – OK.

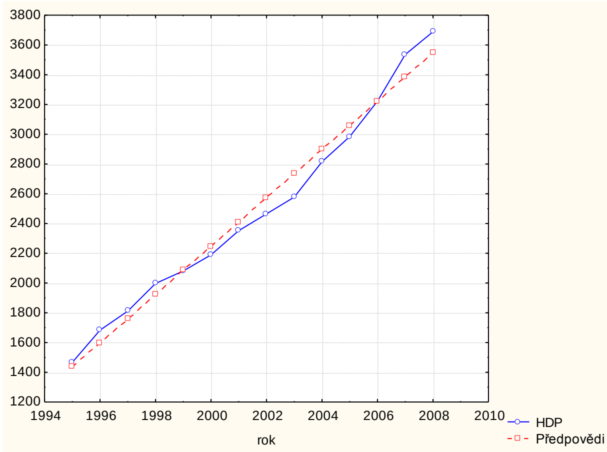


ad b) Odhadneme parametry lineárního trendu.
Statistika – Vícerozměrná regrese – Proměnné závislé: HDP, nezávislé: t – OK – OK. Otevře se nové okno Výsledky – vícenásobná regrese. Na záložce Základní výsledky zvolíme Výpočet: výsledky regrese a získáme tabulku, kde ve sloupci B jsou odhady regresních parametrů.

	Výsledky regrese se závislou proměnnou : HDP (HDP_CR.sta R= ,99052776 R2= ,98114525 Upravené R2= ,97957402 F(1,12)=624,44 p<,00000 Směrod. chyba odhadu : 97,936					
N=14	b*	Sm.chyba z b*	b	Sm.chyba z b	t(12)	p-hodn.
Abs.člen			1273,56	55,2867	23,0356	0,00000
t	0,99052	0,03963	162,25	6,4930	24,9888	0,00000

Odhad trendu: $\hat{f}(t) = 1273{,}565 + 162{,}255t$.
Podle tohoto modelu by tedy HDP v roce 1994 činil 1 273, 565 mil. Kč (realita byla 1 255, 986) a v každém dalším roce by vzrostl o 162, 255 mil. Kč.

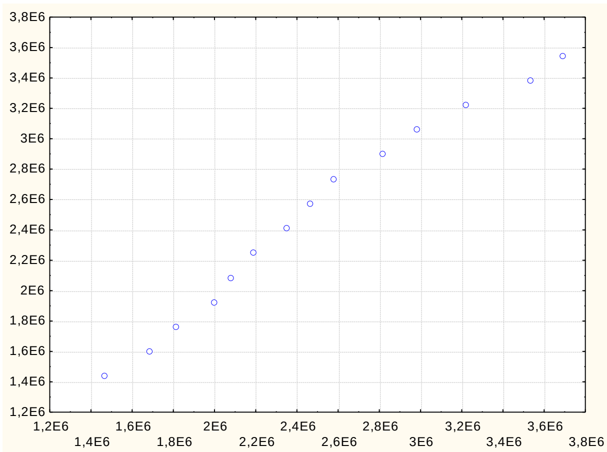
Vytvoření grafu časové řady s odhadnutým trendem:
Na záložce Uložit zvolíme Uložit rezidua a předpovědi. K nim do tabulky uložíme ještě proměnné rok a HDP a pomocí vícenásobného bodového grafu vytvoříme požadovaný graf.



ad c) Odhad HDP roce 2009: Pro výpočet predikované hodnoty zvolíme Rezidua/předpoklady/předpovědi - Předpovědi závisle proměnné čas: 15 - OK. Ve výstupní tabulce je hledaná hodnota označena jako Předpověď: 3 707,392.

	Předpovězené hodnoty (HDP_CR.s proměnné: HDP		
Proměnná	b-váha	Hodnota	b-váha * Hodnot
t	162,255	15,0000	2433,82
Abs. člen			1273,56
Předpověď			3707,39
-95,0%LS			3586,93
+95,0%LS			3827,85

ad d) Index determinace je $ID^2 = 0,981$, jak je uvedeno v záhlaví výstupní tabulky regresní analýzy. Znamená to, že lineární trend vysvětluje variabilitu HDP z 98,1%.
Graf závislosti predikovaných hodnot na hodnotách časové řady vytvoříme tak, že uložíme předpovězené hodnoty. Pak pomocí Bodového grafu vykreslíme závislost predikce na Y.



Jak index determinace, tak graf $\left(y_t, \hat{f}(t)\right)$ svědčí o tom, že model dobře vystihuje charakter dané časové řady.

10.6. Odhad trendu časové řady pomocí klouzavých průměrů

10.6.1. Podstata klouzavých průměrů

Předpokládáme, že časová řada se řídí aditivním modelem
$y_t = f(t) + \varepsilon_t$, t = 1, ..., n. Odhad trendu v bodě t získáme určitým zprůměrováním původních pozorování z jistého okolí uvažovaného časového okamžiku t. Můžeme si představit, že podél dané časové řady klouže okénko, v jehož rámci se průměruje. Nechť toto okénko zahrnuje d členů nalevo od bodu t a d členů napravo od bodu t. Hovoříme pak o vyhlazovacím okénku šířky h = 2d + 1. Prvních a posledních d hodnot trendu neodhadujeme, protože pro $t \in \{1,\ldots,d\} \cup \{n-d+1,\ldots,n\}$ není vyhlazovací okénko symetrické. Odhad trendu ve středu vyhlazovacího okénka je dán vztahem:

$$\hat{f}(t) = \frac{1}{2d+1}\left(y_{t-d} + y_{t-d+1} + \ldots + y_{t+d}\right) = \frac{1}{2d+1}\sum_{k=0}^{2d} y_{t-d+k}, \ t = d+1, ..., n-d.$$

10.6.2. Šířka vyhlazovacího okénka

Velmi důležitou otázkou je stanovení šířky vyhlazovacího okénka. Je-li okénko příliš široké, bude se odhad trendu blížit přímce (říkáme, že je přehlazen) a zároveň se ztratí velký počet členů na začátku a na konci časové řady. Je-li naopak okénko úzké, bude se odhad trendu blížit původním hodnotám (říkáme, že odhad je podhlazen). Nejčastěji se volí šířka okénka h = 3, 5, 7, pro časovou řadu čtvrtletních hodnot 4.

10.6.3. Příklad

Máme k dispozici čtvrtletní časovou řadu průměrných měsíčních mezd v České republice v době od 1/2001 do 3/2009:

čas	mzda	čas	mzda	čas	mzda
1/2002	14204	4/2004	19980	3/2007	21470
2/2002	15772	1/2005	17678	4/2007	23435
3/2002	15422	2/2005	18763	1/2008	22531
4/2002	17315	3/2005	18833	2/2008	23182
1/2003	15407	4/2005	20841	3/2008	23144
2/2003	17084	1/2006	18903	4/2008	25381
3/2003	16522	2/2006	20036	1/2009	22328
4/2003	18697	3/2006	19968	2/2009	22992
1/2004	16722	4/2006	21952	3/2009	23350
2/2004	17817	1/2007	20399		
3/2004	17738	2/2007	21462		

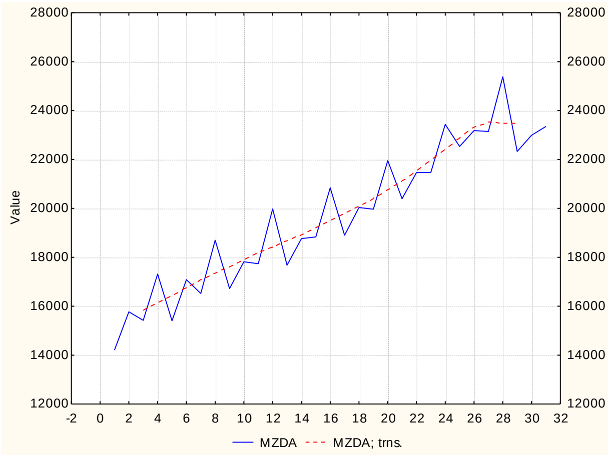
a) Odhadněte trend této časové řady pomocí klouzavých průměrů s vyhlazovacím okénkem šířky 4.
b) Graficky znázorněte průběh časové řady s odhadnutým trendem.

Řešení pomocí systému STATISTICA:
Vytvoříme datový soubor ctvrtletni_mzda.sta o dvou proměnných CAS a MZDA a 31 případech.
Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné Y – OK– OK (transformace, autokorelace, kříž. korelace, grafy) – Vyhlazování – zaškrtneme N-

bod. klouzavý průměr, N = 4 – OK (Transformovat vybrané řady) – vykreslí se graf, vrátíme se do Transformace proměnných – Uložit proměnné. Otevře se nová datová tabulka, kde v proměnné MZDA_1 jsou uloženy klouzavé průměry pro N = 4. K datovému souboru přidáme proměnnou CAS, kterou okopírujeme z původního datového souboru.

	ctvrtletni_mzda.sta		
	1 CAS	2 MZDA	3 MZDA_1
1	1/2002	14204,0	
2	2/2002	15772,0	
3	3/2002	15422,0	15828,6
4	4/2002	17315,0	16143,0
5	1/2003	15407,0	16444,5
6	2/2003	17084,0	16754,7
7	3/2003	16522,0	17091,8
8	4/2003	18697,0	17347,8
9	1/2004	16722,0	17591,5
10	2/2004	17817,0	17903,8
11	3/2004	17738,0	18183,7
12	4/2004	19980,0	18421,5
13	1/2005	17678,0	18676,6
14	2/2005	18763,0	18921,1
15	3/2005	18833,0	19181,8
16	4/2005	20841,0	19494,1
17	1/2006	18903,0	19795,1
18	2/2006	20036,0	20075,8
19	3/2006	19968,0	20401,7
20	4/2006	21952,0	20767,0
21	1/2007	20399,0	21133,0
22	2/2007	21462,0	21506,1
23	3/2007	21470,0	21958,0
24	4/2007	23435,0	22439,5
25	1/2008	22531,0	22863,7
26	2/2008	23182,0	23316,2
27	3/2008	23144,0	23534,1
28	4/2008	25381,0	23485,0
29	1/2009	22328,0	23487,0
30	2/2009	22992,0	
31	3/2009	23350,0	

Pro zobrazení proměnných MZDA a MZDA_1 do jednoho grafu přejdeme na záložku Přehledy a grafy. Vedle možnosti Zobrazit víc proměnných zvolíme Graf. Označíme obě proměnné – OK. Vykreslí se následující graf:



Vidíme, že díky vhodné volbě šířky vyhlazovacího okénka se podařilo odhadnout trend dané časové řady.

Shrnutí

Časovou řadou rozumíme řadu číselných hodnot určitého ukazatele, který se v čase mění. Rozlišujeme *časové řady okamžikové* (příslušný ukazatel udává, kolik jevů existuje v daném časovém okamžiku) a *časové řady intervalové* (příslušný ukazatel udává, kolik jevů vzniklo či zaniklo v určitém časovém intervalu). Nejsou-li jednotlivé časové intervaly stejně dlouhé, musíme provést *očištění časové řady od důsledků kalendářních variací*. Okamžikové časové řady znázorňujeme pomocí *spojnicového diagramu*, intervalové pak pomocí *sloupkového diagramu*.

Okamžikovou časovou řadu charakterizujeme *chronologickým průměrem*, intervalovou *aritmetickým průměrem*.

K popisu časových řad používáme také dynamické charakteristiky: *absolutní* a *relativní přírůstky* a *koeficienty růstu*.

K nejdůležitějším úkolům analýzy časových řad patří odhad trendu, tj. deterministické složky časové řady, která vystihuje popisuje hlavní tendenci dlouhodobého vývoje časové řady. Odhad trendu lze provádět např. metodami regresní analýzy nebo pomocí klouzavých průměrů.

Regresní odhad trendu vyžaduje, aby vývoj časové řady odpovídal nějaké funkci. Její parametry pak odhadujeme metodou nejmenších čtverců (to v případě lineárních či linerizovatelných modelů) nebo vhodnou numerickou metodou (např. Levenbergovou – Marquardtovou). Naproti tomu *metoda klouzavých průměrů*, která patří k tzv. adaptivním metodám, je založena na předpokladu, že časová řada mění v čase svůj charakter, tudíž trend nelze popsat pomocí jediné funkce. Předpokládáme však, že v krátkých úsecích svůj charakter zachovává a tedy každému úseku lze jednu funkci přiřadit. V případě výše popsaných klouzavých průměrů jde o konstantní funkci.

Kontrolní otázky

1. Jak se liší časová řada okamžiková od intervalové?
2. Kdy se používá prostý a kdy vážený chronologický průměr?
3. Jak je definována druhá absolutní diference?
4. Uveďte vzorec pro výpočet průměrného koeficientu růstu.

5. Má-li časová řada kvadratický trend, jak se chovají její první diference?
6. Co to znamená, když náhodná složka časové řady je bílým šumem?
7. Popište princip metody klouzavých průměrů.

Autokorekční test

1. Jaké číslo patří v následující tabulce místo otazníku?

y_t	156	175
k_t	0,975	?

a) 0,891
b) 1,094
c) 1,122

2. Jaké číslo patří v následující tabulce místo otazníku?

y_t	25	32
δ_t	xxx	?

a) 0,219
b) 0,280
c) 0,781

3. Jaký je průměrný absolutní přírůstek za celou dobu sledování?

t	1	…	21
y_t	185	…	249

a) 1,015
b) 3,048
c) 3,200

4. Pokud mají 1. diference časové řady přibližně lineární trend, pak vhodným modelem trendové funkce časové řady je
a) Gomperzova křivka
b) parabola
c) exponenciála

5. Střední kvadratickou chybu odhadu trendu nelze počítat podle vzorce:

a) $MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{f}(t)\right)^2$

b) $MSE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{f}(t)\right|^2$

c) $MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{f}(t)\right)\left|y_t - \hat{f}(t)\right|$

Správné odpovědi: 1c), 2b), 3c), 4b) 5c)

Příklady

1. V jednotlivých čtvrtletích roku 2006 se v ČR uskutečnilo 4 896, 16 545, 23 368 a 8051 sňatků. Vypočtěte očištěné údaje.

Výsledek:
4 964; 16 590,45; 23 177,5; 7985,37

2. V následující tabulce jsou uvedeny údaje o počtu nezaměstnaných (v tisících) v ČR v letech 2001 – 2008. Vypočtěte chronologický průměr.

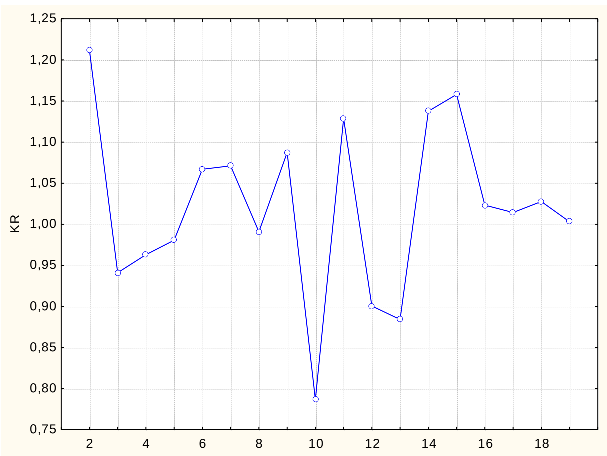
rok	2001	2002	2003	2004	2005	2006	2007	2008
počet	418,3	374,1	399,1	425,9	410,2	371,3	244,5	223,9

Výsledek:
363,7

3. Pro časovou řadu z let 1989 – 2007 spotřeby cigaret na jednoho obyvatele ČR za rok graficky znázorněte průběh koeficientů růstu a vypočtěte a interpretujte průměrný relativní přírůstek.

rok	počet	rok	počet
1989	1776	1999	2090
1990	2152	2000	1882
1991	2025	2001	1664
1992	1950	2002	1893
1993	1912	2003	2192
1994	2040	2004	2243
1995	2185	2005	2275
1996	2165	2006	2338
1997	2354	2007	2345
1998	1852		

Výsledek:
Graf koeficientů růstu



Průměrný relativní přírůstek: 0,0156, tzn., že v letech 1989 – 2007 rostla spotřeba cigaret na jednoho obyvatele za rok v průměru o 1,56 %.

4. Časová řada 112, 149, 238, 354, 580, 867 udává zisk (v tisících dolarů) jisté společnosti v prvních šesti letech její existence.

a) Z grafu časové řady a chování koeficientů růstu lze usoudit, že časová řada má exponenciální trend $f(t) = \beta_0 \beta_1^{\ t}$. Odhadněte jeho parametry.

b) Najděte odhad zisku společnosti v 7. a 8. roce její existence.

e) Vypočtěte index determinace.

Výsledek:

ad a) Model $f(t) = \beta_0 \beta_1^{\ t}$ linearizujeme na model $\ln f(t) = \ln \beta_0 + t \cdot \ln \beta_1$ a metodou nejmenších čtverců získáme odhady ln b_0, ln b_1. Odlogaritmováním dostaneme b_0 = 68,57875, b_1 = 1,522265.

ad b) Odhad zisku společnosti v 7. roce existence: 1299,035 tisíc dolarů, v 8. roce : 1977,476 tisíc dolarů.

ad c) Index determinace je 0,996.

5. Časová řada 215, 219, 222, 235, 202, 207, 187, 204, 174, 172, 201, 272 udává roční objemy vývozu piva (v miliónech litrů) z Československa v letech 1980 až 1991. Odhadněte trend této časové řady pomocí klouzavých průměrů s vyhlazovacím okénkem šířky 5

Výsledek:

rok	1982	1983	1984	1985	1986	1987	1988	1989
kp5	218,6	217	210,6	207	194,8	188,8	187,6	204,6