Úvod do analýzy časových řad

Osnova:

- pojem časové řady
- druhy časových řad a jejich grafické znázornění
- statické a dynamické charakteristiky časové řady
- aditivní model časové řady
- odhad trendu časové řady pomocí klouzavých průměrů
- regresní analýza trendu

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 jev v určitém prostoru a určitém čase (okamžiku či intervalu).

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.

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 $y_1^{(o)} = 2400 \cdot \frac{\frac{365}{12}}{31} = 2354{,}84$,

pro únor $y_2^{(o)} = 2134 \cdot \frac{\frac{365}{12}}{28} = 2318{,}18$.

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	240	3	2354,
2	213	28	2318,
3	240	3	2361,
4	244	30	2478,
5	289	3	2839,
6	335	30	3400,
7	351	3	3448,
8	351	3	3448,
9	322	30	3269,
10	306	3	3005,
11	269	30	2731,
12	260	3	2551,

Grafické znázornění okamžikové časové řady

Použijeme spojnicový diagram. 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.

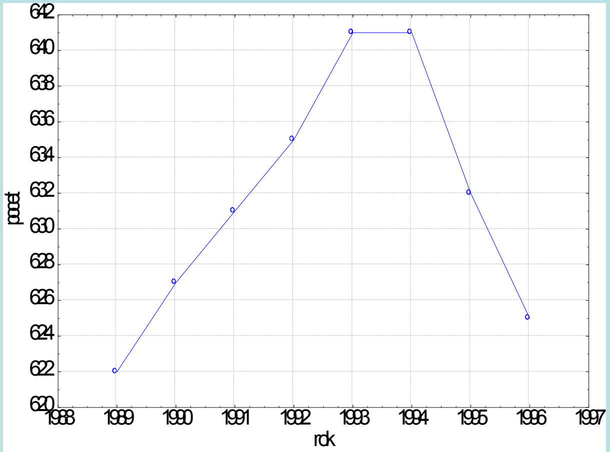
Příklad: Časová řada obsahuje údaje o počtu zaměstnanců určité akciové společnosti v letech 1989 – 1996 vždy k 31.12.

1989	1990	1991	1992	1993	1994	1995	1996
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.



Grafické znázornění intervalové časové řady

Použijeme sloupkový diagram. 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.

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

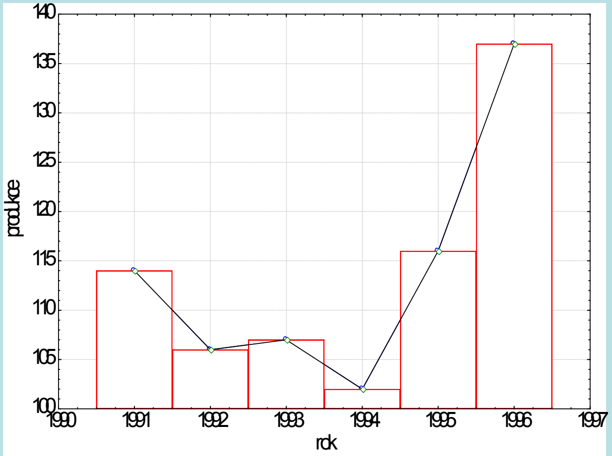
1991	1992	1993	1994	1995	1996
114	106	107	102	116	137

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.



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 $d_i = t_i - t_{i-1}$, $i = 2, \ldots, n$ a použijeme vážený chronologický průměr okamžikové časové řady:

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

Příklad: Časová řada vyjadřuje počet obyvatelstva ČSSR (v tisících) v letech 1965 až 1974 vždy ke dni 31.12.

Rok	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
počet	14194	14271	14333	14387	14443	14345	14419	14576	14631	14738

Charakterizujte tuto časovou řadu chronologickým průměrem.

Řešení: $\bar{y} = \frac{1}{9}\left(\frac{14194}{2} + 14271 + \cdots + 14631 + \frac{14738}{2}\right) = 1443\ldots$

Průměr intervalové časové řady

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

Příklad: Vypočtěte průměrnou hodnotu roční časové řady HDP ČR (v miliardách Kč) v letech 1994 až 2000.

1994	1995	1996	1997	1998	1999	2000
1303,6	1381,1	1447,7	1432,8	1401,3	1390,6	1433,8

Řešení: $\bar{y} = \frac{1}{7}(1303,6 + \dots + 1433,8) = 1398,7.$

Dynamické charakteristiky časových řad

Absolutní přírůstky

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

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

atd.

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

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

Relativní přírůstek

$\delta_i = \dfrac{\Delta_i}{y_{i-1}},\ i = 2, \ldots, n$

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

Koeficient růstu (tempo růstu)

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

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

Průměrný koeficient růstu

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

Průměrný relativní přírůstek

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

Příklad: Pro časovou řadu HDP ČR v letech 1994 až 2000 (v miliardách Kč) vypočtěte základní charakteristiky dynamiky a graficky znázorněte 1. diference a koeficienty růstu.

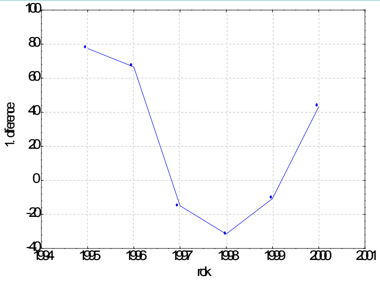
Řešení:

rok	HDP	Δy_i	k_i	δ_i
1994	1303,6	x	x	x
1995	1381,1	77,5	1,059	0,059
1996	1447,7	66,6	1,048	0,048
1997	1432,8	-14,7	0,990	-0,010
1998	1401,3	-31,5	0,978	-0,022
1999	1390,6	-10,7	0,992	-0,008
2000	1433,8	43,2	1,031	0,031

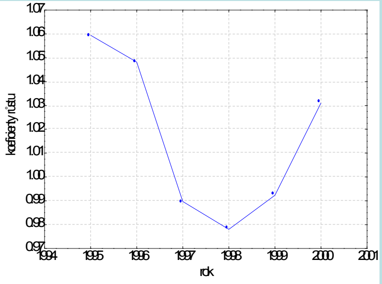
Průměrný absolutní přírůstek: $\bar{\Delta} = \frac{1433,8 - 1303,6}{6} = 21,7$, tzn., že v období 1994 – 2000 rostl HDP průměrně o 21,7 miliard Kč ročně.

Průměrný koeficient růstu: $\bar{k} = \sqrt[6]{\frac{1433,8}{1303,6}} = 1,016$, tzn., že v období 1994 – 2000 rostl HDP průměrně o 1,6% ročně.

Graf 1. diferencí:

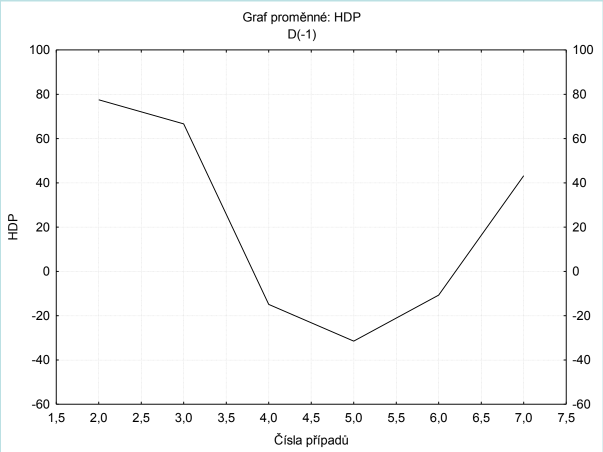


Graf koeficientů růstu:



Výpočet pomocí systému STATISTICA

Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné HDP – OK – OK (transformace, autokorelace, kříž. korelace, grafy) – Diferencování - 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	1303,600	
2	1381,100	77,500
3	1447,700	66,600
4	1432,800	-14,900
5	1401,300	-31,500
6	1390,600	-10,700
7	1433,800	43,200

Výpočet relativních přírůstků: $\delta_i = \frac{\Delta_i}{y_{i-1}}$ 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

	1 HDP	2 HDP	3 HDP	4 RP	5 KR
1	1303,				
2	1381,	77,5	1303,	0,059	1,059
3	1447,	66,6	1381,	0,048	1,048
4	1432,	-14,9	1447,	-0,01	0,989
5	1401,	-31,5	1432,	-0,02	0,978
6	1390,	-10,7	1401,	-0,00	0,992
7	1433,	43,2	1390,	0,031	1,031
8			1433,		

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

Průměrný absolutní přírůstek a průměrný koeficient růstu vypočteme na kalkulačce pomocí vzorců

$\bar{\Delta} = \frac{1433,8 - 1303,6}{6} = 21,7$ a $\bar{k} = \sqrt[6]{\frac{1433,8}{1303,6}} = 1,01.$

Aditivní model časové řady

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).

Odhad trendu časové řady pomocí klouzavých průměrů

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 $t \in \{1, \ldots, d, n-d+1, \ldots, n\}$ není vyhlazovací okénko symetrické. Odhad trendu ve středu vyhlazovacího okénka je dán vztahem:

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

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 čtvrtletní hodnoty pak 4.

Příklad: 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.

a) Odhadněte trend této časové řady pomocí klouzavých průměrů s vyhlazovacím okénkem šířky 3 a poté 5.

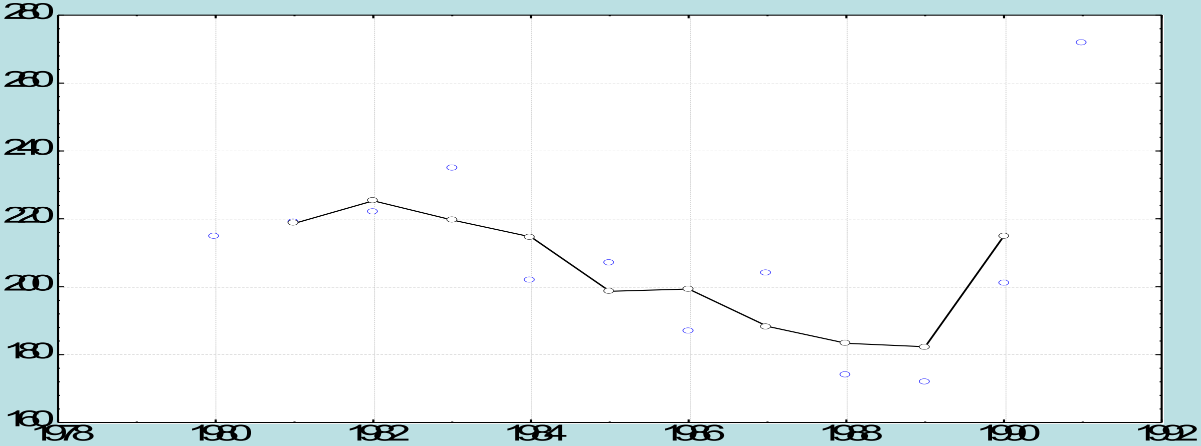
b) Graficky znázorněte průběh časové řady s odhadnutým trendem.

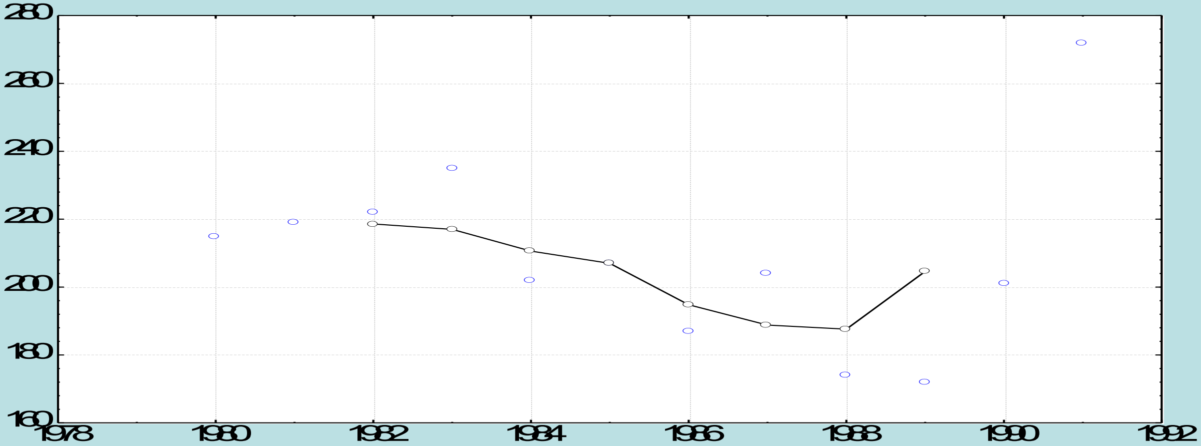
Řešení pomocí systému STATISTICA:

Vytvoříme datový soubor export_piva.sta o dvou proměnných ROK a VYVOZ a dvanácti 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 = 3 – OK (Transformovat vybrané řady) – vykreslí se graf, vrátíme se do Transformace proměnných – Uložit proměnné. Otevře se nový spreadsheet, kde v proměnné VYVOZ_1 jsou uloženy klouzavé průměry pro N = 3. Totéž uděláme pro případ N = 5. Ve spreadsheetu se proměnná VYVOZ_1 přepíše na VYVOZ_2 a nová proměnná se uloží jako VYVOZ_1. Nově vzniklé proměnné nazveme KP3 a KP5. K datovému souboru přidáme proměnnou ROK, do jejíhož Dlouhého jména napíšeme =1979+v0.

	export_piva.sta			
	1 rok	2 VYVC	3 KP3	4 KP5
1	198	215,		
2	198	219,	218,	
3	198	222,	225,	218,
4	198	235,	219,	217,
5	198	202,	214,	210,
6	198	207,	198,	207,
7	198	187,	199,	194,
8	198	204,	188,	188,
9	198	174,	183,	187,
1C	198	172,	182,	204,
11	199	201,	215,	
12	199	272,		

Grafické znázornění časové řady s odhadnutým trendem provedeme pomocí vícenásobných bodových grafů.





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.

$f(t) = g(\beta_0, \beta_1, ..., \beta_k; \varphi_0(t), \varphi_1(t), ..., \varphi_k(t))$.

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.

$\hat{f}(t) = g(b_0, b_1, ..., b_k; \varphi_0(t), \varphi_1(t), ..., \varphi_k(t))$.

Nejdůležitější typy trendových funkcí

Volba typu trendové funkce se provádí

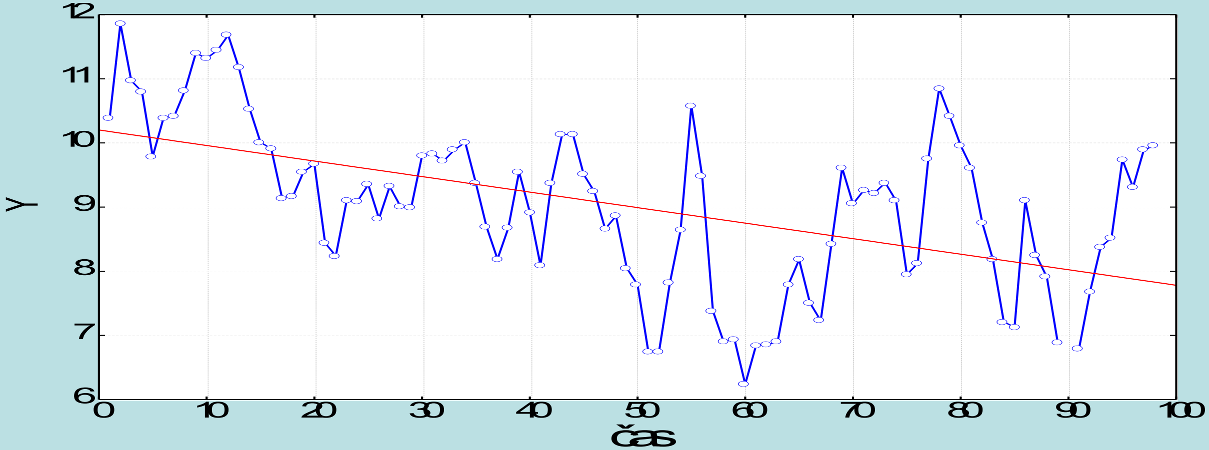
- na základě teoretických znalostí a zkušeností se zkoumanou veličinou Y_t
- pomocí grafu časové řady
- pomocí informativních testů založených na jednoduchých charakteristikách časové řady

a) Lineární trend

Analytické vyjádření: $t(t) = \ldots$

Informativní test: 1. diference ($\Delta = \ldots,\ t = \ldots,n$) jsou přibližně konstantní.

Příklad lineárního trendu:

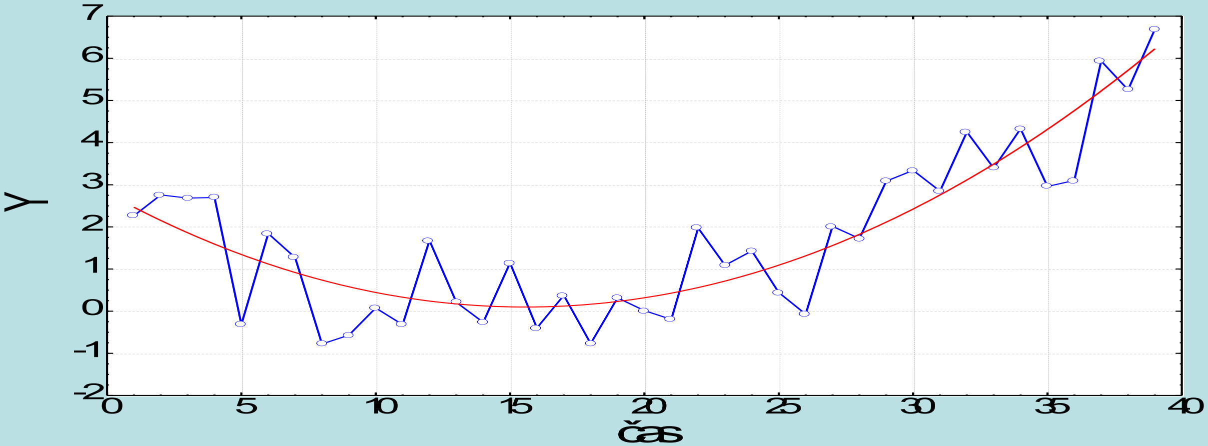


b) Kvadratický trend

Analytické vyjádření: $f(t) = \ldots$

Informativní test: 1. diference mají přibližně lineární trend, 2. diference (... $t = \ldots, n$) jsou přibližně konstantní.

Příklad kvadratického trendu:

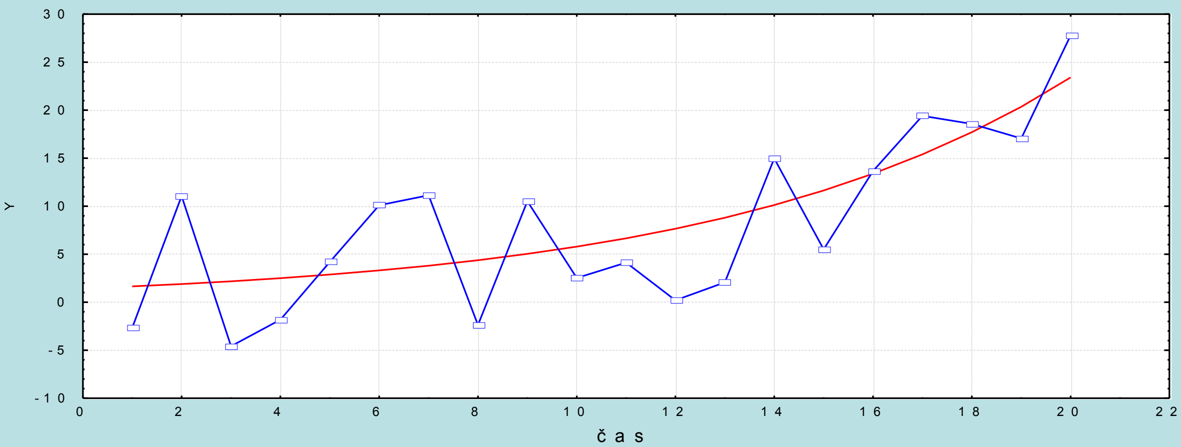


c) Exponenciální trend

Analytické vyjádření: $f(t) =$.

Informativní test: koeficienty růstu ($k_t = \frac{y_t}{y_{t-1}}, t = 2, \ldots, n$) jsou přibližně konstantní.

Příklad exponenciálního trendu:

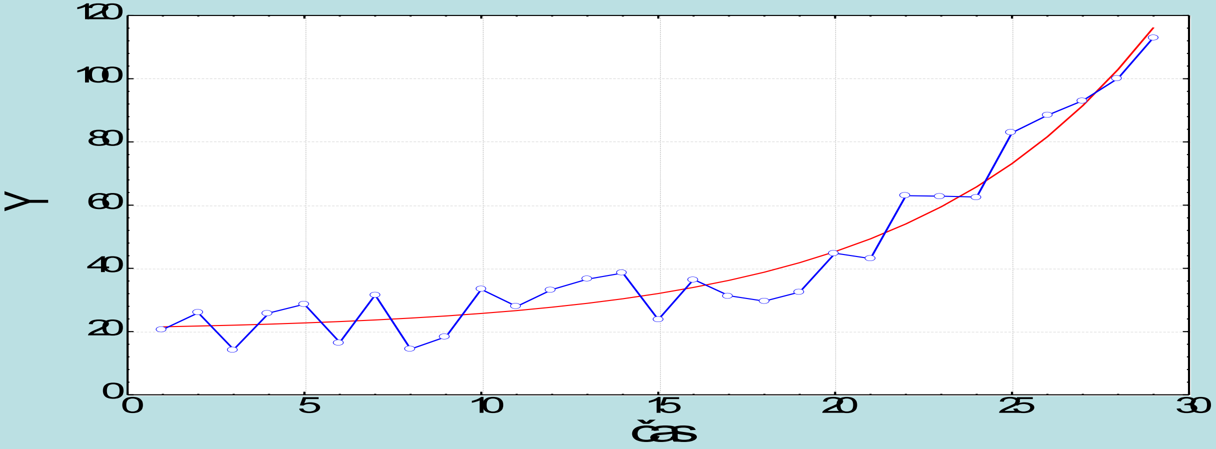


d) Modifikovaný exponenciální trend

Analytické vyjádření: $f(t) =$.

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

Příklad modifikovaného exponenciálního trendu

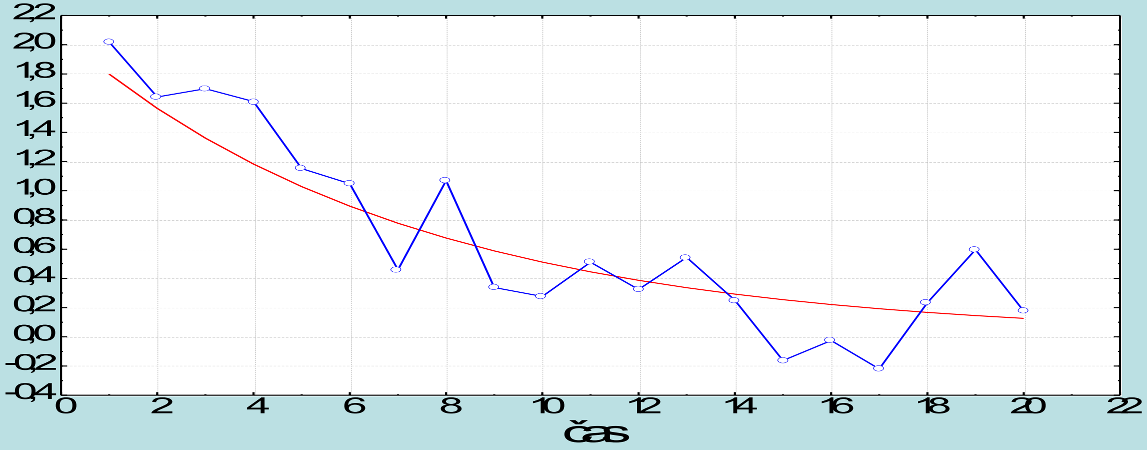


e) Logistický trend

Analytické vyjádření: $f(t) = \ldots \alpha \ldots$

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

Příklad logistického trendu:

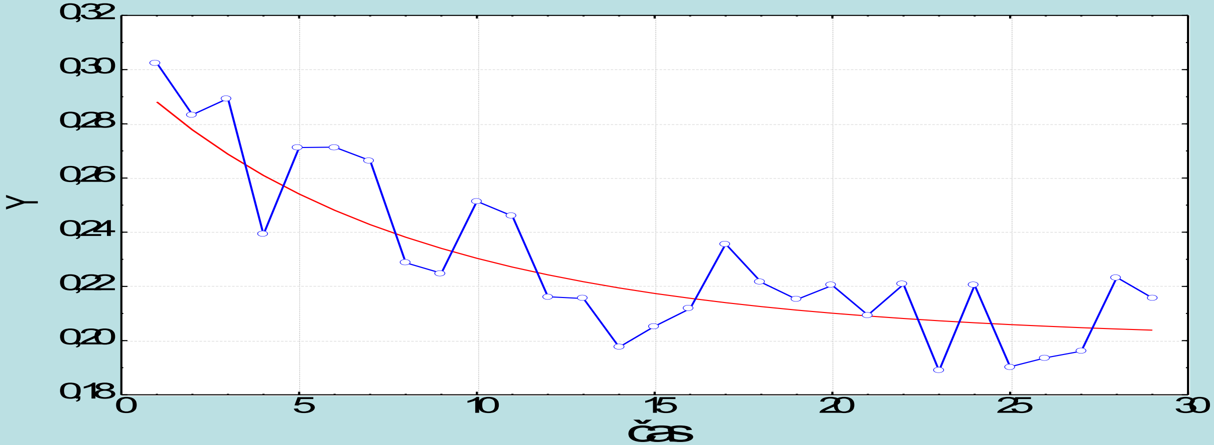


f) Gompertzova křivka

Analytické vyjádření: $f(t) = \beta$

Informativní test: podíly $\frac{\ln y_{t+} - y_{t+}}{\ln y_{t+} - y_{t}}$ jsou přibližně konstantní.

Příklad Gompertzovy křivky



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.

Orientační ověřování kvality modelu

- Index determinace (tj. podíl vysvětlené a celkové variability závisle proměnné veličiny) by měl být blízký 1.
- Body grafu $\left(f(t), \hat{f}(t)\right)$, t = 1, 2, ..., n by se měly řadit do přímky se směrnicí 1.

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

a) Graficky znázorněte průběh této časové řady.

b) Vypočtěte koeficienty růstu

c) Z grafu časové řady a chování koeficientů růstu lze usoudit, že časová řada má exponenciální trend $f(t) =$.

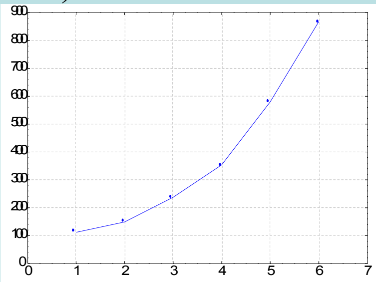
Odhadněte jeho parametry.

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

e) Zjistěte index determinace a sestrojte graf $\left(f(t), \hat{f}(t)\right)$, t = 1, ..., 6.

Řešení: Znovu uveďme hodnoty časové řady: 112, 149, 238, 354, 580, 867

ad a)

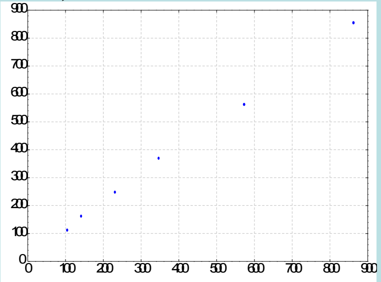


ad b) Koeficienty růstu: 149/112 = 1,33, 238/149 = 1,597, 354/238 = 1,487, 580/354 = 1,628, 867/580 = 1,495. Vidíme, že koeficienty růstu jsou přibližně konstantní.

ad c) Model $f(t) =$ linearizujeme a metodou nejmenších čtverců získáme odhady ln b_0 = 4, 227983, ln b_1 = 0,420199. Odlogaritmováním dostaneme b_0 = 68,57875, b_1 = 1,522265.

ad d) $y_7 = 68{,}57875 \cdot 1{,}522265^7 =$, $y_8 = 68{,}57875 \cdot 1{,}522265^8 =$

ad e) ID^2 = 0,996



Jak index determinace, tak graf $(f(t), \hat{f}(t))$ svědčí o tom, že model byl zvolen správně.

Výpočet pomocí systému STATISTICA:

Vytvoříme datový soubor se dvěma proměnnými čas a Y a 6 případy.

ad a) Časovou řadu znázorníme graficky pomocí Grafy – Bodové grafy.

ad b) Koeficienty růstu získáme pomocí Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce.

ad c) K datovému souboru přidáme novou proměnnou ln Y, kterou získáme zlogaritmováním proměnné Y, v níž jsou uloženy hodnoty zisku společnosti. Provedeme regresní analýzu se závisle proměnnou ln Y a nezávisle proměnnou čas. K výstupní tabulce přidáme novou proměnnou, do jejíhož Dlouhého jména napíšeme =exp(b)

N=6	Výsledky regrese se závislou proměnnou : lnY R= ,99801042 R2= ,99602479 Upravené R2= , F(1,4)=1002,2 p<,00001 Směrod. chyba odhad						
	b*	Sm.chy z b*	b	Sm.chy z b	t(4)	p-hod	NPro =exp(
Abs.čl			4,227	0,051	81,79	0,000	68,57
cas	0,998	0,031	0,420	0,013	31,65	0,000	1,522

Vidíme, že $Y = 68,578 \cdot 1,5222^t$.

ad d) Pro výpočet predikovaného zisku v 7. a 8. roce existence společnosti použijeme STATISTIKU jako kalkulačku.

ad e) Index determinace najdeme ve výstupní tabulce regrese pod označením R2. V našem případě je 0,996.

Pro získání grafu závislosti predikovaných hodnot na naměřených hodnotách přidám ek datovému souboru proměnnou predikce a do jejího Dlouhého jména napíšeme =68,57875*1,52265^cas. Pak vytvoříme Bodový graf.

Můžeme též nakreslit dvourozměrný tečkový diagram s odhadnutou regresní křivkou:
Na liště Detaily vybereme Proložení Exponenciální.