

10. Regresní analýza v časových řadách

Pojem časové řady: Časovou řadou rozumíme řadu hodnot y_{t_1}, \dots, y_{t_n} určité veličiny uspořádanou podle přirozené časové posloupnosti $t_1 < \dots < t_n$. Jsou-li časové intervaly $(t_1, t_2), \dots, (t_{n-1}, t_n)$ stejně dlouhé (ekvidistantní), zjednodušeně zapisujeme časovou řadu jako y_1, \dots, y_n .

Aditivní model časové řady

Předpokládejme, že pro časovou řadu y_1, \dots, y_n platí model

$$y_t = f(t) + \varepsilon_t, \quad t = 1, \dots, n, \text{ kde}$$

$f(t)$ 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**).

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 $\beta_0, \beta_1, \dots, \beta_k$ a známých funkcích $\varphi_0(t), \varphi_1(t), \dots, \varphi_k(t)$, které již neobsahují žádné neznámé parametry, tj. $f(t) = g(\beta_0, \beta_1, \dots, \beta_k; \varphi_0(t), \varphi_1(t), \dots, \varphi_k(t))$. Odhady b_0, b_1, \dots, b_k neznámých parametrů $\beta_0, \beta_1, \dots, \beta_k$ lze získat např. metodou nejmenších čtverců a pak vyjádřit odhad $\hat{f}(t)$ neznámého trendu v bodě t pomocí odhadů b_0, b_1, \dots, b_k a funkcí $\varphi_0(t), \varphi_1(t), \dots, \varphi_k(t)$, tj. $\hat{f}(t) = g(b_0, b_1, \dots, b_k; \varphi_0(t), \varphi_1(t), \dots, \varphi_k(t))$.

Při regresní analýze trendu hrají značnou roli difference a koeficienty růstu, které patří k dynamickým charakteristikám časové řady.

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

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

atd.

Koeficient růstu je dán vztahem $k_i = \frac{y_i}{y_{i-1}}, i = 2, \dots, n$. (Koeficient růstu po

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

Příklad: Je dána časová řada ročních hodnot HDP ČR (v miliardách Kč) v letech 1994 až 2000.

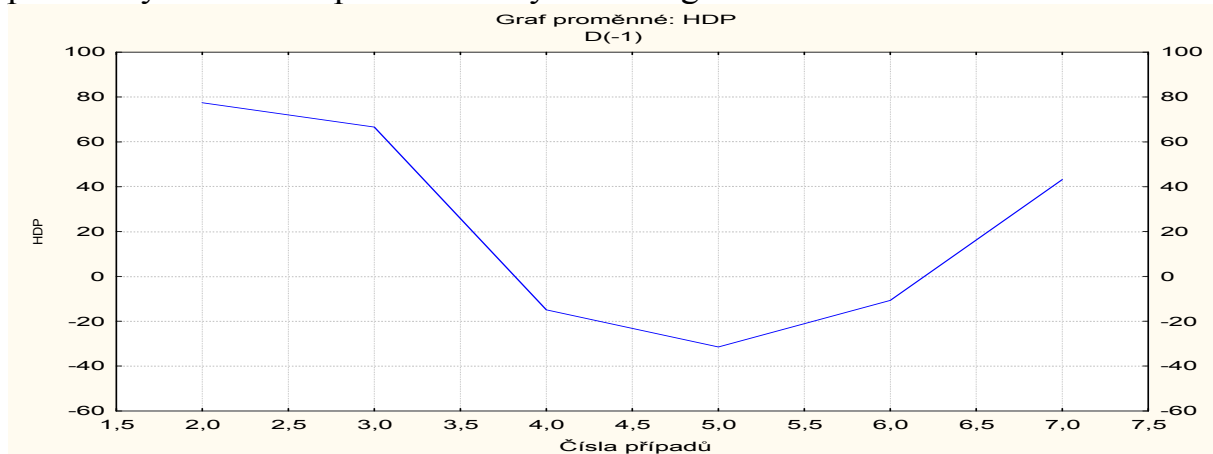
1994	1995	1996	1997	1998	1999	2000
1303,6	1381,1	1447,7	1432,8	1401,3	1390,6	1433,8

Vypočítejte a graficky znázorněte 1. diference a koeficienty růstu.

Řešení pomocí systému STATISTICA:

Výpočet 1. diferencí

Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné HDP – OK – OK (transformace, autokorelace, kříž. korelace, grafy) – Oddělit, sloučit - OK (transformovat vybrané řady) – návrat do transformace proměnných – Uložit proměnné. Vykreslí se graf 1. diferencí:



Ve výstupní tabulce máme proměnné HDP a HDP_1:

	HDP	
	1 HDP	2 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 koeficientů růstu:

Vrátíme se do Transformace proměnných a vybereme záložku Posun. Nastavíme se na proměnnou HDP – OK (transformovat vybrané řady). Vykreslí se graf 1. diferencí. Znovu se vrátíme do Transformace proměnných – Uložit proměnné. Ve výstupní tabulce teď máme proměnné HDP, HDP_2 a HDP_1. K této tabulce přidáme novou proměnnou KR (koeficienty růstu) a do jejího Dlouhého jména napíšeme =v1/v3

	HDP			
	1 HDP	2 HDP_2	3 HDP_1	4 KR
1	1303,600			
2	1381,100	77,500	1303,600	1,059451
3	1447,700	66,600	1381,100	1,048222
4	1432,800	-14,900	1447,700	0,989708
5	1401,300	-31,500	1432,800	0,978015
6	1390,600	-10,700	1401,300	0,992364
7	1433,800	43,200	1390,600	1,031066
8			1433,800	

Graf koeficientů růstu získáme tak, že klikneme pravým tlačítkem na záhlaví proměnné KR – Grafy bloku dat – Spojnicový graf: celé sloupce



Řešení pomocí systému SPSS:

K datovému souboru s proměnnými ROK a HDP přidáme novou proměnnou HDP_1, do níž okopírujeme hodnoty HDP posunuté o 1 rok vzad. Vytvoříme proměnnou DIFERENCE a pomocí Transform – Compute Variable do ní uložíme rozdíl HDP_1 - HDP. Dále vytvoříme proměnnou KR a pomocí Transform – Compute Variable do ní uložíme podíl HDP_1/ HDP.

	rok	HDP	HDP_1	diference	KR
1	1994	1303,6	1381,10	77,50	1,059
2	1995	1381,1	1447,70	66,60	1,048
3	1996	1447,7	1432,80	-14,90	0,990
4	1997	1432,8	1401,30	-31,50	0,978
5	1998	1401,3	1390,60	-10,70	0,992
6	1999	1390,6	1433,80	43,20	1,031
7	2000	1433,8	.	.	.

Grafy vytvoříme pomocí Graphs – Legacy Dialogs – Scatter/Dot.

Nejdůležitější typy trendových funkcí

Volba typu trendové funkce se provádí

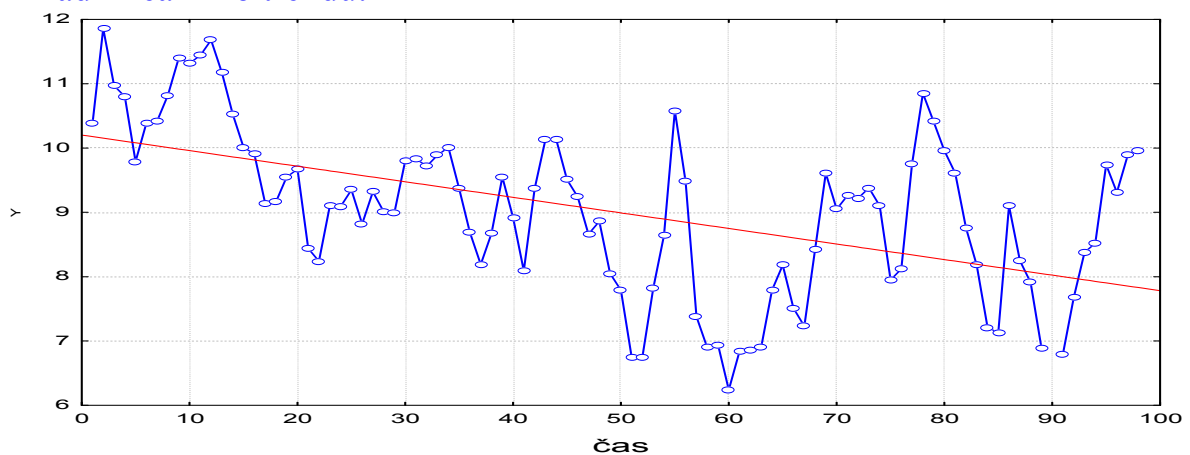
- na základě teoretických znalostí a zkušeností se zkoumanou veličinou Y_t
- pomocí grafu časové řady
- pomocí informativních testů založených na jednoduchých charakteristikách časové řady

a) Lineární trend

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

Informativní test: 1. diference jsou přibližně konstantní.

Příklad lineárního trendu:

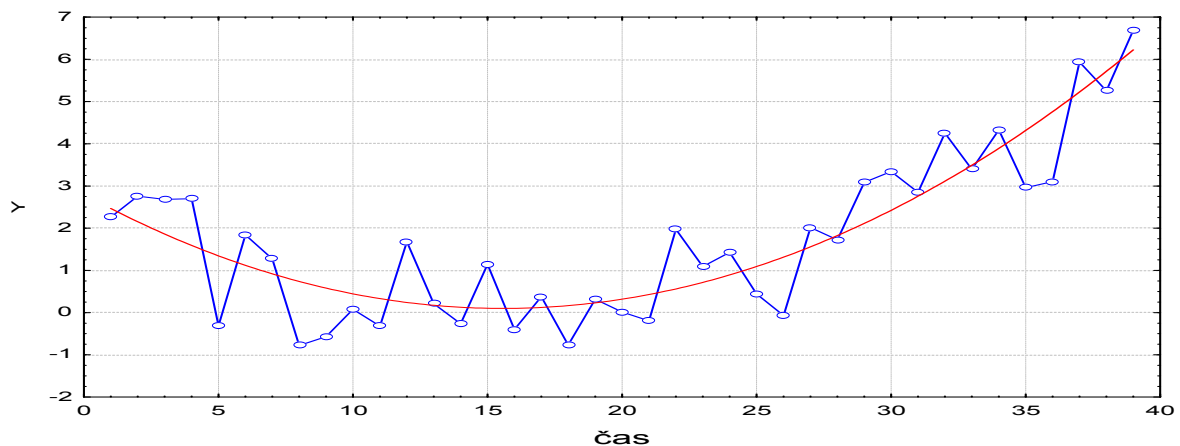


b) Kvadratický trend

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.

Příklad kvadratického trendu:



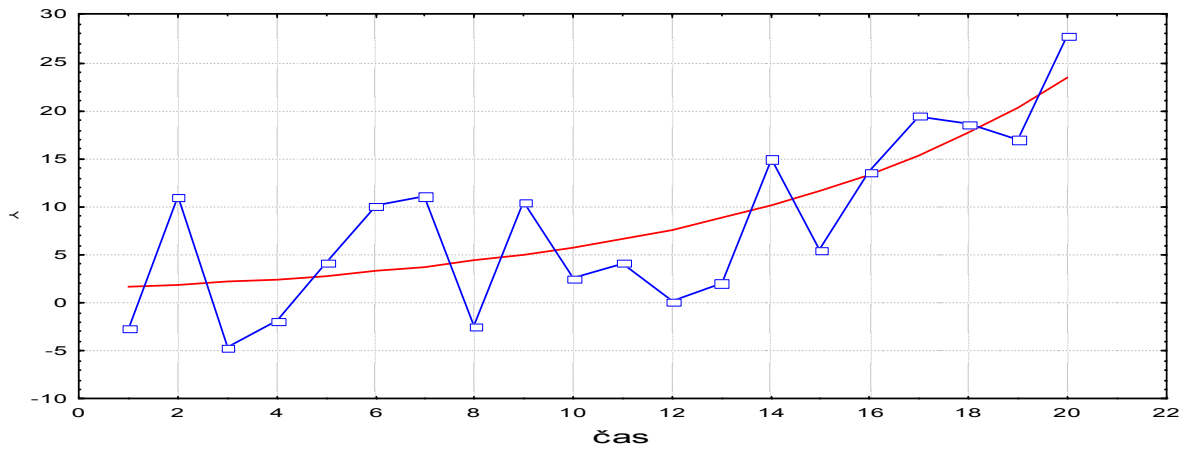
c) Exponenciální trend

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.

Příklad exponenciálního trendu:

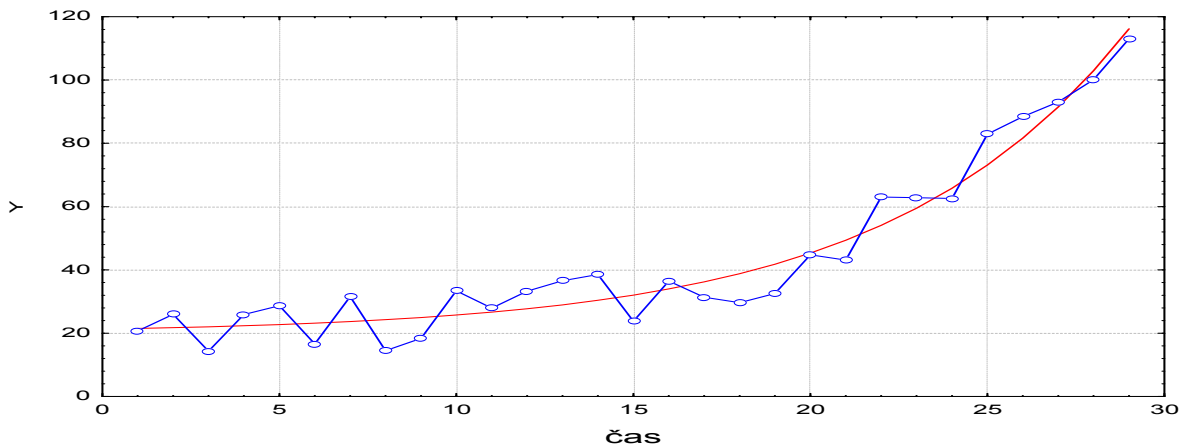


d) **Modifikovaný exponenciální trend**

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

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

Příklad modifikovaného exponenciálního trendu

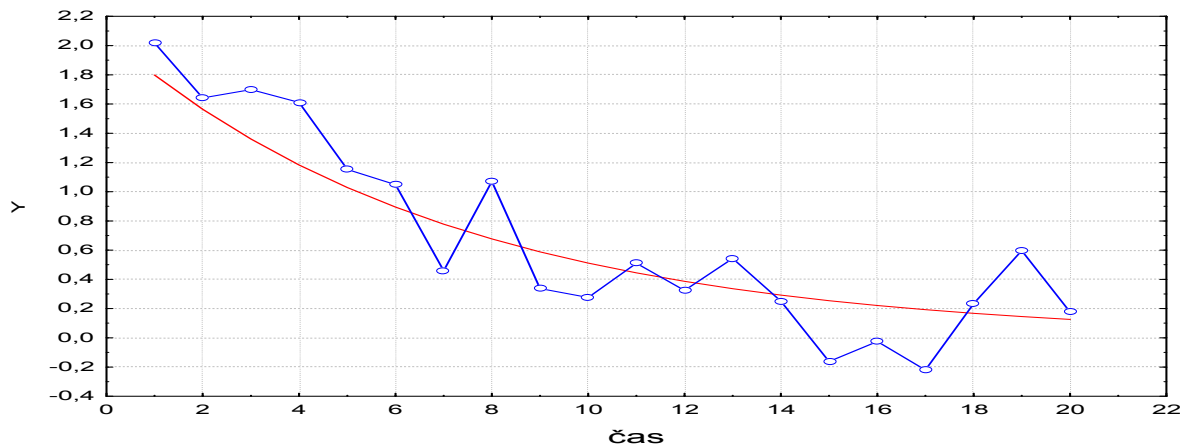


e) **Logistický trend**

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

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

Příklad logistického trendu:

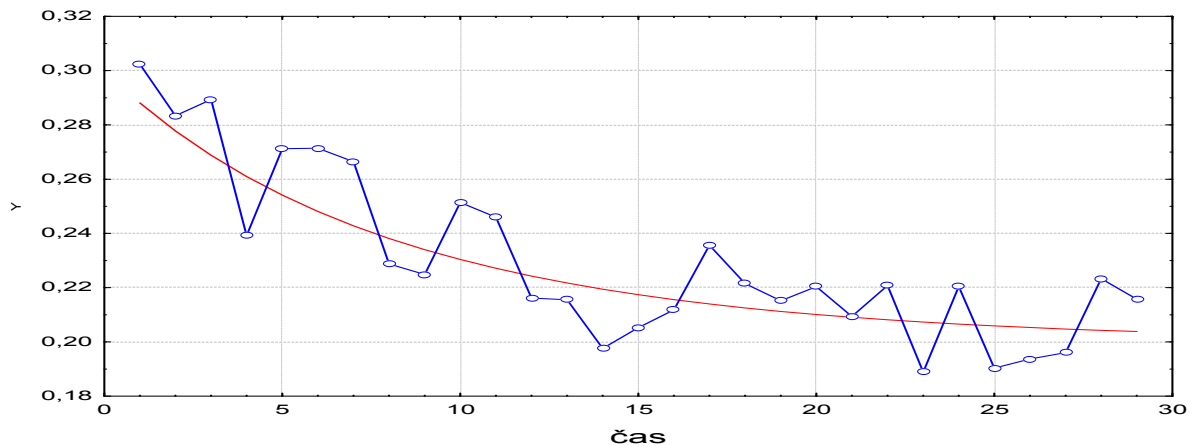


f) Gompertzova křivka

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

Informativní test: podíly $\frac{\ln y_{t+2} - \ln y_{t+1}}{\ln y_{t+1} - \ln y_t}$ 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 $(y_t, \hat{f}(t))$, $t = 1, 2, \dots, n$ by se měly řadit do přímky se směrnici 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) Vypočítejte koeficienty růstu a graficky je znázorněte.

c) 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.

- d) Najděte odhad zisku společnosti v 7. a 8. roce její existence.
 e) Zjistěte index determinace a sestrojte graf $(y_t, \hat{f}(t))_{t=1, \dots, 6}$.

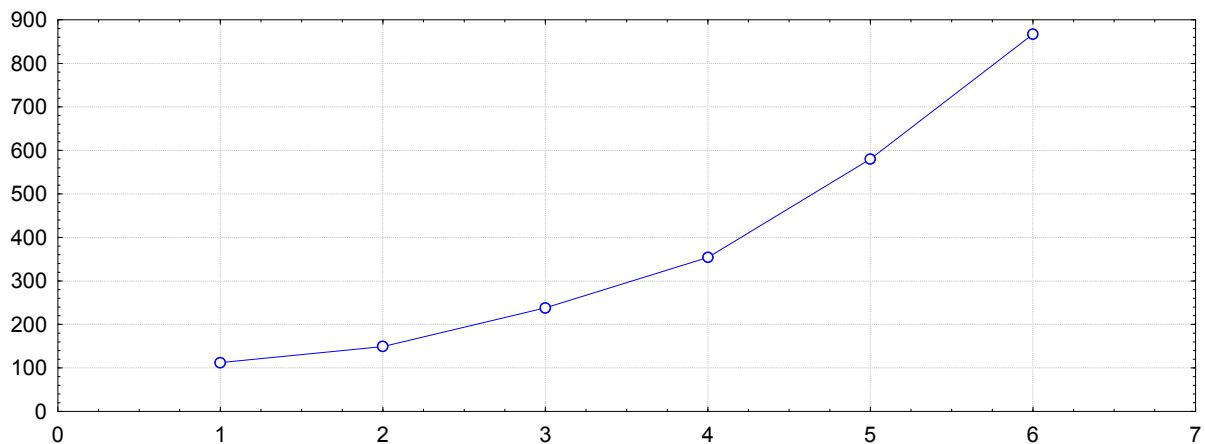
Řešení pomocí systému STATISTICA:

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

	1 čas	2 Y
1	1	112
2	2	149
3	3	238
4	4	354
5	5	580
6	6	867

ad a) Graficky znázorníme průběh této časové řady:

Grafy – Bodové grafy – Proměnné čas, Y – OK – vypneme proložení – OK.



ad b) Výpočet koeficientů růstu:

Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce –
 Proměnné Y – OK – OK (transformace, autokorelace, kříž. korelace, grafy) –
 Posun – Posun řad vzad - OK (transformovat vybrané řady) – návrat do
 transformace proměnných – Uložit proměnné.

Ve výstupní tabulce máme proměnné Y a Y_1:

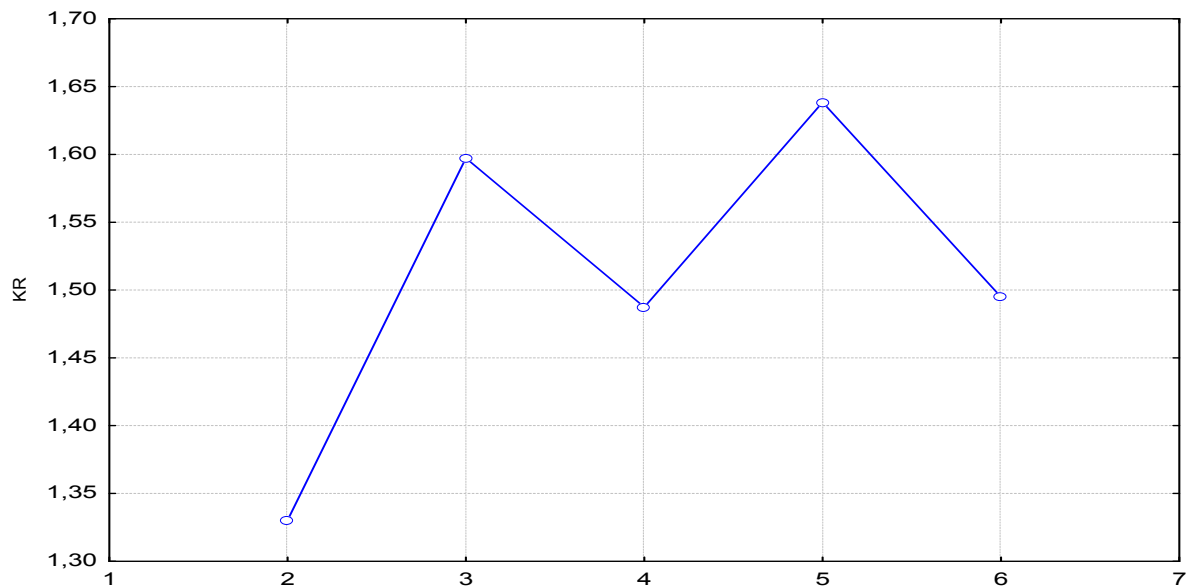
	1 Y	2 Y_1
0		112,000
1	112,000	149,000
2	149,000	238,000
3	238,000	354,000
4	354,000	580,000
5	580,000	867,000
6	867,000	
7		

Za proměnnou Y_1 přidáme proměnnou KR a do jejího Dlouhého jména napíšeme =v2/v1.

	1 Y	2 Y_1	3 KR
0		112,000	
1	112,000	149,000	1,330357
2	149,000	238,000	1,597315
3	238,000	354,000	1,487395
4	354,000	580,000	1,638418
5	580,000	867,000	1,494828
6	867,000		
7			

Vytvoření grafu koeficientů růstu:

Klikneme pravým tlačítkem na název proměnné KR – Grafy bloku dat –
Spojnicový graf: celé sloupce



Vidíme, že koeficienty růstu jsou přibližně konstantní.

ad c) Model $f(t) = \beta_0 \beta_1^t$ linearizujeme a metodou nejmenších čtverců získáme odhady $\ln b_0$, $\ln b_1$. Odlogaritmováním dostaneme $b_0 = 68,57875$, $b_1 = 1,522265$. K datovému souboru přidáme proměnnou $\ln Y$. Do jejího Dlouhého jména napíšeme =log(Y).

	1 čas	2 Y	3 lnY
1	1	112	4,718499
2	2	149	5,003946
3	3	238	5,472271
4	4	354	5,869297
5	5	580	6,363028
6	6	867	6,765039

Provedeme regresní analýzu se závisle proměnnou ln Y a nezávisle proměnnou čas.

Výsledky regrese se závisle proměnnou : lnY (Tabulka4)						
R= ,99801042 R2= ,99602479 Upravené R2= ,99503099						
F(1,4)=1002,2 p<,00001 Směrod. chyba odhadu : ,05553						
N=6	Beta	Sm.chyba beta	B	Sm.chyba B	t(4)	Úroveň p
Abs.člen			4,227983	0,051691	81,79336	0,000000
čas	0,998010	0,031525	0,420199	0,013273	31,65812	0,000006

Vidíme, že $\ln b_0 = 4,227983$, $\ln b_1 = 0,420199$.

K této tabulce přidáme proměnnou expB a do jejího Dlouhého jména napíšeme =exp(B).

Výsledky regrese se závislou proměnnou : lnY (Tabulka4)							
R= ,99801042 R2= ,99602479 Upravené R2= ,99503099							
F(1,4)=1002,2 p<,00001 Směrod. chyba odhadu : ,05553							
N=6	Beta	Sm.chyba beta	B	Sm.chyba B	t(4)	Úroveň p	expB =exp(B)
Abs.člen			4,227983	0,051691	81,79336	0,000000	68,57875
čas	0,998010	0,031525	0,420199	0,013273	31,65812	0,000006	1,522265

Získáme odhady $b_0 = 68,57875$, $b_1 = 1,522265$.

ad d) Odhad zisku společnosti v 7. roce existence: Pro výpočet predikované hodnoty zvolíme Rezidua/předpoklady/předpovědi - Předpovědi závisle proměnné čas: 7 - OK. Ve výstupní tabulce je hledaná hodnota označena jako Předpověď: 7,169377. K výstupní tabulce přidáme novou proměnnou a do jejího Dlouhého jména napíšeme =exp(v3).

Proměnná	B-váž.	Hodnota	B-váž. * Hodnot	NProm =exp(v3)
čas	0,420199	7,000000	2,941394	18,94224
Abs. člen			4,227983	68,57875
Předpověď			7,169377	1299,035
-95,0%LS			7,025860	1125,362
+95,0%LS			7,312894	1499,511

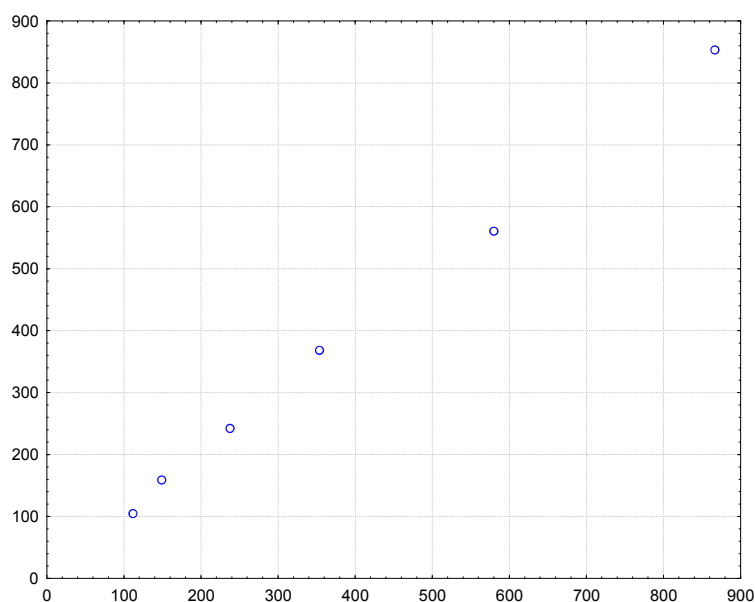
Předpověď zisku v 7. roce existence společnosti je tedy 1299,035 tisíc dolarů. Analogicky pro 8. rok zjistíme, že predikce zisku je 1977,47567

ad e) Index determinace je $ID^2 = 0,996$, jak je uvedeno v záhlaví výstupní tabulky regresní analýzy.

Graf závislosti predikovaných hodnot na hodnotách časové řady vytvoříme tak, že uložíme předpovězené hodnoty. K datovému souboru s předpovězenými hodnotami přidáme novou proměnnou predikce a do jejího Dlouhého jména napíšeme =exp(v3).

	1 čas	2 Y	3 Předpovědi	4 predikce
1	1	112	4,65	104,395
2	2	149	5,07	158,9169
3	3	238	5,49	241,9135
4	4	354	5,91	368,2565
5	5	580	6,33	560,584
6	6	867	6,75	853,357

Pak pomocí Bodového grafu vykreslíme závislost predikce na Y.



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

Řešení pomocí systému SPSS:

ad a) Graficky znázorníme průběh této časové řady:

Graphs – Legacy Dialogs – Scatter/Dot – Define – X Axes cas, Y Axes Y – OK.

ad b) Výpočet koeficientů růstu:

K datovému souboru s proměnnými cas a Y přidáme novou proměnnou Y_1, do níž okopírujeme hodnoty Y posunuté o 1 rok vzad. Vytvoříme proměnnou Y_1. Vytvoříme proměnnou KR a pomocí Transform – Compute Variable do ní uložíme podíl Y_1 / Y .

ad c) Získání odhadů regresních koeficientů v modelu $f(t) = \beta_0 \beta_1^t$:

K datovému souboru přidáme nezávisle proměnnou lnY. Její hodnoty získáme pomocí Transform – Compute Variable.

Provedeme regresní analýzu se závisle proměnnou ln Y a nezávisle proměnnou čas.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,228	,052		81,793	,000
	cas	,420	,013	,998	31,658	,000

a. Dependent Variable: lnY

Vidíme, že $\ln b_0 = 4,228$, $\ln b_1 = 0,420$. Odlogaritmováním získáme odhady $b_0 = 68,58$ a $b_1 = 1,52$.

ad d) Odhad zisku společnosti v 7. roce existence:

SPSS použijeme jak inteligentní kalkulačku. Pomocí Transform – Compute variable zapíšeme vzorec $68,58 * 1,52^{**7}$. Získáme odhad 1286.

ad e) Index determinace je $ID^2 = 0,996$, jak je uvedeno v tabulce Model Summary.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,998 ^a	,996	,995	,0555251

a. Predictors: (Constant), cas

Graf závislosti predikovaných hodnot na hodnotách časové řady vytvoříme tak, že uložíme předpovězené hodnoty. Ty se uloží do proměnné PRE_1.

K datovému souboru s předpovězenými hodnotami přidáme novou proměnnou predikce a pomocí Transform – Compute variable získáme její hodnoty jako $\exp(\text{PRE}_1)$

Odhad trendu časové řady pomocí klouzavých průměrů

Podstata klouzavých průměrů

Předpokládáme, že časová řada se řídí aditivním modelem

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

$t \in \{1, \dots, d\} \cup \{n - d + 1, \dots, n\}$ není vyhlazovací okénko symetrické. Odhad trendu ve

středu vyhlazovacího okénka je dán vztahem:

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

Šířka vyhlazovacího okénka

Velmi důležitou otázkou je stanovení šířky vyhlazovacího okénka. Je-li okénko příliš široké, bude se odhad trendu blížit průměru (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 $h = 3, 5, 7$.

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.

- Odhadněte trend této časové řady pomocí klouzavých průměrů s vyhlazovacím okénkem šířky 3 a poté 5.
- Graficky znázorněte průběh časové řady s odhadnutým trendem.

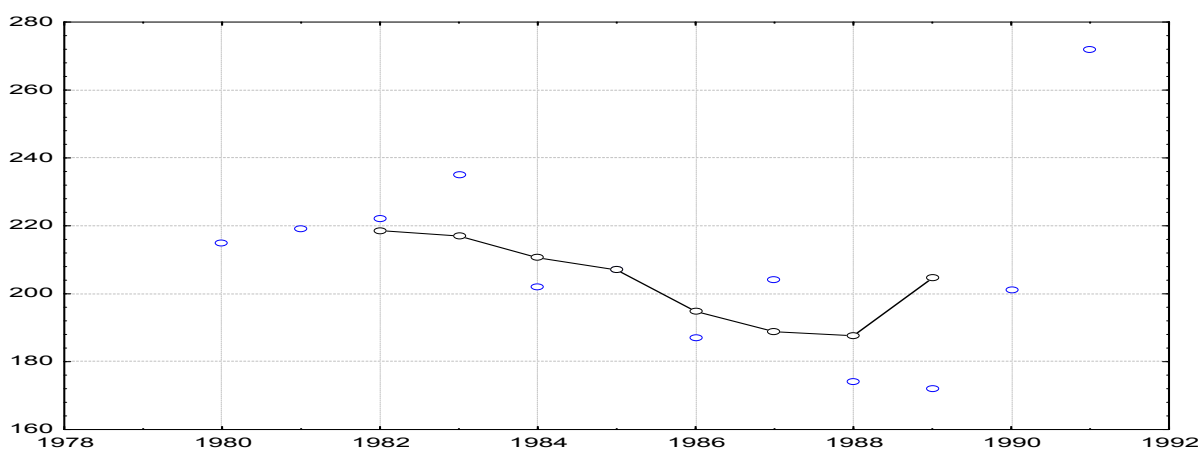
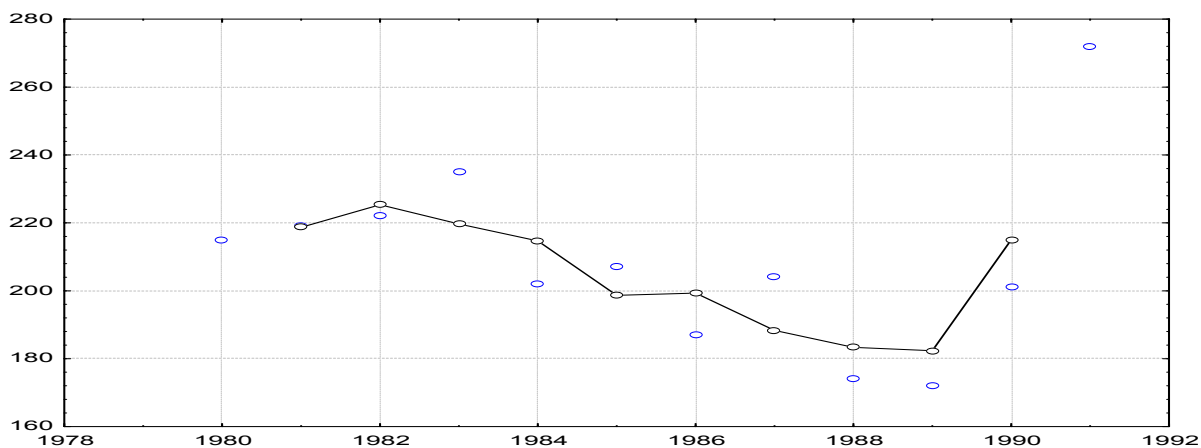
Řešení pomocí systému STATISTICA:

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 VYVOZ	3 KP3	4 KP5
1	1980	215,000		
2	1981	219,000	218,667	
3	1982	222,000	225,333	218,600
4	1983	235,000	219,667	217,000
5	1984	202,000	214,667	210,600
6	1985	207,000	198,667	207,000
7	1986	187,000	199,333	194,800
8	1987	204,000	188,333	188,800
9	1988	174,000	183,333	187,600
10	1989	172,000	182,333	204,600
11	1990	201,000	215,000	
12	1991	272,000		

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



Řešení pomocí systému SPSS:

Načteme datový soubor export_piva.sav.

Transform – Create Time series – New Variable VYVOZ – Function Centered
 Movin Average – Span 3 – klikneme na Change – OK

	ROK	VYVOZ	VYVOZ_1
1	1980	215	.
2	1981	219	218,67
3	1982	222	225,33
4	1983	235	219,67
5	1984	202	214,67
6	1985	207	198,67
7	1986	187	199,33
8	1987	204	188,33
9	1988	174	183,33
10	1989	172	182,33
11	1990	201	215,00
12	1991	272	.

Časovou řadu s odhadnutým trendem znázorníme pomocí Scatter/Dot, Overlay Scatter.