

Cvčení 14: Úvod do analýzy časových řad

Úkol 1: 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.

Návod: Načteme datový soubor `obyvatele_CSSR.sta` o 11 proměnných a jednom případě. Do Dlouhého jména poslední proměnné napíšeme

$$=(v1/2+\text{sum}(v2:v9)+v10/2)/9$$

Dostaneme výsledek 14430,11.

Úkol 2: Pro časovou řadu HDP ČR v letech 1994 až 2000 (v miliardách Kč) vypočtete základní charakteristiky dynamiky a graficky znázorníte relativní přírůstky a koeficienty růstu.

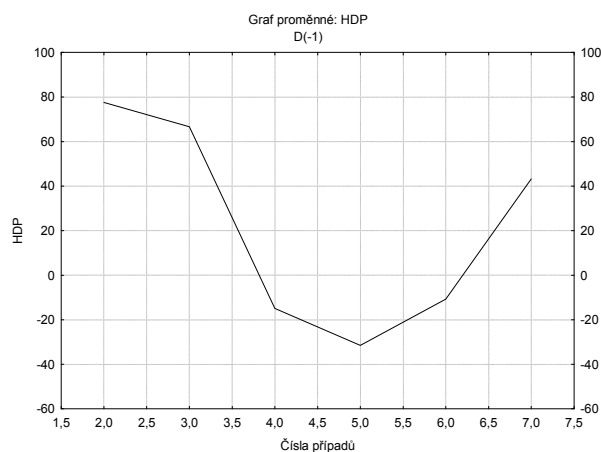
Návod: Načteme datový soubor `HDP.sta`.

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

Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné Y – OK

– OK (transformace, autokorelace, kříž. 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	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: $\delta_i = \frac{\Delta y_i}{y_{i-1}}$ pro $i = 2, \dots, n$

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 $i = 2, \dots, n$

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

Získáme tabulku

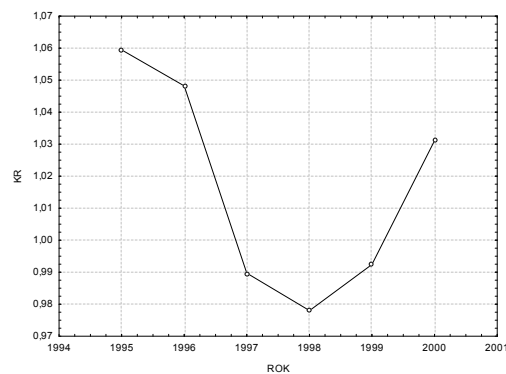
	HDP	HDP_2	HDP_1	RP	KR
1	1303,600				
2	1381,100	77,500	1303,600	0,059451	1,059451
3	1447,700	66,600	1381,100	0,048222	1,048222
4	1432,800	-14,900	1447,700	-0,010292	0,989708
5	1401,300	-31,500	1432,800	-0,02198	0,978015
6	1390,600	-10,700	1401,300	-0,00764	0,992364
7	1433,800	43,200	1390,600	0,031066	1,031066
8			1433,800		

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

Graf relativních přírůstků



Graf koeficientů růstu



Průměrný absolutní přírůstek a průměrný koeficient růstu vypočteme na kalkulačce pomocí

$$\text{vzorců } \bar{\Delta} = \frac{1433,8 - 1303,6}{6} = 21,7 \text{ a } \bar{k} = \sqrt[6]{\frac{1433,8}{1303,6}} = 1,016.$$

Úkol 3.: Je dána časová řada potratů (v tisících) v ČR v letech 1986 až 1996: 99,5 126,7 129,3 126,5 126,1 120,1 109,3 85,4 67,4 61,6 60.

Předpokládejte, že tato časová řada má kvadratický trend. Odhadněte parametry trendové funkce.

Vypočtete index determinace ID^2 .

Proveďte celkový F-test. (Popis celkového F- testu: Na hladině významnosti α testujeme

$H_0: (\beta_1, \dots, \beta_p)' = (0, \dots, 0)'$ proti $H_1: (\beta_1, \dots, \beta_p)' \neq (0, \dots, 0)'$, přičemž p je počet odhadovaných regresních parametrů (bez parametru β_0) (Nulová hypotéza říká, že dostačující je model konstanty.)

Testová statistika $F = \frac{S_R/p}{S_E/(n-p-1)}$ má rozložení $F(p, n-p-1)$, pokud H_0 platí. Přitom

$S_E = \sum_{t=1}^n (y_t - \hat{y}_t)^2$ je reziduální součet čtverců a $S_R = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2$ je regresní součet čtverců,

kde $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$.

Kritický obor: $W = \langle F_{1-\alpha}(p, n-p-1), \infty \rangle$.

$F \in W \Rightarrow H_0$ zamítáme na hladině významnosti α .

Proveďte dílčí t-testy. (Popis dílčích t-testů: Na hladině významnosti α pro $j = 0, 1, \dots, p$ testujeme hypotézu

$H_0: \beta_j = 0$ proti $H_1: \beta_j \neq 0$.

Testová statistika: $T_j = \frac{b_j}{s_{b_j}}$ má rozložení $t(n-p-1)$, pokud H_0 platí. Přitom s_{b_j} je směrodatná

chyba odhadu b_j .

Kritický obor: $W = (-\infty, -t_{1-\alpha/2}(n-p-1)) \cup \langle t_{1-\alpha/2}(n-p-1), \infty \rangle$.

$T_j \in W \Rightarrow H_0$ zamítáme na hladině významnosti α .)

Ověřte normalitu reziduí.

Sestrojte 95% intervaly spolehlivosti pro parametry trendové funkce. (Vzorec pro meze 100(1- α)% intervalu spolehlivosti pro β_j : $b_j \pm t_{1-\alpha/2}(n-p-1)s_{b_j}$)

Stanovte střední absolutní procentuální chybu predikce (MAPE). MAPE se počítá podle

vzorce $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$.

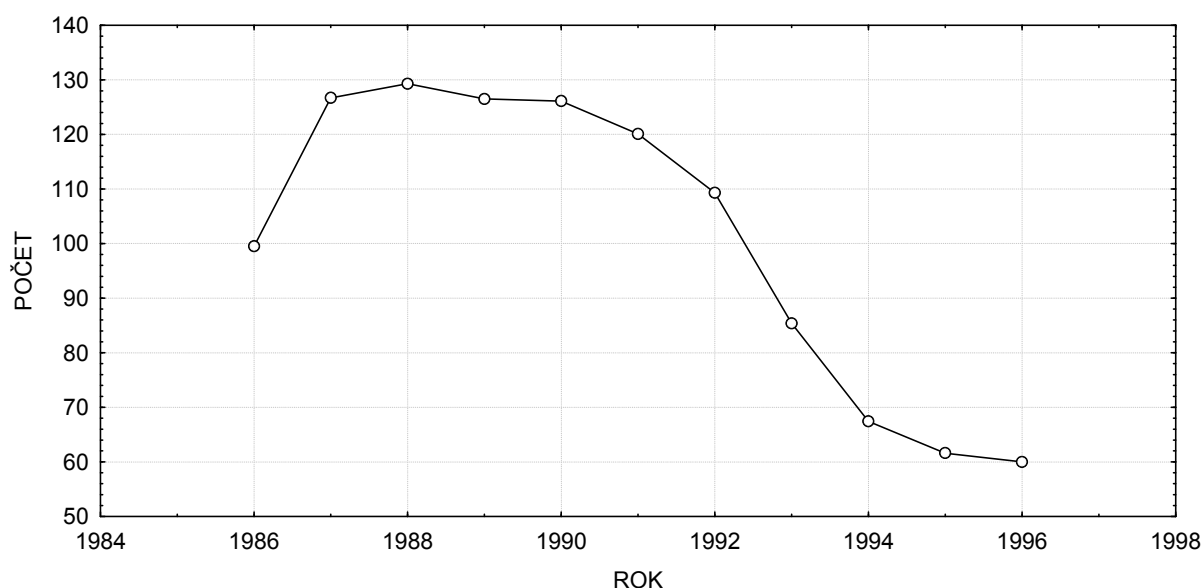
Graficky znázorněte průběh časové řady s odhadnutým trendem, 95% pásem spolehlivosti a 95% predikčním pásem.

Návod:

Načteme datový soubor potraty.sta. Pro lepší orientaci znázorníme časovou řadu graficky.

Grafy – Bodové grafy – Proměnné X ROK, Y POCET – OK – vypneme Lineární proložení– OK.

Formát – Všechny možnosti – Graf: Obecné – zaškrtneme Spojnice – OK. Vznikne spojnicový diagram.



Trendová funkce $\hat{f}(t) = \beta_0 + \beta_1 t + \beta_2 t^2$

Odhady parametrů:

Statistiky – Vícenásobná regrese – Proměnné Závislé, Nezávislé t, tkv - OK

Výsledky regrese se závislou proměnnou : POCET (potraty.sta)						
R= ,94015284 R2= ,88388736 Upravené R2= ,85485920						
F(2,8)=30,449 p<,00018 Směrod. chyba odhadu : 10,629						
N=11	Beta	Sm.chyba beta	B	Sm.chyba B	t(8)	Úroveň p
Abs.člen			103,2418	11,67235	8,84499	0,000021
t	1,30140	0,531476	10,9470	4,47060	2,44866	0,040020
tkv	-2,16020	0,531476	-1,4748	0,36285	-4,06453	0,003611

Odhadnutá trendová funkce má tedy tvar:

$$\hat{f}(t) = 103,2418 + 10,947t - 1,4748t^2, \text{ kde } t = 1, \dots, 11.$$

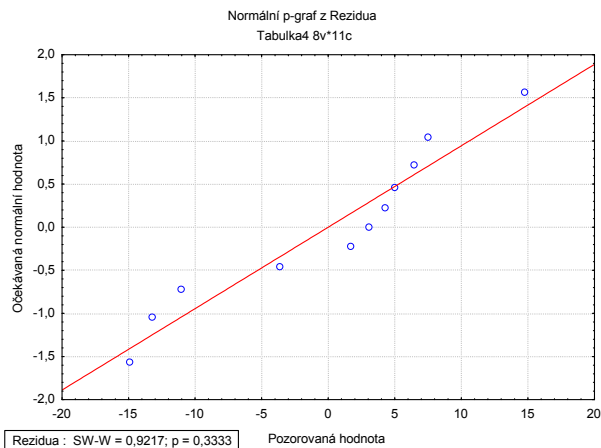
Index determinace je 0,884, což znamená, že kvadratická trendová funkce vysvětluje variabilitu dané časové řady z 88,4%.

Testová statistika celkového F-testu je 30,449, p-hodnota je blízká 0, tedy na hladině významnosti 0,05 zamítáme hypotézu o nevýznamnosti modelu jako celku.

Všechny tři dílčí t-testy mají p-hodnoty menší než 0,05, tedy na hladině významnosti 0,05 zamítáme hypotézy o nulovosti parametrů β_0 , β_1 , β_2 .

Ověření normality reziduí:

Na záložce Rezidua/předpoklady/předpovědi zvolíme Reziduální analýza – Uložit – Uložit rezidua & předpovědi. Sestrojíme N-P plot reziduí a současně provedeme S-W test:



S-W test poskytuje p-hodnotu 0,333, tedy na hladině významnosti 0,05 nezamítáme hypotézu o normalitě reziduí.

Sestrojení 95% intervalů spolehlivosti pro parametry trendu:

Ve výstupní tabulce výsledků regrese přidáme za proměnnou Úroveň p dvě nové proměnné dm (pro dolní meze 95% intervalů spolehlivosti) a hm (pro horní meze 95% intervalů spolehlivosti). Do Dlouhého jména proměnné dm resp. hm napíšeme: $=v_3-v_4*VStudent(0,975;8)$ resp. $=v_3+v_4*VStudent(0,975;8)$

Výsledky regrese se závislou proměnnou : POCET (potraty.sta)								
R= ,94015284 R2= ,88388736 Upravené R2= ,85485920								
F(2,8)=30,449 p<,00018 Směrod. chyba odhadu : 10,629								
N=11	Beta	Sm.chyba beta	B	Sm.chyba B	t(8)	Úroveň p	dm =v3-v4*	hm =v3+v4
Abs.člen			103,2418	11,67235	8,84499	0,000021	76,32533	130,1583
t	1,30140	0,531476	10,9470	4,47060	2,44866	0,040020	0,637767	21,25622
tkv	-2,16020	0,531476	-1,4748	0,36285	-4,06453	0,003611	-2,31156	-0,63809

Vidíme, že $76,32 < \beta_0 < 130,16$ s pravděpodobností aspoň 0,95, $0,64 < \beta_1 < 21,26$ a $-2,31 < \beta_2 < -0,64$ s pravděpodobností aspoň 0,95.

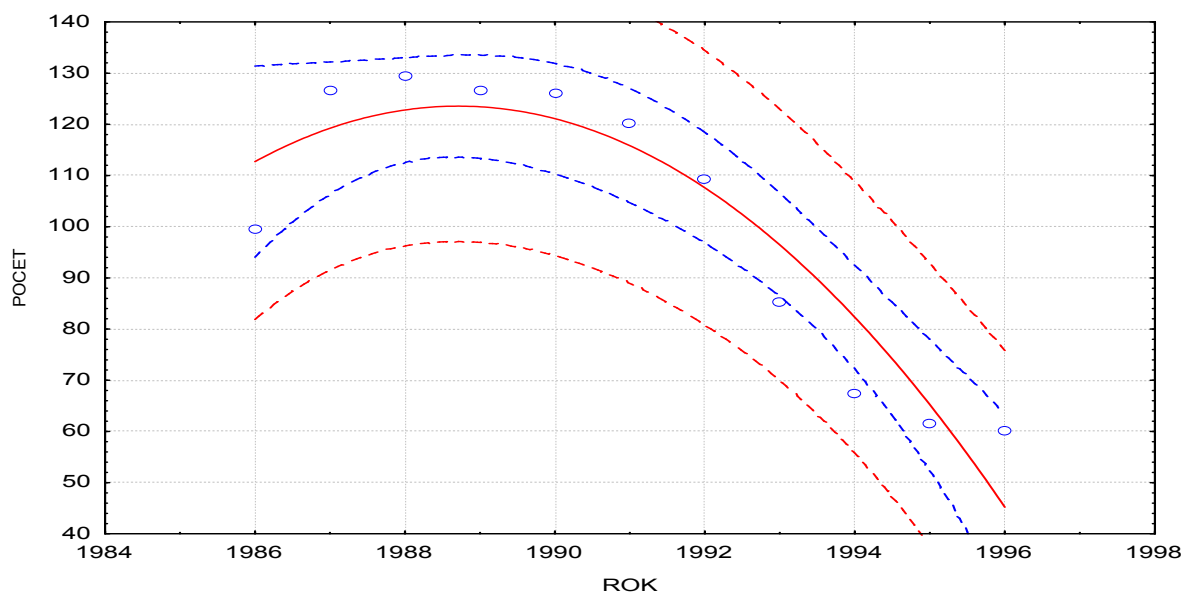
Výpočet MAPE:

Ve výsledcích Vícenásobné regrese zvolíme záložku Residua / předpoklady / předpovědi – Residuální analýza – Uložit – Uložit rezidua a předpovědi – Vybrat vše – OK. Ve vzniklé tabulce odstraníme proměnné 7 – 12, přidáme proměnnou chyby a do jejího Dlouhého jména napíšeme $=100*abs(v_6/v_2)$. Pak spočteme průměr této proměnné a zjistíme, že MAPE = 9,21%.

Graf časové řady s proloženým kvadratickým trendem získáme takto:

Grafy – Bodové grafy – Proměnné X ROK, Y POCET – OK – Detaily Proložení

Polynomiální. Ve vytvořeném grafu 2x klikneme na pozadí, vybereme Graf: Regresní pásy – Přidat nový pár pásů – Typ Spolehlivostní – OK. Totéž provedeme ještě jednou a nyní zaškrtneme Typ Predikční.



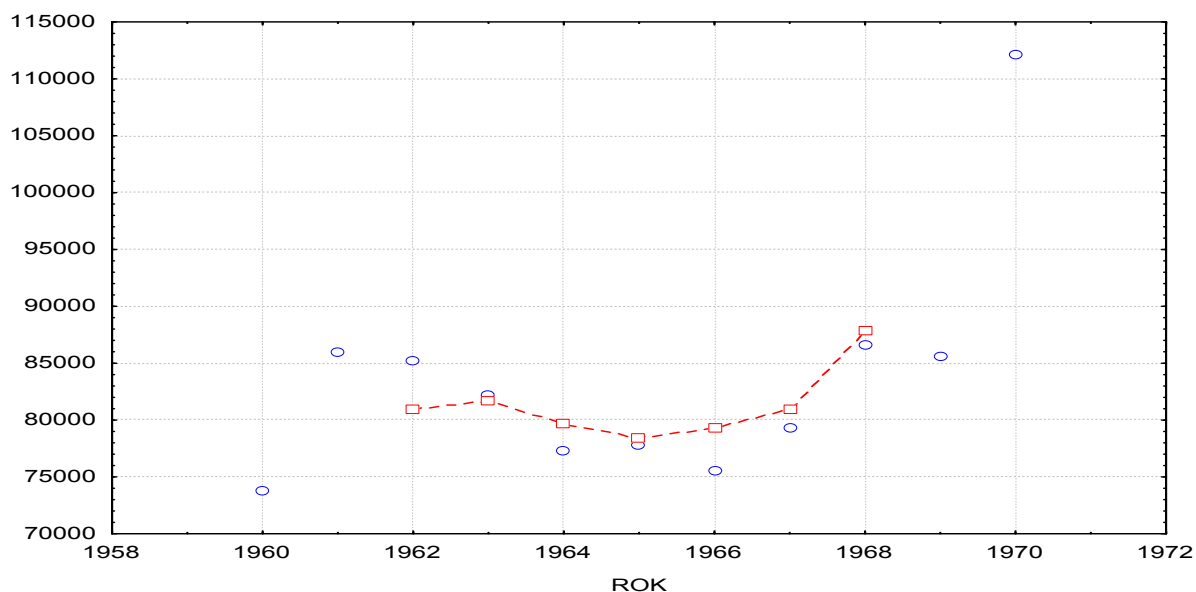
Úkol 4.: Máme k dispozici údaje o počtu bytů předaných do užívání v Československu v letech 1960 až 1970: 73 766 86 032 85 221 82 189 77 301 77 818 75 576 79 297 86 571 85 656 112 135. Odhadněte trend této časové řady pomocí klouzavých průměrů s vyhlazovacím okénkem šířky 5 a graficky znázorněte.

Návod:

Načteme datový soubor byty.sta o dvou proměnných ROK a POCET a jedenácti případech. Statistiky – Pokročilé lineární/nelineární modely – Časové řady/predikce – Proměnné POCET – OK– OK (transformace, autokorelace, kříž. korelace, grafy) – Vyhlazování – zaškrtneme N-bod. klouzavý průměr, N = 5 – OK (Transformovat vybrané řady) – vykreslí se graf, vrátíme se do Transformace proměnných – Uložit proměnné. Otevře se nový spreadsheet, kde v proměnné POCET_1 jsou uloženy klouzavé průměry pro N = 5. Proměnnou POCET_1 okopírujeme do původního datového souboru do nové proměnné KP5 (pozor – roky 1960, 1961, 1969 a 1970 nemají přiřazený odhad).

	1 ROK	2 POCET	3 KP5
1	1960	73766	
2	1961	86032	
3	1962	85221	80901,8
4	1963	82189	81712,2
5	1964	77301	79621,0
6	1965	77818	78436,2
7	1966	75576	79312,6
8	1967	79297	80983,6
9	1968	86571	87847,0
10	1969	85656	
11	1970	112135	

Pomocí Grafy – Bodové grafy – Vícenásobný graf vytvoříme graf časové řady počtu bytů s odhadnutým trendem.



Příklady k samostatnému řešení:

Příklad 1.: V průběhu jednoho roku byla čtyřikrát provedena inventarizace skladových zásob. Určete průměrný stav zásob ve sledovaném roce, jestliže hodnoty jsou uvedeny v tabulce. Pro jednoduchost počítejte, že každý měsíc má 30 dní.

inventarizace	1	2	3	4
datum	2.1.	2.3.	12.9.	30.12.
zásoby [tis. Kč]	752	652	925	426

Výsledky: použijeme vážený chronologický průměr a zjistíme, že průměrný stav zásob je 739,9 [tis. Kč].

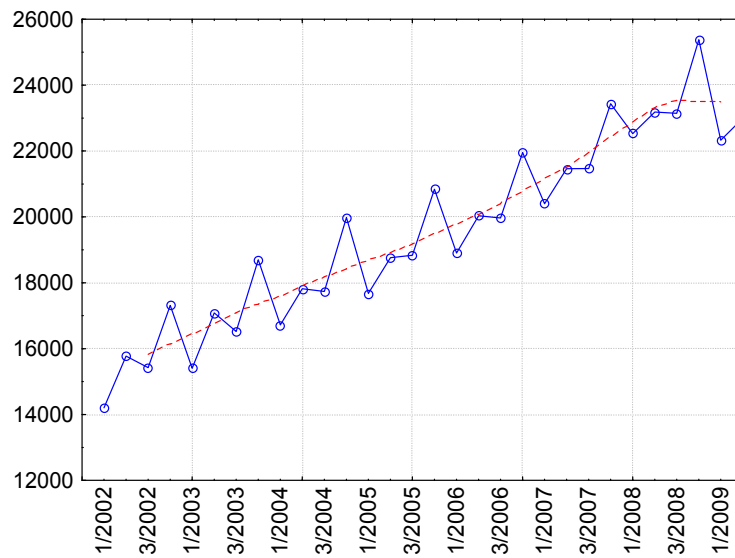
Příklad 2.: Máme k dispozici čtvrtletní časovou řadu průměrných měsíčních mezd v České republice v době od 1/2001 do 3/2009 (datový soubor ctvrtletni_mzda.sta):

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.

Výsledky:



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