

Trendová analýza – řešený příklad

Máme k dispozici údaje o počtu dokončených bytů v tis. v ČR 2007-2018. Vyrovnajte tyto údaje pomocí vhodné trendové funkce. U všech testů uvažujte $\alpha = 0,05$.

Rok	Počet dokončených bytů
2007	41 649
2008	38 380
2009	38 473
2010	36 442
2011	28 630
2012	29 467
2013	25 238
2014	23 954
2015	25 095
2016	27 322
2017	28 569
2018	33 850

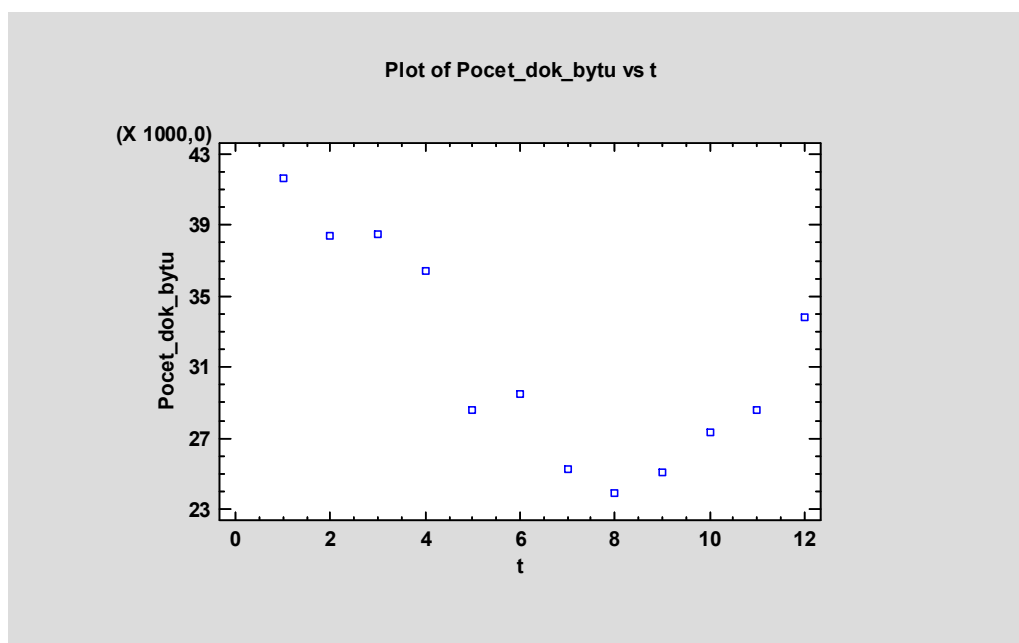
Zdroj: www.czso.cz

Řešení:

Pokud máme sami určit, jaká funkce má být vhodná k vyhlazení (vyrovnání) dat, je nejjednodušší zobrazit si průběh ukazatele v čase do bodového diagramu. Vzhledem k časové náročnosti výpočtů už zde není ruční řešení příkladu, je zde uveden pouze postup výpočtu přes program SGP. Pokud tedy chceme zkonstruovat bodový diagram, je potřeba vložit do prvního sloupce data o počtu dokončených bytů v ČR a do druhého časovou proměnnou t , která zde nabývá hodnot 1, 2, ..., 12 – viz Obr. 1. Vycházíme z toho, že v jednorozměrném modelu časové řady je jediným faktorem udávající dynamiku sledovaného ukazatele čas. Potom vybereme posloupnost procedur **Plot – Scatterplots – X-Y Plot ...** Bodový diagram najdeme na Obr. 2. Podle toho, jak vypadá průběh daného ukazatele, můžeme usuzovat, že by mohla být k vyrovnání hodnot vhodná parabola.

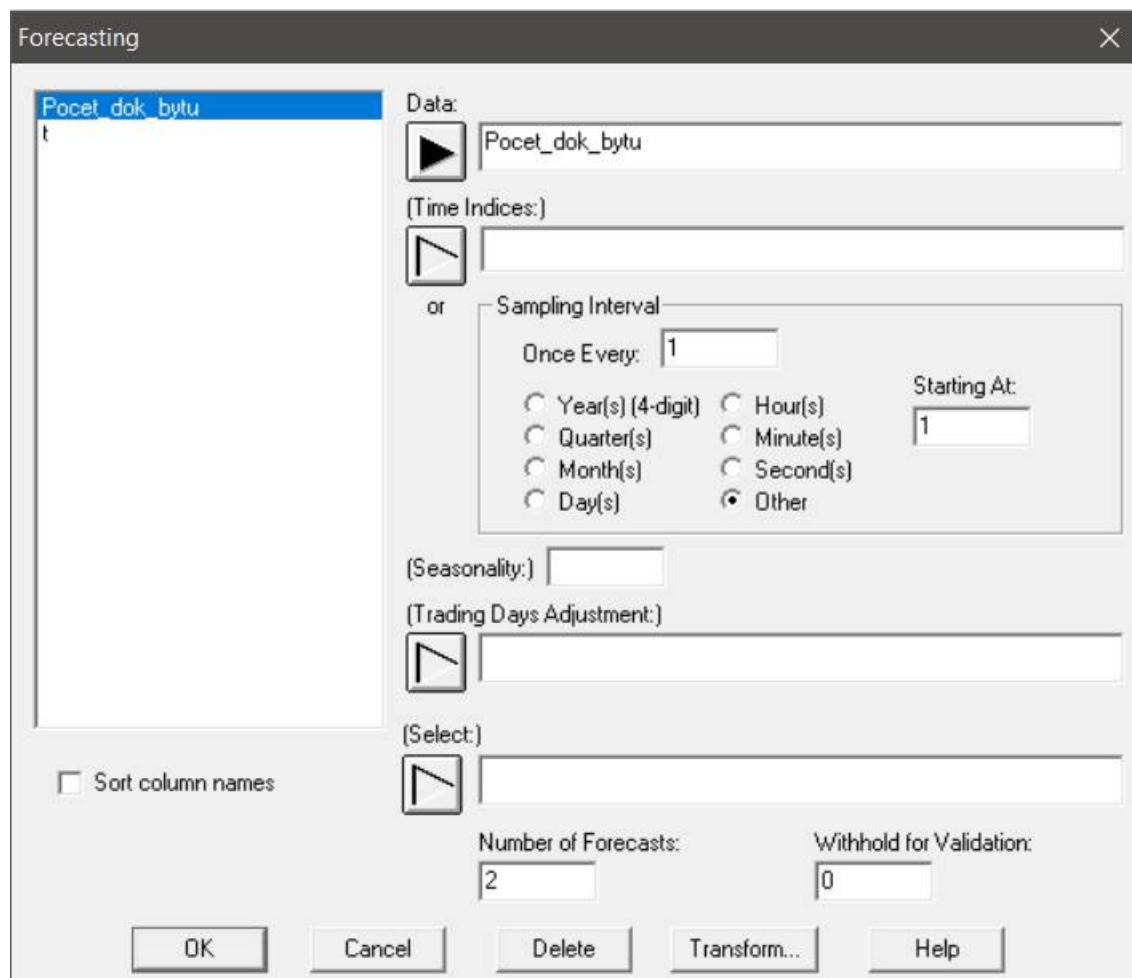
	Pocet_dok_bytu	t
	Numeric	Numeric
1	41649	1
2	38380	2
3	38473	3
4	36442	4
5	28630	5
6	29467	6
7	25238	7
8	23954	8
9	25095	9
10	27322	10
11	28569	11
12	33850	12

Obrázek 1 – Zadání dat pro vytvoření bodového diagramu



Obrázek 2 – Bodový diagram

Jedna z procedur, které se používají v analýze časových řad, je procedura Forecast. Zvolíme tedy **Forecast – User-Specified Model ...** Do řádku *Data* zadáme `Pocet_dok_bytu` a jinak nic měnit a doplňovat nemusíme. Ve spodní části panelu můžeme upravit počet předpovědí na 2 (Pole *Number of Forecasts*), protože 12 jich určitě potřebovat nebudeme. Potvrdíme klávesou OK. Panel pro zadání proměnné v proceduře *Forecast* vidíme na Obr. 3.



Obrázek 3 – Zadání dat ve vstupním panelu Forecasting

Poté se objeví panel, ve kterém můžeme zadat specifika požadovaného modelu – viz Obr. 4. Je zde možné vyvolat výsledky pro celkem 5 modelů, kterým přidělíme písmena A až E v oddílu *Model* vlevo nahoře. Pokud je označen model A, přesuneme puntík v oddílu *Type* na *Quadratic Trend*. Pokud nás budou zajímat i jiné funkce, přesuneme puntík na model B a v oddílu *Type* označíme např. *Linear* atd. Po stisknutí tlačítka OK, se objeví nabídka *Tables and Graphs*, kterou nemusíme nijak měnit.

Když se podíváme na výstup *Analysis Summary*, vidíme v úvodu rovnici trendové paraboly, kterou vyrovnáváme data. Ta má tvar: $\hat{T}_t = 49510,6 - 5709,5t + 351,205t^2$. Pod ní jsou pak hodnoty charakteristik, které se používají k posouzení kvality trendové funkce. Následují pak t-testy o nulové hodnotě parametrů trendové funkce. Jak můžeme vidět, všechny jsou významné na hladině významnosti 5 %. Celkový F-test o vhodnosti celého modelu tu ovšem chybí. Graf trendové paraboly najdeme v pravé části pole s výstupy pod názvem *Time Sequence Plot*. Je vidět na Obr. 6.

Obrázek 4 – Specifikace modelu trendu

Forecasting - Pocet dok bytu

Data variable: Pocet_dok_bytu

Number of observations = 12

Start index = 1,0

Sampling interval = 1,0

Forecast Summary

Forecast model selected: Quadratic trend = $49510,6 + -5709,5t + 351,205t^2$

Number of forecasts generated: 2

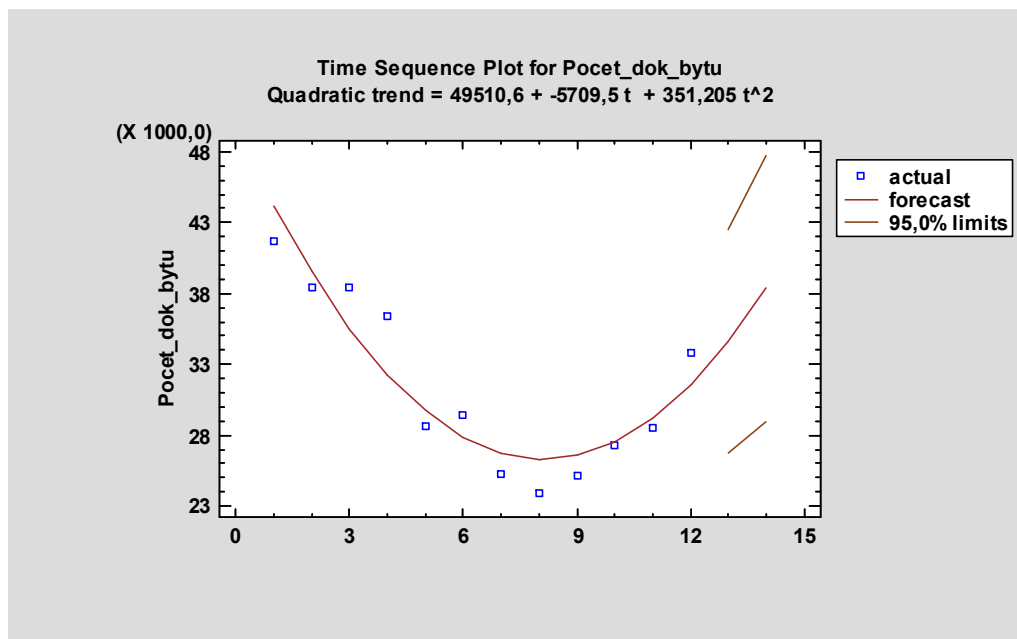
Number of periods withheld for validation: 0

	Estimation	Validation
Statistic	Period	Period
RMSE	2420,02	
MAE	1821,7	
MAPE	5,71734	
ME	-3,03165E-12	
MPE	-0,539353	

Trend Model Summary

Parameter	Estimate	Std. Error	t	P-value
Constant	49510,6	2501,16	19,7951	0,000000
Slope	-5709,5	884,602	-6,45432	0,000118
Quadratic	351,205	66,2417	5,30187	0,000493

Obrázek 5 – Výstup Analysis Summary



Obrázek 6 – Trendová parabola

Ve výstupu *Model Comparisons* (obr. 7) najdeme srovnání vybraných charakteristik používaných k posouzení kvality trendových funkcí a rovnice srovnávaných modelů.

Nevýhodou zde použité procedury k vyhledání vhodného typu trendové funkce je, že neobsahuje všechny potřebné údaje, na základě kterých je možné vybrat vhodný typ trendové funkce. Existuje však jednoduché a navíc už známé řešení – použít proceduru užívanou v regresní analýze. Tato úvaha se odvíjí od myšlenky, že v jednoduchém modelu časové řady nejde vlastně o nic jiného, než o popis závislosti sledovaného ukazatele na časové proměnné t .

Model Comparison

Data variable: Pocet_dok_bytu
Number of observations = 12
Start index = 1,0
Sampling interval = 1,0

Models

(A) Quadratic trend = $49510,6 + -5709,5t + 351,205t^2$
(B) Linear trend = $38857,3 + -1143,84t$
(C) Exponential trend = $\exp(10,5633 + -0,0345757t)$
(D) S-curve trend = $\exp(10,2021 + 0,527589t)$
(E) Simple exponential smoothing with $\alpha = 0,9999$

Estimation Period

Model	RMSE	MAE	MAPE	ME	MPE
(A)	2420,02	1821,7	5,71734	-3,03165E-12	-0,539353
(B)	4661,9	3660,91	12,2125	-2,42532E-12	-1,91261
(C)	4499,93	3586,09	11,7817	301,193	-0,940639
(D)	4302,42	3327,12	10,5133	220,528	-0,768957
(E)	3461,35	2454,26	8,26999	-649,998	-2,31262

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	2420,02	OK	OK	OK	OK	OK
(B)	4661,9	OK	*	*	OK	OK
(C)	4499,93	OK	*	*	OK	OK
(D)	4302,42	OK	OK	OK	OK	OK
(E)	3461,35	OK	OK	OK	OK	OK

Key:

RMSE = Root Mean Squared Error
RUNS = Test for excessive runs up and down
RUNM = Test for excessive runs above and below median
AUTO = Ljung-Box test for excessive autocorrelation
MEAN = Test for difference in mean 1st half to 2nd half
VAR = Test for difference in variance 1st half to 2nd half
OK = not significant ($p \geq 0,05$)
* = marginally significant ($0,01 < p \leq 0,05$)
** = significant ($0,001 < p \leq 0,01$)

Obrázek 7 – Výstup Model Comparison

Alternativní postup řešení (doporučený)

Zvolíme posloupnost procedur **Relate – One Factor – Polynomial Regression ...** Za proměnnou y zadáme data o počtu dokončených bytů a za proměnnou x zadáme časovou proměnnou t . Potom v panelu *Polynomial Options* potvrdíme nabízené údaje beze změny (Order: 2, Shift: 0,0) a stiskneme OK.

Ve výstupu *Analysis Summary* (viz Obr. 8) najdeme v první tabulce výběrové parametry trendové paraboly, na jejichž základě sestavíme rovnici výběrové trendové paraboly:

$\hat{T}_t = 49510,6 - 5709,5t + 351,205t^2$. V této tabulce jsou i výsledky individuálních t-testů o nulové hodnotě trendových parametrů. V nulové hypotéze vždy testujeme, že parametr trendové paraboly není statisticky významný, tj. nemá přínos pro danou trendovou funkci.

Polynomial Regression - Pocet_dok_bytu versus t

Dependent variable: Pocet_dok_bytu

Independent variable: t

Order of polynomial = 2

Number of observations: 12

		Standard	T	
Parameter	Estimate	Error	Statistic	P-Value
CONSTANT	49510,6	2501,16	19,7951	0,0000
t	-5709,5	884,602	-6,45432	0,0001
t^2	351,205	66,2417	5,30187	0,0005

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	3,5172E8	2	1,7586E8	30,03	0,0001
Residual	5,27083E7	9	5,85647E6		
Total (Corr.)	4,04428E8	11			

R-squared = 86,9672 percent

R-squared (adjusted for d.f.) = 84,071 percent

Standard Error of Est. = 2420,02

Mean absolute error = 1821,7

Durbin-Watson statistic = 1,4409 (P=0,0299)

Lag 1 residual autocorrelation = 0,170794

Obrázek 8 – Výstup Analysis Summary v Polynomial Regression

t-testy

$$H_0: \alpha_0 = 0$$

$$H_1: \alpha_0 \neq 0$$

$$P\text{-Value} = 0,0000$$

$P\text{-Value} < \alpha$, proto zamítáme H_0 a přijímáme H_1 .

$$H_0: \alpha_1 = 0$$

$$H_1: \alpha_1 \neq 0$$

$$P\text{-Value} = 0,0001$$

$P\text{-Value} < \alpha$, proto zamítáme H_0 a přijímáme H_1 .

$$H_0: \alpha_2 = 0$$

$$H_1: \alpha_2 \neq 0$$

$$P\text{-Value} = 0,0005$$

$P\text{-Value} < \alpha$, proto zamítáme H_0 a přijímáme H_1 .

Souhrnně můžeme říci, že všechny t-testy jsou významné, což znamená, že všechny parametry trendové paraboly jsou přínosné pro tuto funkci.

Podívejme se nyní na další tabulku. V tabulce nadepsané Analysis of Variance najdeme výsledek celkového F-testu, jak už víme z regresní analýzy.

$$H_0: \alpha_0 = c, \alpha_1 = \alpha_2 = 0$$

$$H_1: \text{non } H_0$$

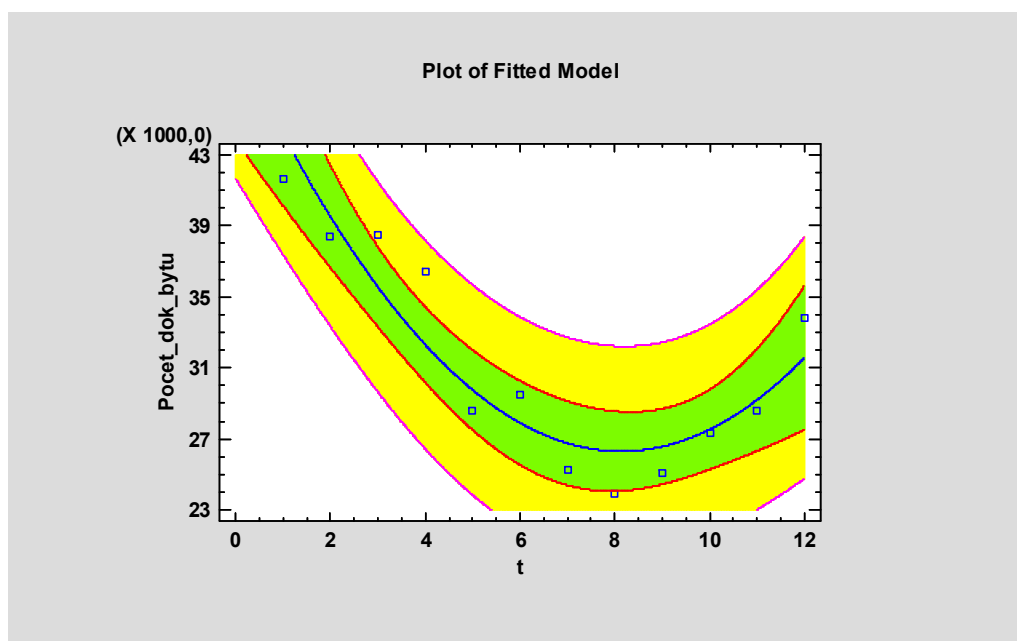
P-Value = 0,0001

P-Value < α , proto zamítáme H_0 a přijímáme H_1 .

Na hladině významnosti 5 % bylo prokázáno, že parabola je vhodná k vyrovnání hodnot této časové řady.

Tím, že jsme zde prokázali, že je parabola vhodná k vyrovnání hodnot sledované časové řady, práce ještě nekončí. Je opět potřeba tuto informaci doplnit, do jaké míry je parabola k tomuto účelu vhodná. K zodpovězení této otázky můžeme použít index determinace, který najdeme rovněž ve výstupu *Analysis Summary* pod tabulkou Analysis of Variance (R-Squared). Hodnota indexu determinace, zaokrouhlená na 3 desetinná místa, je zde 0,870 (ve výstupu je uveden v procentech: 86,9672 %). Znamená to, že 87 % variability ukazatele počtu dokončených bytů je možné vysvětlit pomocí této trendové funkce. Je to opravdu výstižný model.

Grafický výstup pak nabízí pravá část obrazovky a výstup nazvaný Plot of Fitted Model – viz Obr. 9.



Obrázek 9 – Graf trendové paraboly

Pokud bychom měli pochybnost, zda jsme opravdu vybrali nejlepší možný model, můžeme samozřejmě využít i posloupnost procedur **Relate – One Factor – Simple Regression ...**, kde v nabídce *Tables and Graphs* vybereme položku *Comparison of Alternative Models* a tam se orientačně podíváme na hodnoty indexů determinace pro všechny modely, které tato procedura obsahuje. Pokud bychom tam našli hodnoty výrazně vyšší než je ta u paraboly, mohli bychom se podívat na daný model blíže a rozhodnout, zda by nebyl vhodnější než parabola.