

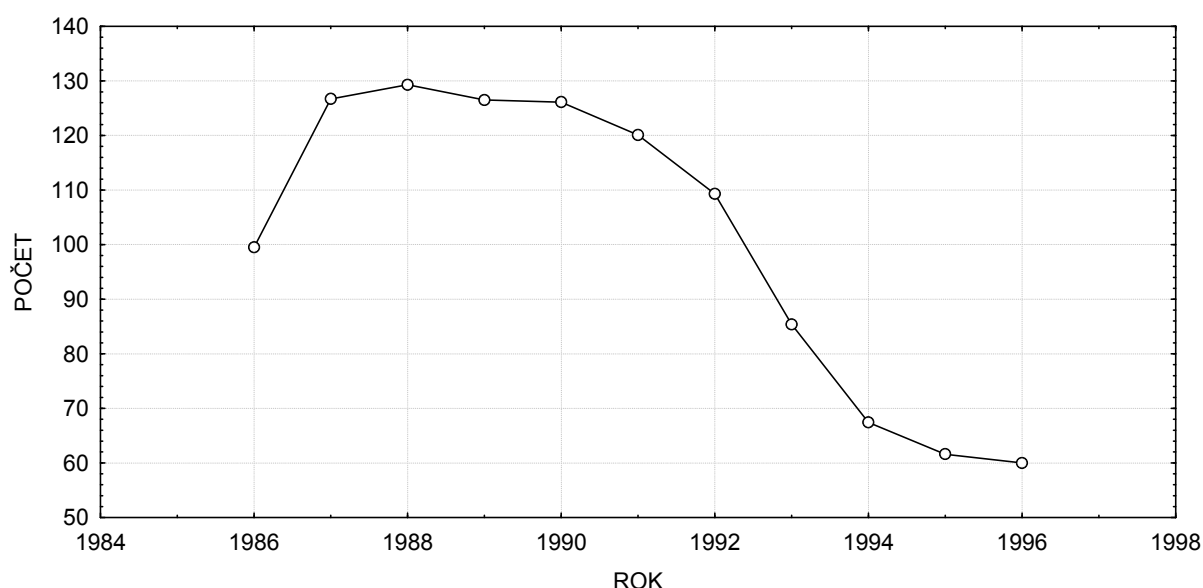
Téma č. 10: Regresní analýza v časových řadách

Příklad 1.: 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, proveďte celkový F-test a dílčí t-testy. Ověřte normalitu reziduí a sestrojte 95% intervaly spolehlivosti pro parametry trendové funkce. Stanovte střední absolutní procentuální chybu predikce (MAPE). Graficky znázorněte průběh časové řady s odhadnutým trendem, 95% pásem spolehlivosti a 95% predikčním pásem.

Řešení:

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

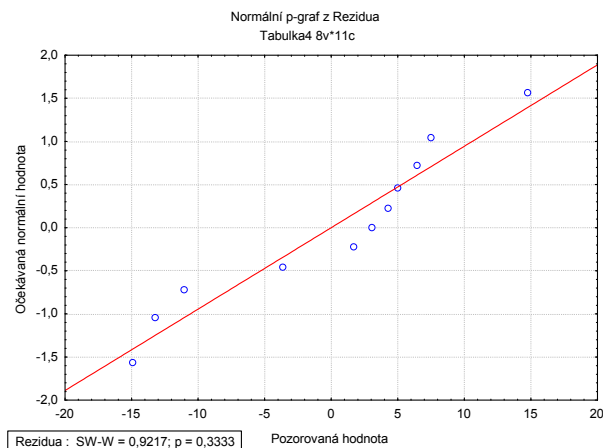
Odhadnutá trendová funkce má tedy tvar:

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

Index determinace je 0,884, tedy 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:

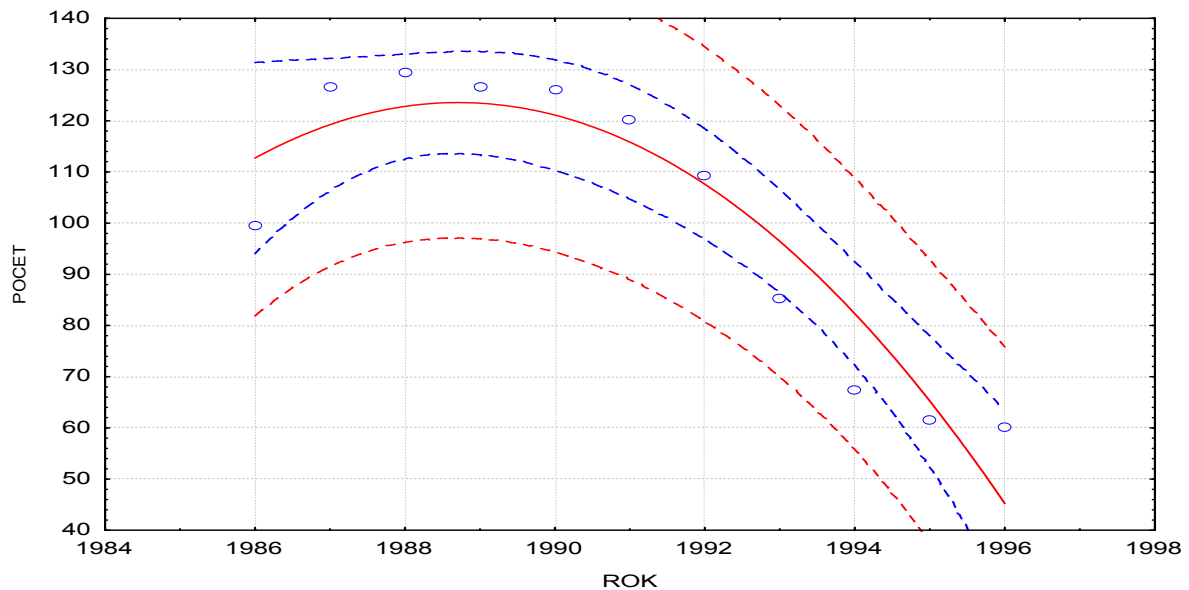
Ve výsledcích Vícenásobné regrese zvolíme záložku Rezidua / předpoklady / předpovědi – Reziduální analýza – Uložit – Uložit rezidua a předpovědi – Vybrat vše – OK. Ve vzniklé tabulce odstraníme proměnné 7 – 12, přidáme proměnnou chyby a do jejího Dlouhého jména napíšeme $=100*abs(v_6/v_2)$. Pak spočteme 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.



Příklad 2.: Máte 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.

Řešení:

Načteme datový soubor byty.sta o dvou proměnných ROK a POCET a jedenácti případech. Statistika – 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.

