

Univerzita Pardubice

Fakulta chemicko-technologická, Katedra analytické chemie



Licenční studium GALILEO – Interaktivní statistická analýza dat

Semestrální práce z předmětu

„Tvorba nelineárních regresních modelů v analýze dat“

Vypracoval

Ing. Ladislav Menšík, Ph.D.

Ústav ekologie lesa
Mendelova univerzita v Brně
Lesnická a dřevařská fakulta



Brno

červen, 2016

Obsah

Úvod	3
1. Porovnání růstových modelů a nalezení nejvhodnějšího modelu pro zadaná data.	4
2. Nalezení parametrů Michajlovy růstové funkce pro data dřeviny buk (BK) z mladého smíšeného porostu z oblasti Dražanské vrchoviny	10
Použitá literatura	14

Úvod

Při lesnickém výzkumu dochází k pořizování různých typů výzkumných dat s velkými výběry ($n > 30$) i data velmi malých výběrů ($n > 10$; $n > 4$). Pro hodnocení musí být použity různé korektní postupy a metody statistické analýzy dat včetně různých typů software.

1. Porovnání růstových modelů a nalezení nejvhodnějšího modelu pro zadaná data.

Zadání

V rámci řešení projektu COST LD14018 - Udržitelné hospodaření smíšených lesů vrchovin „Toky látek a biogeochemické koloběhy živin“ v roce 2014 byla změřena výčetní tloušťka ($d_{1,3}$) a výška (h) stromů ve smíšeném porostu modřínu s bukem ve věku cca 33 let v oblasti Dražanské vrchoviny.

Cílem úlohy je ověřit použitelnost některých známých růstových funkcí pro modelování základních taxačních veličin výčetní tloušťky a výšky.

Data - výčetní tloušťky ($d_{1,3}$) a výšky stromů (h) dřeviny buk (BK) v cm a m

Tab. 1: Vstupní data - výčetní tloušťky ($d_{1,3}$) a výšky (h) stromů dřeviny buku (BK) v cm a m

Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)
1,40	4,00	4,85	8,40	8,35	11,20	12,35	16,60	16,25	19,40
1,50	4,50	5,15	8,60	8,80	13,20	12,50	14,20	17,25	16,30
1,85	4,00	5,30	8,50	9,25	13,50	12,90	16,30	17,50	19,00
2,05	2,80	5,80	8,10	9,40	12,30	13,10	11,50	17,95	16,50
2,75	5,50	5,90	9,00	9,50	18,20	13,10	16,20	18,30	15,40
2,80	7,20	7,20	11,30	9,70	14,90	13,20	17,30	18,35	20,00
3,50	4,20	7,25	13,10	10,00	17,10	14,00	16,30	21,40	19,10
3,65	9,00	7,50	10,20	10,25	16,40	14,35	16,10		
4,00	9,00	7,70	12,60	10,85	18,60	14,40	15,70		
4,50	8,50	8,00	11,50	10,95	14,80	15,75	15,60		

Užitý program

- ADSTAT
- QC Expert

Řešení

Úloha byla řešena v programu ADSTAT v modulu „Růstové křivky“. Paralelně také pomocí SW QCExpert. Jako vstupní odhady parametrů v QCE byly použity hodnoty nalezené v ADSTAT.

Modul „Růstové křivky“ obsahuje následující růstové modely:

- Schnuteho model /fce: $Y = p[1] * [1 + p[4] * \exp(-p[3] * (x - p[2]))]^{-1/p[4]}$; pro $p[4] > 1/$
- Mitscherlichův model /fce: $Y = p[1] * [1 + p[4] * \exp(-p[3] * (x - p[2]))]^{-1/p[4]}$; pro $p[4] = -1/$
- Richardsův model /Y = $p[1] * [1 + p[4] * \exp(-p[3] * (x - p[2]))]^{-1/p[4]}$; pro $p[4] < 1/$
- Gompertzův model /Y = $p[1] * \exp(-\exp(-p[2] * (x - p[3])))$
- Logistický model /Y = $p[1] * [1 + p[4] * \exp(-p[3] * (x - p[2]))]^{-1/p[4]}$; pro $p[4] = 1/$

Odhady parametrů ADSTAT:

Schnuteho model:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	1.7717E+01	9.2951E-01	1.9926E-01	1.1246E+00
p[2]	4.9529E+00	1.8649E+00	-1.7516E-01	-3.5366E+00

p[3] 3.0892E-01 1.4717E-01 4.3452E-02 1.4066E+01
 p[4] 7.7237E-01 1.3797E+00 3.3812E-01 4.3777E+01

Mitscherlich–v model:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	1.7476E+01	6.8605E-01	5.6351E-02	3.2245E-01
p[2]	2.1867E+00	1.6061E-01	-9.3814E-03	-4.2902E-01
p[3]	2.4336E-01	3.0466E-02	1.2172E-03	5.0014E-01

Richards–v model:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	1.7717E+01	9.2951E-01	1.9926E-01	1.1246E+00
p[2]	4.9529E+00	1.8649E+00	-1.7516E-01	-3.5366E+00
p[3]	3.0892E-01	1.4717E-01	4.3452E-02	1.4066E+01
p[4]	7.7237E-01	1.3797E+00	3.3812E-01	4.3777E+01

Gompertz–v model:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	1.8277E+01	8.2293E-01	9.0868E-02	4.9716E-01
p[2]	2.2738E-01	3.3489E-02	2.0191E-03	8.8797E-01
p[3]	3.6666E+00	3.4235E-01	2.4998E-02	6.8177E-01

Logistický model:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	1.7609E+01	6.2129E-01	5.2860E-02	3.0018E-01
p[2]	5.2363E+00	3.7537E-01	2.2983E-02	4.3892E-01
p[3]	3.3287E-01	4.3020E-02	3.2065E-03	9.6329E-01

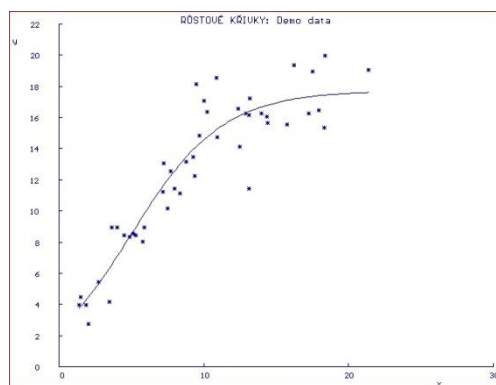
Základní statistické charakteristiky ADSTAT:

Model	RSC	Regresní rabat (D^2)
Schnuteho model	1.2753E+02	8.8031E+01
Mitscherlichův model	1.7360E+02	8.3708E+01
Richardsův model	1.2753E+02	8.8031E+01
Gompertzův model	1.2830E+02	8.7959E+01
Logistický model	1.2759E+02	8.8026E+01

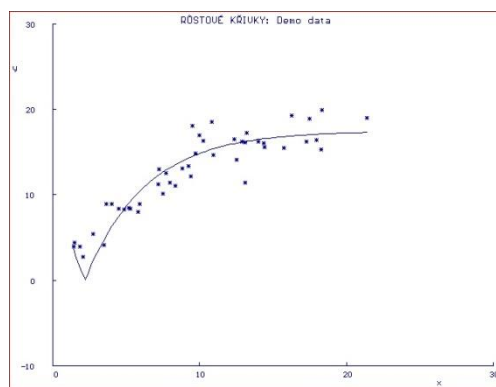
Závěr:

Jako nejlepší model pro data buku (BK) naměřená ve smíšeném porostu v oblasti Dražanské vrchoviny se jeví Richardsův růstový model (Schnuteho model).

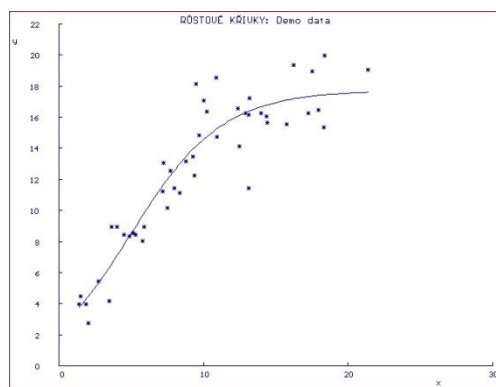
Grafické výstupy ADSTAT:



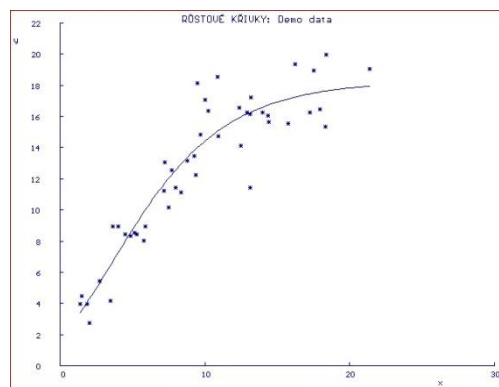
Regresní křivka - Schnuteho model



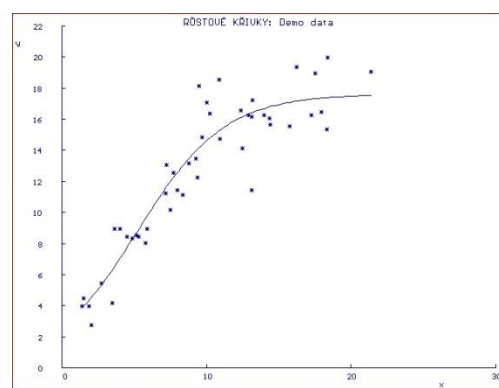
Regresní křivka - Mitscherlichův model



Regresní křivka - Richardsův model



Regresní křivka - Gompertzův model



Regresní křivka - Logistický model

Odhady parametrů QC Expert:

Richards–v model:

$$\text{Model : } [M801y] \sim P1 * (1 + P4 * \exp(-P3 * ([M801x] - P2)))^{(-1/P4)}$$

Počáteční hodnoty parametrů :

- P1 20
- P2 3
- P3 1
- P4 1

Odhady parametrů	Parametr	Směr. odchylka	Dolní mez	Horní mez
P1	17,74773568	0,9662045077	15,79919859	19,69627278
P2	4,973162422	1,968450833	1,003402983	8,94292186
P3	0,305781573	0,1517121589	-0,000175154	0,6117383003
P4	0,8214983457	1,468074637	-2,139156323	3,782153015

Statistické charakteristiky regrese

- Vícenásobný korel. koef. R : 0,9354487431
- Koeficient determinace R² : 0,875064351
- Predikovaný korel. koef. Rp : 0,8535887264
- Stř. kvadratická chyba predikce MEP : 3,25345999
- Akaikeho informační kritérium : 55,99161701

- Reziduální součet čtverců : 130,4833766
- Průměr absolutních reziduí : 60,60187229
- Reziduální směr. odchylka : 1,741980806
- Reziduální rozptyl : 3,03449713

Šikmost reziduí : 0,001907726898
Špičatost reziduí : 3,502899048

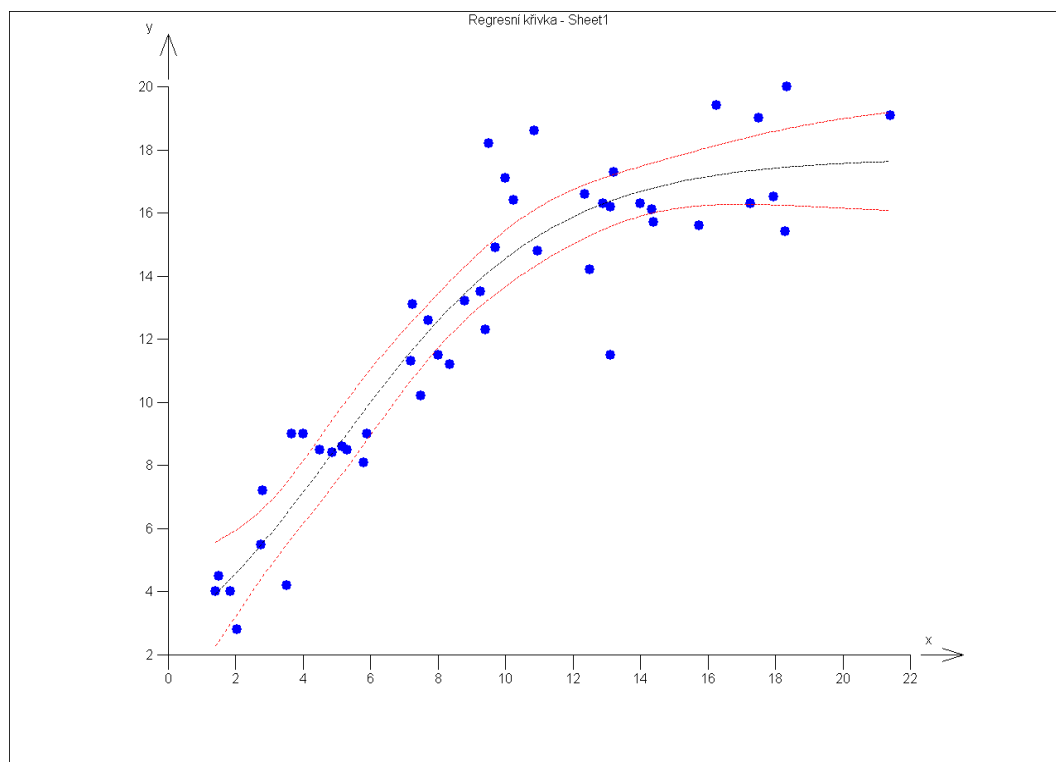
Cook-Weisbergův test heteroskedasticity
Hodnota kritéria CW : 1,638931209
Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
Pravděpodobnost : 0,2004721737
Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality
Hodnota kritéria JB : 0,5102209542
Kvantil $\chi^2(1-\alpha,2)$: 5,991464547
Pravděpodobnost : 0,7748308922
Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace
Hodnota kritéria WA : 0,01961685243
Kvantil $\chi^2(1-\alpha,1)$: 3,841458829
Pravděpodobnost : 0,2004721737
Závěr : Autokorelace je nevýznamná

Znaménkový test reziduí
Hodnota kritéria Sg : 0,2678288261
Kvantil $N(1-\alpha/2)$: 1,959963999
Pravděpodobnost : 0,7888310811
Závěr : V reziduích není trend.

Grafické výstupy QC Expert:



Regresní křivka - Richardsův model

Základní statistické charakteristiky QC Expert:

Parametr	Schnuteho model	Mitscherlichův model	Richardsův model	Gompertzův model	Logistický model
R	0,9321955798	-	0,9354487431	0,9340443246	0,9340443246
R²	0,8689885989	-	0,875064351	0,8724388003	0,8724388003
R_p	0,8451646296	-	0,8535887264	0,849849317	0,849849317
MEP	3,440655013	-	3,25345999	3,336554812	3,336554812
AIC	58,22343806	-	55,99161701	56,9690972	56,9690972

Pozn.: R – vícenásobný korelační koeficient, R² - koeficient determinace, R_p – predikovaný korelační koeficient, MEP - střední kvadratická chyba predikce, AIC - akaikeho informační kritérium

Závěr:

Nejlepší model pro data buku (BK) naměřená ve smíšeném porostu v oblasti Dražanské vrchoviny se jeví Richardsův růstový model (Schnuteho model).

2. Nalezení parametrů Michajlovy růstové funkce pro data dřeviny buk (BK) z mladého smíšeného porostu z oblasti Dražanské vrchoviny

Zadání

V rámci řešení projektu COST LD14018 - Udržitelné hospodaření smíšených lesů vrchovin „Toky látek a biogeochemické koloběhy živin“ v roce 2014 byla změřena výčetní tloušťka ($d_{1,3}$) a výška (h) stromů ve smíšeném porostu modřínu s bukem ve věku cca 33 let v oblasti Dražanské vrchoviny.

Cílem úlohy je nalezení (zpřesnění) parametrů Michajlovy růstové funkce pro dřevinu buk (BK) v mladém smíšeném porostu z oblasti Dražanské vrchoviny pro modelování základních taxačních veličin výčetní tloušťky a výšky.

Data - výčetní tloušťky ($d_{1,3}$) a výšky stromů (h) dřeviny buku (BK) v cm a m

Tab. 1: Vstupní data - výčetní tloušťky ($d_{1,3}$) a výšky (h) stromů dřeviny buku (BK) v cm a m

Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)	Tloušťka (cm)	Výška (m)
1,40	4,00	4,85	8,40	8,35	11,20	12,35	16,60	16,25	19,40
1,50	4,50	5,15	8,60	8,80	13,20	12,50	14,20	17,25	16,30
1,85	4,00	5,30	8,50	9,25	13,50	12,90	16,30	17,50	19,00
2,05	2,80	5,80	8,10	9,40	12,30	13,10	11,50	17,95	16,50
2,75	5,50	5,90	9,00	9,50	18,20	13,10	16,20	18,30	15,40
2,80	7,20	7,20	11,30	9,70	14,90	13,20	17,30	18,35	20,00
3,50	4,20	7,25	13,10	10,00	17,10	14,00	16,30	21,40	19,10
3,65	9,00	7,50	10,20	10,25	16,40	14,35	16,10		
4,00	9,00	7,70	12,60	10,85	18,60	14,40	15,70		
4,50	8,50	8,00	11,50	10,95	14,80	15,75	15,60		

Užitý program

- ADSTAT
- QC Expert

Řešení

Základní úloha byla řešena v programu ADSTAT v modulu „Nelineární regrese“. Upravená rovnice Michajlovy růstové funkce byla hledána pomocí SW QCExpert v modulu „Nelineární regrese“.

Standardní Michajlova růstová funkce: /fce: $Y = (p1*x)^{-p2/x} + 1.3/$

Odhady parametrů ADSTAT:

BODOVÉ ODHADY PARAMETRŮ:

Parametr	Bodový odhad	Směrodatná odchylka	Absolutní vychýlení	Relativní vychýlení
p[1]	2.3131E-01	1.0018E-02	2.9542E-04	1.2772E-01
p[2]	-3.2447E+01	1.5691E+00	-3.7526E-02	1.1565E-01

STATISTICKÉ CHARAKTERISTIKY REGRESE:

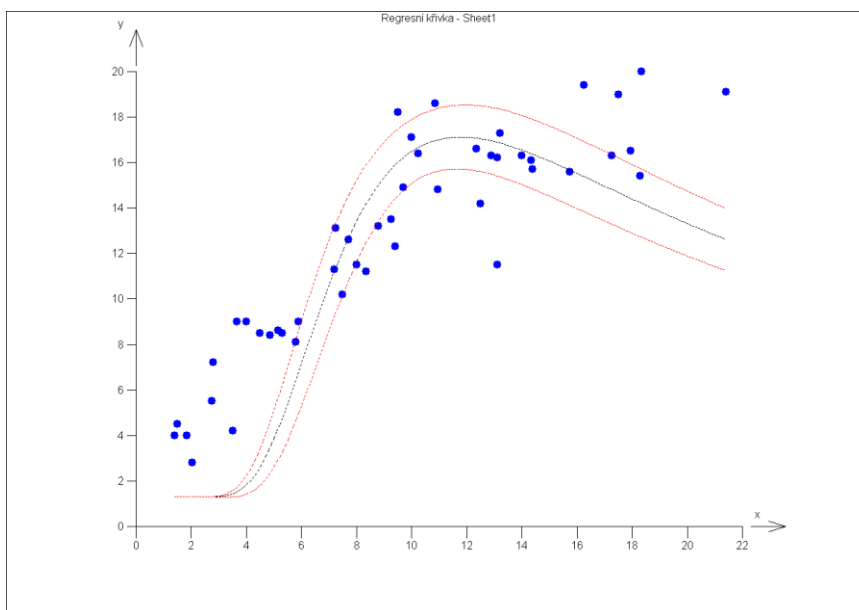
Reziduální součet čtverců, RSC	: 4.8865E+02
Regresní rabat, D^2 [%]	: 5.4141E+01
Akaikeho informační kritérium, AIC	: 1.1405E+02

Odhady parametrů QC Expert:

Odhady parametrů	Parametr	Směr.	odchylka	Dolní mez	Horní mez
P1	0,2310673638	0,01018282196	0,2105581076	0,2515766201	
P2	-32,48395953	1,599446696	-35,70541054	-29,26250852	

Statistické charakteristiky regrese

Vícenásobný korel. koef. R : 0,7181372015
 Koeficient determinace R² : 0,5157210402
 Predikovaný korel. koef. Rp : 0,4869418413
 Stř. kvadratická chyba predikce MEP : 11,40085835
 Akaikeho informační kritérium : 115,6701455



Regresní křivka - standardní Michajlova růstová funkce: /fce: $Y = (p1*x)^{-p2/x} + 1.3/$

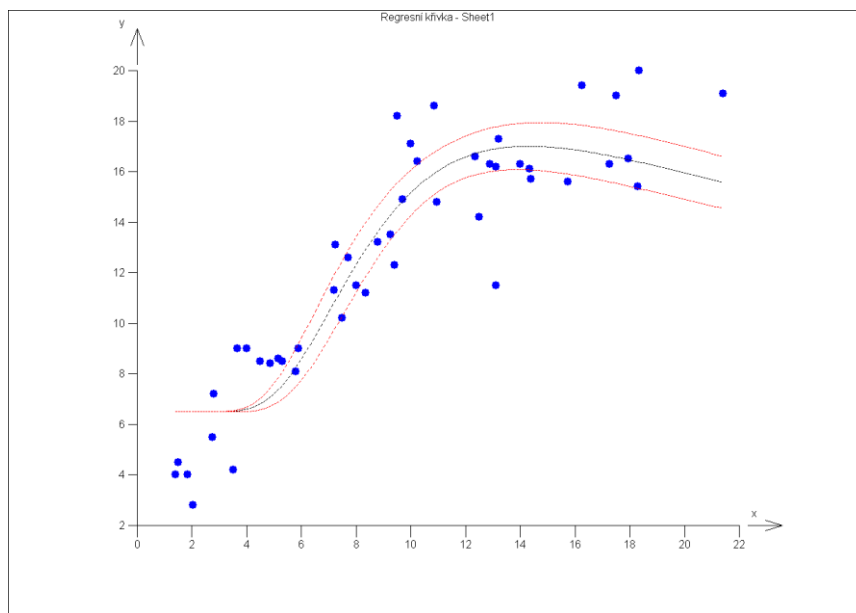
Základní statistické charakteristiky upravené rovnice Michajlovy růstové funkce - QC Expert:

Parametr	$Y = (p1*x)^{-p2/x} + 1.3$	$Y = (p1*x)^{-p2/x} + 6.0$	$Y = (p1*x)^{-p2/x} + 6.4$	$Y = (p1*x)^{-p2/x} + 6.5$	$Y = (p1*x)^{-p2/x} + 6.6$
R	0,7181372015	0,9087643568	0,9100583061	0,9100471632	0,9099015746
R2	0,5157210402	0,8258526561	0,8282061205	0,8281858392	0,8279208755
Rp	0,4869418413	0,8105047466	0,813251469	0,8133153599	0,8131291334
MEP	11,40085835	4,210845312	4,14980936	4,148389616	4,152527826
AIC	115,6701455	67,60045559	66,96095659	66,96650489	67,03893026

Pozn.: R – vícenásobný korelační koeficient, R² - koeficient determinace, Rp – predikovaný korelační koeficient, MEP - střední kvadratická chyba predikce, AIC - akaikeho informační kritérium

Závěr:

Jako nejlepší upravená rovnice Michajlovy růstové funkce pro data buku (BK) naměřená ve smíšeném porostu v oblasti Dražanské vrchoviny se jeví rovnice $Y = (p1*x)^{-p2/x} + 6.5$.



Upravená Michajlova růstová funkce: /fce: $Y = (p1*x)^{-p2/x}+6.5/$

Odhady parametrů	Parametr	Směr. odchylka	Dolní mez	Horní mez
P1	0,1900490335	0,007812898335	0,1743130484	0,2057850185
P2	-33,63690733	1,718278257	-37,09769739	-30,17611727

Statistické charakteristiky regrese

Vícenásobný korel. koef. R : 0,9100471632
 Koeficient determinace R² : 0,8281858392
 Predikovaný korel. koef. Rp : 0,8133153599
 Stř. kvadratická chyba predikce MEP : 4,148389616
 Akaikeho informační kritérium : 66,96650489

Reziduální součet čtverců : 179,4435137
 Průměr absolutních reziduí : 74,01887823
 Reziduální směr. odchylka : 1,996906016
 Reziduální rozptyl : 3,987633639
 Šikmost reziduí : 0,02703781163
 Špičatost reziduí : 2,918665014

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW : 0,0979730728
 Kvantil Chi²(1-alfa,1) : 3,841458829
 Pravděpodobnost : 0,7542757193
 Závěr : Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB : 0,2247513102
 Kvantil Chi²(1-alfa,2) : 5,991464547
 Pravděpodobnost : 0,8937084683
 Závěr : Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA : 2,43528132

Kvantil $\chi^2(1-\alpha,1)$: 3,841458829

Pravděpodobnost : 0,7542757193

Závěr : Autokorelace je nevýznamná

Znaménkový test reziduí

Hodnota kritéria Sg : 1,748764688

Kvantil $N(1-\alpha/2)$: 1,959963999

Pravděpodobnost : 0,0803317031

Závěr : V reziduích není trend.

Použitá literatura

Meloun, M., Militký, J. 2012: Interaktivní statistická analýza dat. 4. vyd. Praha: Karolinum Praha. 955 s. ISBN 978-80-246-2173-9.

Meloun, M., Militký, J. 2012: Kompendium statistického zpracování dat. 3. vyd. Praha: Karolinum Praha. 985 s. ISBN 978-80-246-2196-8.