

UNIVERZITA PARDUBICE
Fakulta chemicko-technologická
Katedra analytické chemie

TVORBA LINEÁRNÍCH REGRESNÍCH MODELŮ PŘI ANALÝZE DAT

Semestrální práce

Licenční studium Galileo – Interaktivní statistická analýza dat

Brno 2015

Ing. Petra Hlaváčková, Ph.D.
Mendelova univerzita v Brně

Obsah

1	Úloha 1 Porovnání dvou regresních přímek u jednoduchého lineárního regresního modelu.....	4
1.1	Zadání a vstupní data	4
1.2	Návrh modelu.....	4
1.3	Odhadování parametrů	4
1.4	Základní statistické charakteristiky	5
1.5	Regresní diagnostika	5
1.5.1	Kritika dat (model 1).....	5
1.5.2	Kritika dat (model 2).....	6
1.5.3	Kritika modelu	7
1.5.4	Kritika metody	8
1.6	Konstrukce zpřesněného modelu.....	9
1.7	Porovnání regresních modelů	10
1.8	Závěr.....	12
2	Úloha 2 Určení stupně polygonu	13
2.1	Zadání a vstupní data	13
2.2	Návrh modelu.....	13
2.3	Odhadování parametrů	15
2.4	Regresní diagnostika	15
2.4.1	Kritika metody	16
2.5	Závěr.....	17
3	Úloha 3 Validace nové analytické metody	18
3.1	Zadání a vstupní data	18
3.2	Návrh modelu.....	18
3.3	Předběžná analýza dat	19
3.4	Odhadování parametrů	19
3.5	Základní statistické charakteristiky	19
3.6	Regresní diagnostika	19
3.6.1	Kritika dat	19
3.6.2	Kritika modelu	21
3.6.3	Kritika metody	21
3.7	Konstrukce zpřesněného modelu.....	22
3.8	Zhodnocení kvality modelu	23
3.9	Závěr.....	23
4	Úloha 4 Vícerozměrný lineární regresní model.....	24
4.1	Zadání a vstupní data	24
4.2	Návrh modelu.....	24
4.3	Předběžná analýza dat	24

4.4	Odhadování parametrů	25
4.5	Základní statistické charakteristiky	25
4.6	Regresní diagnostika	26
4.6.1	Kritika dat	26
4.6.2	Kritika modelu	27
4.6.3	Kritika metody	28
4.7	Zhodnocení kvality modelu	29
4.8	Závěr	29

1 Úloha 1 Porovnání dvou regresních přímek u jednoduchého lineárního regresního modelu

1.1 Zadání a vstupní data

U deseti odvozních souprav byla zjištěna spotřeba paliva, množství dopraveného dříví a počet ujetých kilometrů. Cílem je zjistit, zda a jak velká existuje závislost mezi těmito veličinami. Závislost spotřeby paliva na dopraveném množství a ujetých kilometrech bude popsána regresními křivkami. Vstupní data analýzy jsou vedena v tabulce 1.1. Použitým programem byl QC Expert.

Tab. 1.1 Vstupní data analýzy

Spotřeba paliva (l)	Množství dopraveného dříví (m ³)	Počet ujetých kilometrů
25 436	18 299	35 523
59 134	31 419	107 940
88 026	27 144	170 313
175 521	94 219	288 034
176 692	99 439	287 518
473 217	233 246	784 100
817 000	336 950	1 427 102
1 330 002	517 241	2 368 490
1 814 832	764 622	3 229 222
1 859 244	694 312	3 412 844

1.2 Návrh modelu

V případě spotřeby paliva odvozních souprav byl navržen přímkový regresní model pro závislost na množství dopraveného dříví, resp. počtu ujetých kilometrů. Rovnice pro navržené modely je následující:

$$y = \beta_0 + \beta_1 x_1$$

Nejprve bude testován regresní triplet zvlášť pro oba modely. Poté budou modely porovnány a bude testována statistická hypotéza H_0 o shodě modelů:

$$H_0: \beta_A = \beta_B$$

Regresní parametry pro model 1 jsou:

$$\beta_A = (\beta_{1,A}, \beta_{0,A})$$

Regresní parametry pro model 2 jsou:

$$\beta_B = (\beta_{1,B}, \beta_{0,B})$$

1.3 Odhadování parametrů

Odhad parametrů byl proveden klasickou metodou nejmenších čtverců (MNČ), jejíž pomocí byly stanoveny regresní parametry úsek β_0 a směrnice β_1 . Analýza probíhala na hladině významnosti alfa rovno 0,05. S využitím Studentova t-testu bude zjištěno, zda lze úseky (absolutní členy) a směrnice u modelů považovat za statisticky významné či nikoli. Odhady parametrů pro model 1 jsou uvedeny v tabulce 1.2, pro model 2 v tabulce 1.3.

Tab. 1.2 Odhady regresních parametrů modelu 1

Proměnná	Odhad	Směr. odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
y	-40182,14711	32604,75822	Nevýznamný	0,25278765	-115368,8544	35004,56016
x	2,563438014	0,083546557	Významný	1,38304E-09	2,370779307	2,756096721

Tab. 1.3 Odhady regresních parametrů modelu 2

Proměnná	Odhad	Směr. odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
y	13824,28155	9203,821918	Nevýznamný	0,171493	-7399,76985	35048,33296
x	0,551631884	0,005266287	Významný	7,70495E-14	0,539487805	0,563775963

Z tabulek 1.2 a 1.3 vyplývá, že s využitím Studentova t-testu bylo zjištěno, že absolutní člen (úsek) lze považovat za statisticky nevýznamný u obou modelů, zatímco směrnice je statisticky významná také u obou modelů, jelikož t je větší než kvantil Studentova rozdělení $t_{krit} = 2,3060$.

1.4 Základní statistické charakteristiky

Základní statistické charakteristiky obou modelů uvádí tabulka 1.4.

Tab. 1.4 Základní statistické charakteristiky modelů 1 a 2

Statistické charakteristiky regrese	Model 1	Model 2
Vícenásobný korelační koeficient (R)	0,995778038	0,999635638
Koeficient determinace (R^2)	0,991573901	0,999271409
Predikovaný korelační koeficient (R_p)	0,964653600	0,997017513
Střední kvadratická chyba predikce (MEP)	8597388293	721500164,6
Akaikeho informační kritérium (AIC)	225,2782012	200,7984328

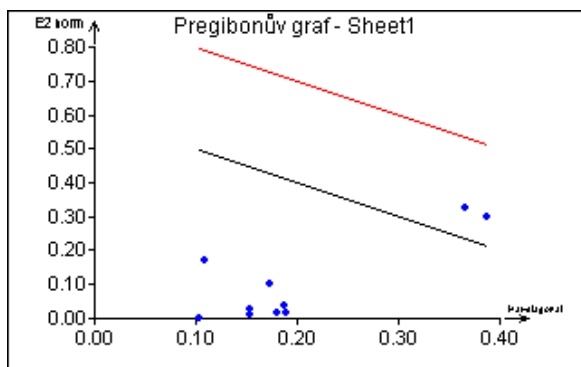
U obou modelů byla zjištěna vysoká hodnota korelačního koeficientu, což značí významnost modelu. Koeficient determinace je opět u obou modelů velmi vysoký, to znamená, že většina experimentálních bodů vyhovuje navrhovaným modelům.

1.5 Regresní diagnostika

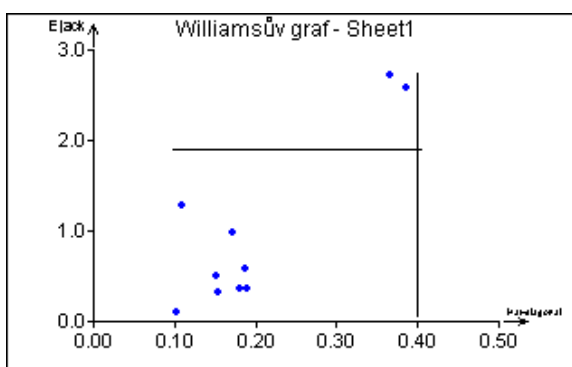
Regresní diagnostiku tvoří tzv. regresní triplet – kritika dat, modelu a metody. Pro modely 1 a 2 bude provedena zvlášť.

1.5.1 Kritika dat (model 1)

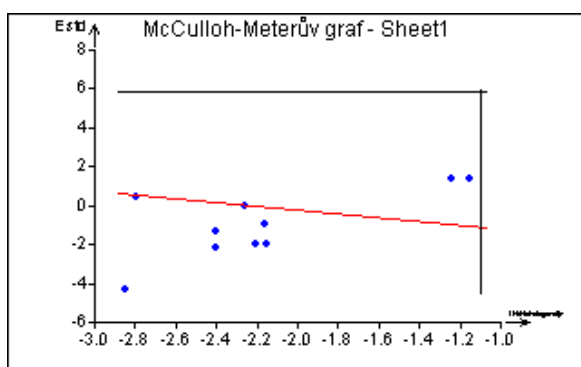
V této části regresního tripletu budou identifikovány vlivné body, které zahrnují odlehlé body a extrémy. Odlehlé body by měly být vyloučeny, aby mohl být sestaven zpřesněný model. Kritika dat bude provedena na grafech, které znázorňují vlivné body. Případně je možné využít indexových a rankitových grafů. Pro model 1 jsou uvedeny vybrané grafy na obrázcích 1.1 až 1.6.



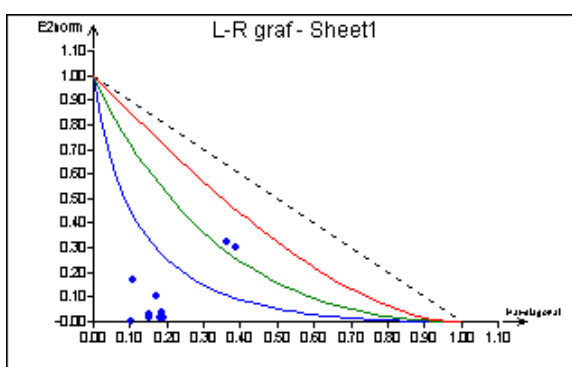
Obr. 1.1 Pregibonův graf (model 1)



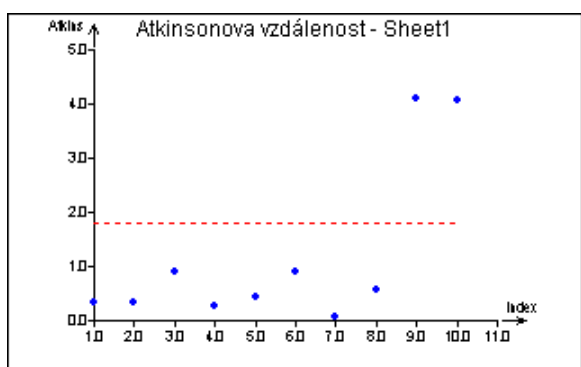
Obr. 1.2 Williamsův graf (model 1)



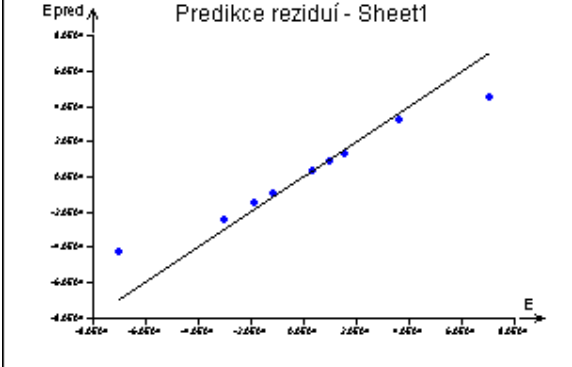
Obr. 1.3 McCullohův-Meeterův graf (model 1)



Obr. 1.4 L-R graf (model 1)



Obr. 1.5 Graf Atkinsonovy vzdálenosti (model 1)



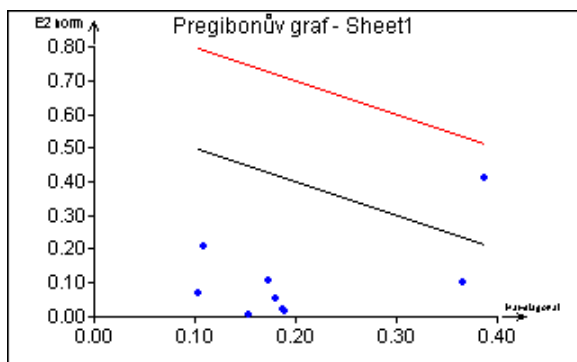
Obr. 1.6 Graf predikovaných reziduí (model 1)

Graf na obrázku 1.1 slouží k posouzení vlivných bodů. Z obrázku je zřejmé, že jsou v modelu 2 odlehle body. Jedná se o body 9 a 10. Graf na obrázku 1.2 identifikoval 2 odlehle body, žádné extrémy. Jsou to body 9 a 10. McCullohův-Meeterův graf odhalil dva podezřelé body (9, 10). L-R graf odhalil opět stejné odlehle body 9 a 10. Totéž platí i u Atkinsonovy vzdálenosti (obr. 1.5) a grafu predikovaných reziduí (obr. 1.6).

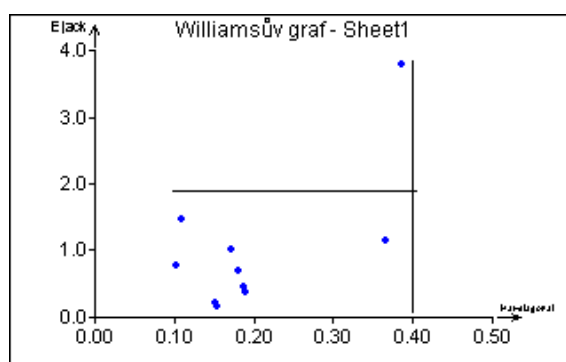
Kritika dat při tvorbě modelu 1 odhalila 2 odlehle body, toto tvrzení potvrdilo minimálně 5 grafických diagnostik. Pro zpřesnění modelu budou tedy body 9 a 10 z datového souboru vyloučeny a bude sestaven zpřesněný model 1.

1.5.2 Kritika dat (model 2)

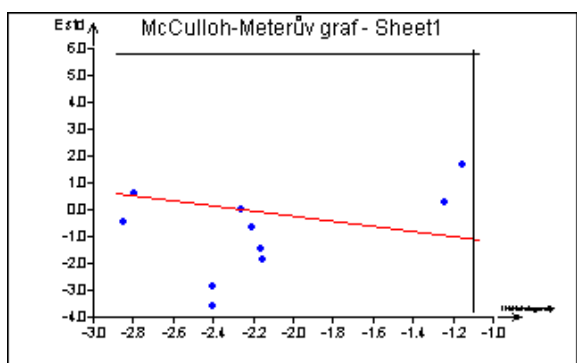
Stejně jako u modelu 1, bude provedena kritika dat pomocí grafického znázornění vlivných bodů. Pro model 2 jsou uvedeny grafy na obrázcích 1.7 až 1.12.



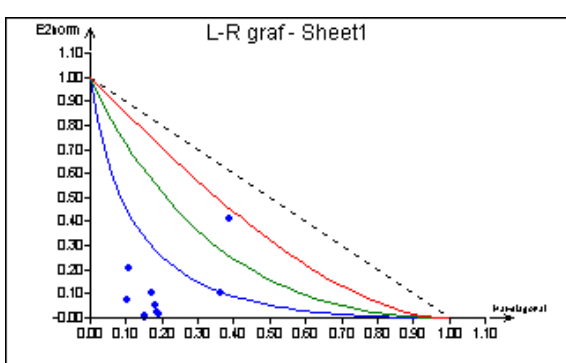
Obr. 1.7 Pregibonův graf (model 2)



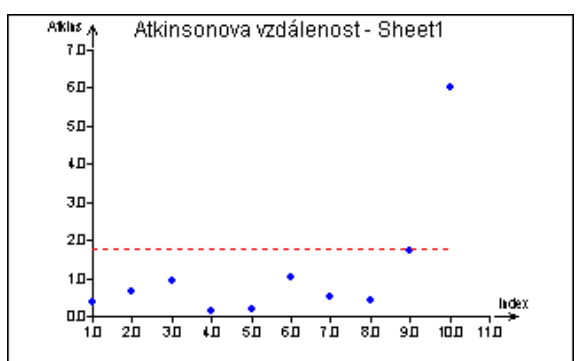
Obr. 1.8 Williamsův graf (model 2)



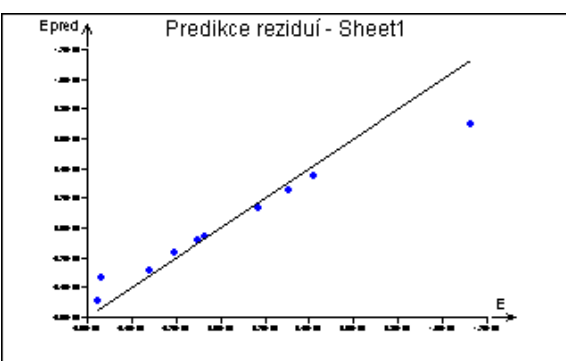
Obr. 1.9 McCullohův-Meeterův graf (model 2)



Obr. 1.10 L-R graf (model 2)



Obr. 1.11 Graf Atkinsonovy vzdálenosti (model 2)



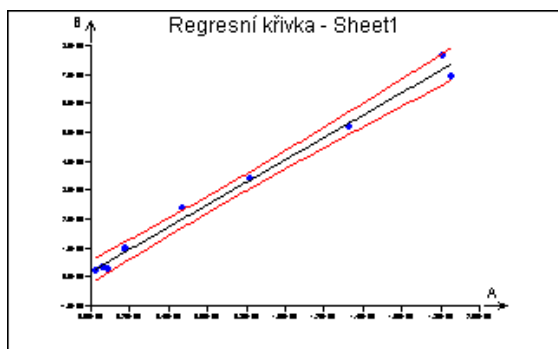
Obr. 1.12 Graf predikovaných reziduí (model 2)

V případě modelu 2 se pravděpodobně vyskytuje 1 vlivný bod (bod č. 10). Dle Pregibonova grafu (obr. 1.7) je bod č. 10 vlivným bodem. Dle Williamsova grafu (obr. 1.8) se jedná v případě bodu 10 o odlehlý bod). McCullohův graf (obr. 1.9) označuje za podezřelé body 9 a 10. U grafu L-R (obr. 1.10) je bod 10 bodem odlehlým a bod 9 je na hranici extrému. Atkinsonova vzdálenost (obr. 1.11) i graf predikovaných reziduí (obr. 1.12) znázorňují jako odlehlý bod č. 10.

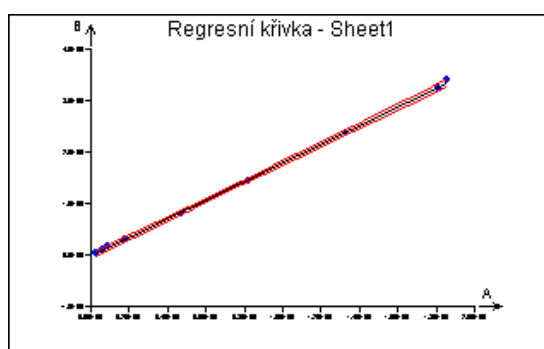
Lze tedy konstatovat, že při kritice dat modelu 2 byla prokázána odlehlost u bodu č. 10, jelikož minimálně 5 grafických charakteristik potvrdilo toto tvrzení. Při konstrukci zpřesněného modelu bude bod č. 10 vyloučen z datového souboru. Dalším vlivným bodem by mohl být bod č. 9.

1.5.3 Kritika modelu

Dle grafů na obrázcích 1.13 a 1.14 bude posouzeno, zda byla dodržena přímková závislost.



Obr. 1.13 Regresní přímka (model 1)



Obr. 1.14 Regresní přímka (model 2)

Dle obrázků 1.13 a 1.14 lze konstatovat, že lineární regresní model splňuje přímkovou závislost. Navržený model je tedy vhodný. Pás spolehlivosti predikce je u modelu 2 užší než u modelu 1.

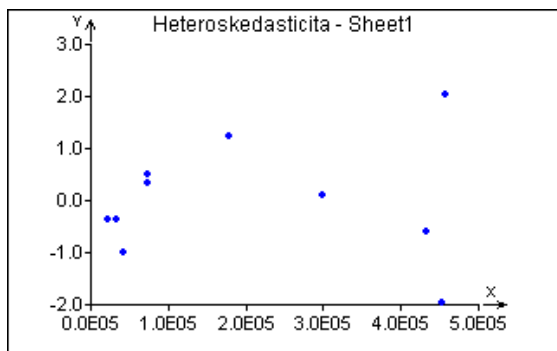
1.5.4 Kritika metody

Kritika metody představuje posouzení, zda modely splňují základní předpoklady MNČ. Tabulka 1.5 uvádí hodnoty jednotlivých kritérií v rozdělení na model 1 a 2. Dále je kritika doplněna grafickými diagnostikami – grafy heteroskedasticity pro model 1 uvádí obrázek 1.15, pro model 2 obrázek 1.16, grafy predikovaných reziduí jsou uvedeny pro model 1 na obrázku 1.17, pro model 2 na obrázku 1.18.

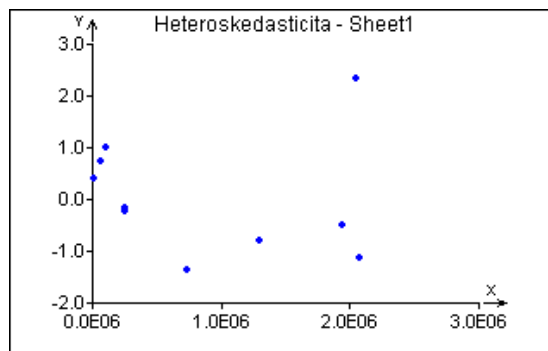
Tab. 1.5 Kritika metody modelů 1 a 2

Parametr	Model 1	Model 2
Fisher-Snedocorův test významnosti modelu, F		
hodnota kritéria F	941.4310593	10972.09929
tabulkový kvantil $F(1-\alpha, m-1, n-m)$	5.317655072	5.317655072
Pravděpodobnost	1.38304E-09	7.70761E-14
Závěr	Model je významný	Model je významný
Cook-Weisbergův test heteroskedasticity, Sf		
hodnota kritéria Sf	2.593436565	1.419147581
tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.841458829	3.841458829
Pravděpodobnost	0.107307282	0.233543758
Závěr	Rezidua vykazují homoskedasticitu	Rezidua vykazují homoskedasticitu
Jarque-Berrův test normality reziduí, L(e)		
hodnota kritéria L(e)	0.165190508	0.559053328
tabulkový kvantil, $\chi^2(1-\alpha, 2)$:	5.991464547	5.991464547
Pravděpodobnost	0.920723731	0.756141566
Závěr	Rezidua mají normální rozdělení	Rezidua mají normální rozdělení
Waldův test autokorelace, Wa		
hodnota kritéria Wa	0.848260436	0.165353637
tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.841458829	3.841458829
Pravděpodobnost	0.357044918	0.68427463
Závěr	Autokorelace je nevýznamná	Autokorelace je nevýznamná

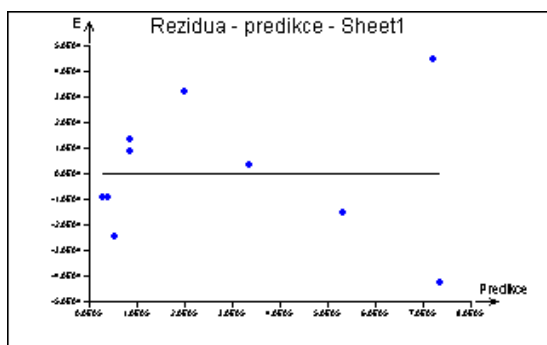
Znaménkový test reziduí, Sg		
hodnota kritéria Sg	0.335410197	1.615610525
kvantil N(1-alfa/2)	1.959963999	1.959963999
Pravděpodobnost	0.737315677	0.106178547
Závěr	V reziduích není trend	V reziduích není trend



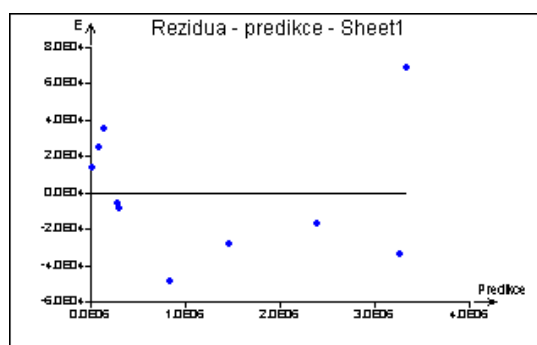
Obr. 1.15 Graf hereroskedasticity (model 1)



Obr. 1.16 Graf heteroskedasticity (model 2)



Obr. 1.17 Graf predikovaných reziduí (model 1)



Obr. 1.18 Graf predikovaných reziduí (model 2)

Kritika metody prokázala, že oba modely dle Fisher-Snedecorova testu významnosti lze považovat za významné, jelikož hodnota testovacího kritéria F významně přesáhla kritickou hodnotu tabulkového kvantilu. Podle Cook-Weisbergova testu heteroskedasticity i dle grafických diagnostik (obr. 1.15 a 1.16) je rozptyl reziduí u obou modelů konstantní, tzv. rezidua vykazují homoskedasticitu. Testem Jarque-Berrova testu normality bylo u obou modelů zjištěno normální rozdělení reziduí. Waldův test dále zjistil nevýznamnost autokorelace a znaménkový test pak určuje, zda v reziduích není trend. U obou modelů bylo zjištěno, že trend v reziduích není, což je také zřejmé z grafu predikovaných reziduí (obr. 1.17 a 1.18).

1.6 Konstrukce zpřesněného modelu

Zpřesněný model byl konstruován pro oba modely. V případě modelu 1 byly z datového souboru odstraněny 2 body. Jednalo se o body č. 9 a 10. V případě modelu 2 byl z datového souboru odstraněn bod č. 10. Zda toto rozhodnutí vedlo ke zlepšení modelu, bude zřejmé z dvou kritérií – z hodnoty střední kvadratické predikce (MEP), příp. z hodnoty Akaikeho informačního kritéria (AIC). Základní statistické charakteristiky zpřesněných modelů 1 a 2 uvádí tabulka 1.6. Upravené hodnoty odhadů regresních parametrů uvádí tabulky 1.7 a 1.8.

Tab. 1.6 Základní statistické charakteristiky zpřesněných modelů

Statistické charakteristiky regrese	Model 1	Model 2
Vícenásobný korelační koeficient (R)	0.994855660	0.999818829
Koeficient determinace (R^2)	0.989737784	0.99963769
Predikovaný korelační koeficient (R_p)	0.950084846	0.998949071
Střední kvadratická chyba predikce (MEP)	4732647619	192400118.1
Akaikeho informační kritérium (AIC)	175.0105833	172.3273733

Tab. 1.7 Odhady regresních parametrů modelu 1

Proměnná	Odhad	Směr. odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
y	-41814,58869	25439,12583	Nevýznamný	0,151340605	-104061,8872	20432,70978
x	2,562337916	0,106517578	Významný	3,39041E-07	2,301698793	2,822977039

Tab. 1.8 Odhady regresních parametrů modelu 2

Proměnná	Odhad	Směr. odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
y	9571.496839	5842.099349	Nevýznamný	0.14535532	-4242.872962	23385.86664
x	0.560310523	0.004031794	Významný	2.63567E-13	0.550776846	0.569844201

Při porovnání základních statistických charakteristik je zřejmé, že v případě modelu 1 došlo zpřesněním modelu vynecháním bodu č. 9 a 10 k požadovanému výsledku, jelikož hodnota střední kvadratické chyby i hodnota Akaikeho informačního kritéria klesla. V případě modelu 2 došlo, odstraním bodu č. 10, k zpřesnění modelu, jelikož hodnota MEP i hodnota AIC klesla.

Z pohledu kritiky metody nedošlo ke změně závěrů testů.

Dále je třeba konstatovat, že u obou modelů byly testovány dva regresní parametry, tzn., že dostatečný počet bodů by měl být u obou modelů 10. Z provedené analýzy vyplývá, že by bylo vhodnější v datových souborech pro regresi zachovat všechny body.

1.7 Porovnání regresních modelů

Pro porovnání regresních modelů byl sestaven společný model pro naměřené hodnoty, nebylo uvažováno vyloučení odlehlých bodů. Odhad regresních parametrů tohoto modelu uvádí tabulka 1.9.

Tab. 1.9 Odhad regresních parametrů společného modelu

Proměnná	Odhad	Směr. odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
y	716.9060376	4479.866932	Nevýznamný	0.877378623	-9876.295955	11310.10803
x1	0.55027232	0.090525888	Významný	0.000501552	0.336212609	0.76433203
x2	0.434467875	0.01940529	Významný	8.96959E-08	0.388581654	0.480354095

Pro společný model byla prokázána nevýznamnost směrnice a významnost směrnice. Základní statistické charakteristiky uvádí tabulka 1.10.

Tab. 1.10 Základní statistické charakteristiky společného modelu

Statistické charakteristiky regrese	Společný model
Vícenásobný korelační koeficient	0.999941976
Koeficient determinace (R^2)	0.999883955
Predikovaný korelační koeficient (R_p)	0.999208179
Střední kvadratická chyba predikce (MEP)	191446156.1
Akaikeho informační kritérium (AIC)	184.427099

Z tabulky 1.10 je zřejmý velký význam modelu. Korelační koeficient vykazuje vysokou hodnotu. Koeficient determinace dokládá, že většina experimentálních bodů vyhovuje společnému modelu.

Dále byla provedena kritika modelu, výsledky jsou uvedeny v tabulce 1.11.

Tab. 1.11 Kritika metody společného modelu

Fisher-Snedocorův test významnosti modelu, F	
hodnota kritéria F	30157.18588
tabulkový kvantil $F(1-\alpha, m-1, n-m)$	4.737414126
Pravděpodobnost	1.68343E-14
Závěr	Model je významný
Cook-Weisbergův test heteroskedasticity, Sf	
hodnota kritéria Sf	1.334223833
tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.841458829
Pravděpodobnost	0.248055181
Závěr	Rezidua vykazují homoskedasticitu
Jarque-Berruv test normality reziduí, L(e)	
hodnota kritéria $L\epsilon$	1.352350872
tabulkový kvantil, $\chi^2(1-\alpha, 2)$:	5.991464547
Pravděpodobnost	0.508558291
Závěr	Rezidua mají normální rozdělení
Waldův test autokorelace, Wa	
hodnota kritéria Wa	0.336018926
tabulkový kvantil, $\chi^2(1-\alpha, 1)$	3.841458829
Pravděpodobnost	0.562136237
Závěr	Autokorelace je nevýznamná
Znaménkový test reziduí, Sg	
hodnota kritéria Sg	1.39113905
kvantil $N(1-\alpha/2)$	1.959963999
Pravděpodobnost	0.164183265
Závěr	V reziduích není trend

Dle tabulky 1.11 je zřejmé, že společný model je významný, rezidua vykazují homoskedasticitu a mají normální rozdělení, autokorelace je nevýznamná a v reziduích není trend.

Dále bylo provedeno porovnání regresních přímk a test shody rozptylů:

Model 1: $y = -40182,14(32604,76) + 2,5634(0,0835)x$
Reziduální součet čtverců (RSC_2): 40737106177
Reziduální směrodatná odchylka: 71359,22
 $\beta_0 = (-11538,85 \ 35004,56)$
 $\beta_1 = (2,3708 \ 2,7561)$

Model 2: $y = 13824,28(9203,82) + 0,551631(0,00527)x$
Reziduální součet čtverců (RSC_2): 3522470322
Reziduální směrodatná odchylka: 20983,54
 $\beta_0 = (-7399,77 \ 35048,33)$
 $\beta_1 = (0,5395 \ 0,56377)$

Test shody rozptylů:

$$F = \frac{\max(\sigma_1^2, \sigma_2^2)}{\min(\sigma_1^2, \sigma_2^2)} = 11,565$$

Pro porovnání dvou výběrů byly dále testovány shody rozptylů pomocí Fisher-Snedocova F-testu. Tabulková hodnota $F_{0,95} = 5,317$. Testační statistika je větší než tabulková hodnota F-rozdělení, tj. rozptyly nejsou shodné.

Dále bude proveden Chowův test pro porovnání regresních přímk s kritickou hodnotou F ($m; n - 2m$); reziduální součet čtverců $RSC = 561035643,7$:

$$F_c = \frac{(RSC - RSC_1 - RSC_2) \times (n - 2m)}{(RSC_1 + RSC_2) \times m} = -29,05$$

Hodnota testačního kritéria $F_c = -29,05$ je menší než tabulková hodnota $F_{0,95} = 4,737$. Z toho vyplývá, že lineární závislosti jsou shodné.

1.8 Závěr

V úloze č. 1 byla řešena závislost spotřeby paliva na dopraveném množství dříví a ujetých kilometrech. Nejprve byla provedena regresní diagnostika, testován regresní triplet, kdy byl zjištěn fakt, že v obou modelech existují odlehlé body. Model 1 závislost spotřeby paliva na dopraveném množství vykazoval 2 odlehlé body – bod 9 a bod 10. Model 2 závislost spotřeby paliva na ujetých kilometrech vykazoval jeden odlehlý bod 10. Byly stanoveny zpřesněné modely pro obě závislosti. Dále však bylo zohledněno, že je doporučováno pro každý regresní parametr použít alespoň 5 experimentálních dat, a také, že nebyly zohledněny další parametry odvozních souprav, proto pro další analýzy nebyl zpřesněný model využit.

U obou modelů by bylo možné vyloučit úsek, jelikož byl zjištěn jako nevýznamný. Směrodatná odchylka je významná. Výsledné rovnice lineárních regresních přímk lze napsat ve tvaru:

- pro model 1: $y = 2,5634(0,0835)x$
- pro model 2: $y = 0,551631(0,00527)x$

Testováním shody reziduálních rozptylů obou regresních přímk bylo zjištěno, že se rozptyly neshodují. Na základě Chowova testu lze konstatovat, že závislost spotřeby paliva na dopravovaném množství dříví i ujetých kilometrech se shoduje.

2 Úloha 2 Určení stupně polygonu

2.1 Zadání a vstupní data

Cílem úlohy je nalézt vhodný lineární regresní model pro závislost návštěvnosti dané lokality na denní době. Lokalitou bylo vybrané území Školního lesního podniku Masarykův les Křtiny. Sledovaná denní doba byla od třetí do osmnácté hodiny. Předpokladem je polynomická regresní závislost. Vstupní údaje uvádí tabulka 2.1. Použitým programem byl QC Expert.

Tab. 2.1 Vstupní data analýzy

Denní doba (hodiny)	Počet návštěvníků
3	0
4	2
5	7
6	10
7	9
8	15
9	8
10	30
11	8
12	8
13	11
14	17
15	30
16	51
17	43
18	63

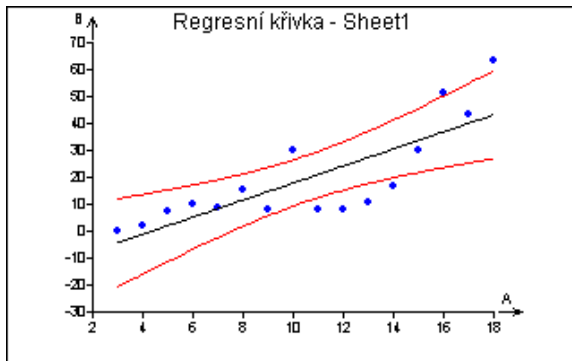
2.2 Návrh modelu

Pro závislost návštěvnosti území na denní době je třeba sestavit vhodný regresní model. Testována byla závislost od přímkové po několik stupňů polygonu. O vhodném stupni polygonu bude rozhodovat střední kvadratická chyba predikce (MEP), Akaikovo informační kritérium (AIC) a reziduální směrodatná odchylka (SD). Hodnoty jednotlivých ukazatelů pro testovanou přímkovou závislost a stupně polygonu uvádí tabulka 2.2.

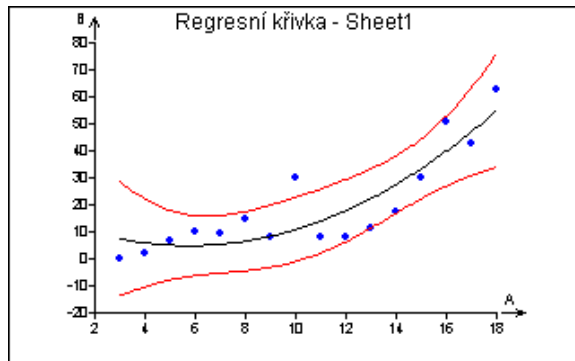
Tab. 2.2 Hodnoty MEP, AIC a SD pro přímkovou závislost a stupně polygonu

Model	MEP	AIC	SD
přímkový	146,6917	79,3714	11,2698
polynom 2. stupně	110,119	74,2886	9,3731
polynom 3. stupně	67,3871	68,7439	7,7067
polynom 4. stupně	87,6031	70,7301	8,0459
polynom 5. stupně	123,2700	70,0511	7,7609

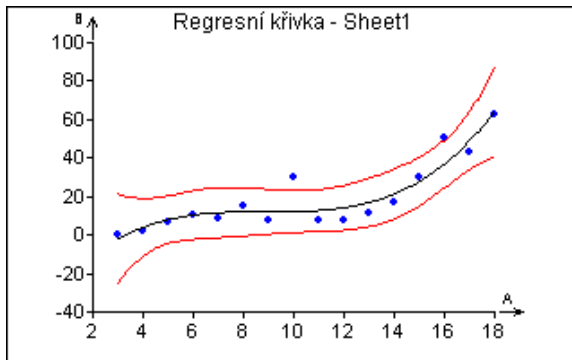
Z tabulky je zřejmé, že hodnota MEP, AIC i SD klesala, kdy nejnižší jsou všechny hodnoty pro polygon 3. stupně. U vyšších stupňů polygonu je kombinace všech tří ukazatelů horší než v případě polygonu 3. stupně. Z toho vyplývá, že nejvhodnějším regresním modelem je polygon 3. stupně. Toto tvrzení bude dále demonstrováno na grafech (obrázky 2.1 až 2.5).



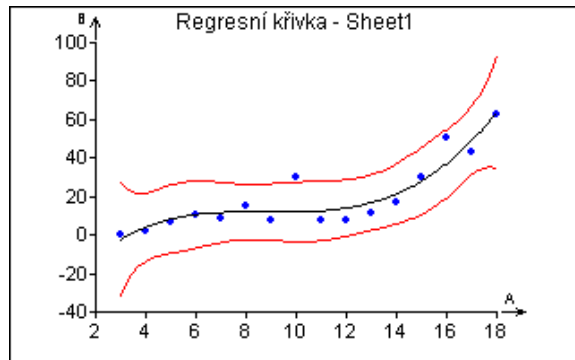
Obr. 2.1 Regresní závislost – přímka



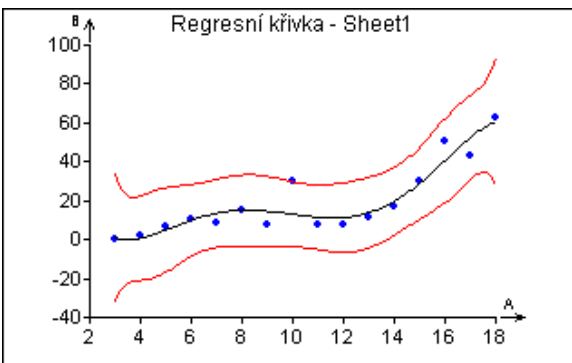
Obr. 2.2 Regresní závislost – polygon 2. stupně



Obr. 2.3 Regresní závislost – polygon 3. stupně



Obr. 2.4 Regresní závislost – polygon 4. stupně



Obr. 2.5 Regresní závislost – polygon 5. stupně

Z regresních křivek závislosti je zřejmé, že polygon 3. stupně vykazuje dostatečné proložení, pásy spolehlivosti jsou nejužší. U polygonů 4. a 5. stupně lze již pozorovat rozšíření pásů spolehlivosti na obou koncích regresní závislosti.

2.3 Odhadování parametrů

Pomocí metody nejmenších čtverců byly stanoveny odhady regresních parametrů. Analýzy dat probíhaly na hladině významnosti alfa rovno 0,05. Údaje z analýzy jsou uvedeny v tabulce 2.3.

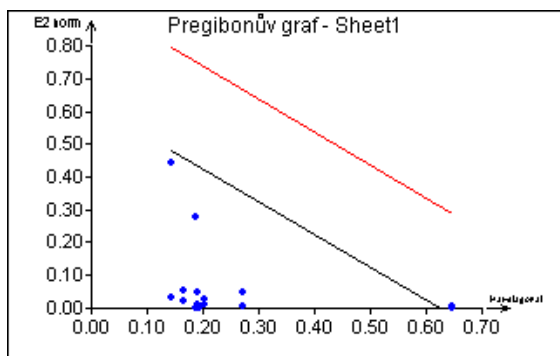
Tab. 2.3 Odhady regresních parametrů

Proměnná	Odhad	Směrodatná odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-36.2552252	21.94718628	Nevýznamný	0.12445463	-84.07403621	11.56358588
x	16.3027705	7.801433309	Nevýznamný	0.05860494	-0.695092483	33.30063348
x^2	-1.83370757	0.8125034	Významný	0.043456231	-3.604000404	-0.063414741
x^3	0.068806065	0.02558982	Významný	0.01971332	0.013050638	0.124561492

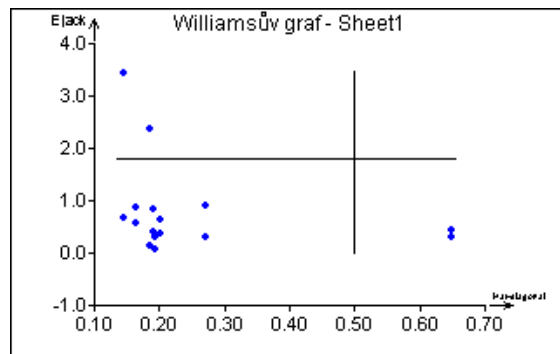
S využitím Studentova t-testu bylo zjištěno, že úsek, stejně jako směrnici v případě x lze považovat za nevýznamné, zatímco ostatní parametry jsou statisticky významné.

2.4 Regresní diagnostika

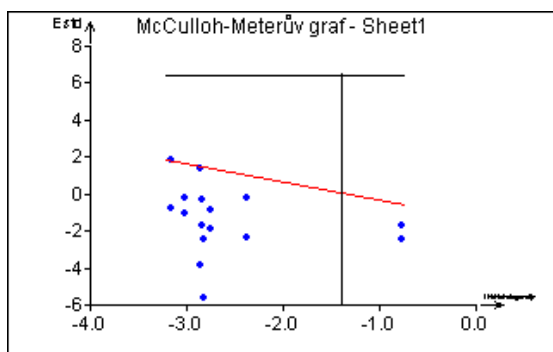
Dále byl testován regresní triplet, resp. především byly identifikovány odlehlé body. Výsledky uvádí obrázky 2.6 – 2.11.



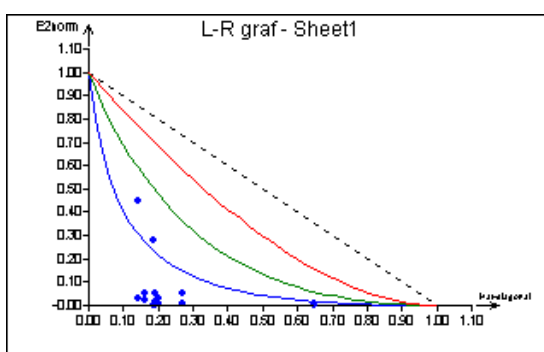
Obr. 2.6 Pregibonův graf



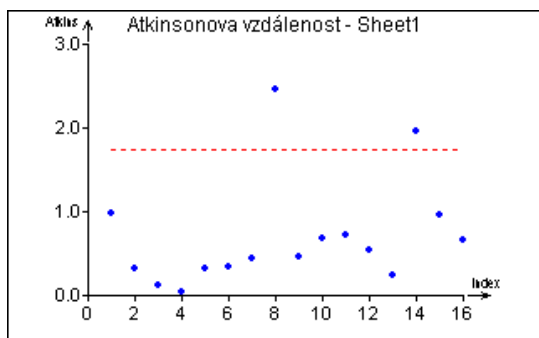
Obr. 2.7 Williamsův graf



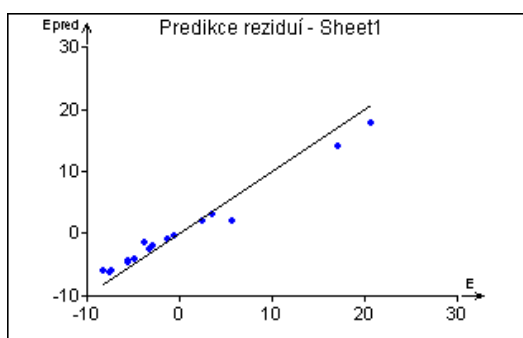
Obr. 2.8 McCullohův-Meeterův graf



Obr. 2.9 L-R graf



Obr. 2.10 Graf Atkinsonovy vzdálenosti



Obr. 2.11 Graf predikovaných reziduí

Pregibohův graf (2.6) identifikoval 2 středně vlivné body (1 a 16). Williamsův graf (2.7) odhalil 4 vlivné body, z nichž body 8 a 14 jsou pravděpodobně odlehlé body a body 1 a 16 jsou extrémní. Tytéž body identifikoval i graf predikovaných reziduí (2.11). McCullohův-Meeterův graf (2.8) také identifikoval body 1 a 16 jako extrémní. Dle Atkinsonovy vzdálenosti jsou body 8 a 14 odlehlými body. Dle grafické diagnostiky není zcela zřejmé, které body je možné z analýzy vyloučit, dalo by se předpokládat, že body 8 a 14. Avšak s ohledem na relevanci dat, není možné vyloučit z analýzy žádný bod. Tabulka 2.4 uvádí základní statistické regresní charakteristiky modelu.

Tab. 2.4 Základní statistické charakteristiky regrese

Statistické charakteristiky regrese	Hodnota
Vícenásobný korelační koeficient (R)	0,928888633
Koeficient determinace (R^2)	0,862834093
Predikovaný korelační koeficient (R_p)	0,628049132
Střední kvadratická chyba predikce (MEP)	67,38708335
Akaikeho informační kritérium (AIC)	68,74386513

Dle hodnoty párového korelačního koeficientu je navržený model statisticky významný. I koeficient determinace dosahuje vysoké hodnoty, tudíž většina experimentálních bodů vyhovuje navrženému modelu.

2.4.1 Kritika metody

Dále je třeba určit, zda model splňuje předpoklady MNČ.

1. Fisher-Snedecorův test významnosti modelu, F

- hodnota kritéria F: 17,3159
- tabulkový kvantil $F(1-\alpha, m-1, n-m)$: 3,3567
- závěr: model je významný
- pravděpodobnost: 0,0001

2. Cook-Weisbergův test heteroskedasticity, CW

- hodnota kritéria CW: 0,0421
- tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3,8414
- závěr: rezidua vykazují homoskedasticitu
- pravděpodobnost: 0,8373

4. Jarque-Berrův test normality reziduí, JB

- hodnota kritéria JB: 8,4718
- tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5,9915
- závěr: rezidua nemají normální rozdělení
- pravděpodobnost: 0,014

5. Waldův test autokorelace, WA

- hodnota kritéria WA: 0,3548
- tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3,8415
- závěr: autokorelace je nevýznamná
- pravděpodobnost: 0,551

6. Znaménkový test reziduí Sg

- hodnota kritéria Sg: 0.3809
- tabulkový kvantil, $N(1-\alpha/2)$: 1,9600
- závěr: v reziduích není trend
- pravděpodobnost: 0,703

Dle Fisher-Snedecorova testu je model významný, jelikož hodnota testovacího kritéria přesáhla hodnotu tabulkového kvantilu. Rezidua vykazují homoskedasticitu a nemají normální rozdělení. Autokorelace je nevýznamná. Znaménkový test prokázal, že rezidua nevykazují trend.

2.5 Závěr

Cílem úlohy bylo najít vhodný regresní model pro závislost návštěvnosti území na denní době. Byl stanoven polynomický regresní model. Pomocí MEP, AIC a SD byl vybrán vhodný stupeň polynomu – polynom 3. stupně. Byla zjištěna nevýznamnost úseku a jedné směrnice. Nalezený regresní model lze zapsat:

$$y = -1,8337(0,8125)x^2 + 0,0688(0,2559)x^3$$

3 Úloha 3 Validace nové analytické metody

3.1 Zadání a vstupní data

V rámci projektu Interní grantové agentury Lesnické a dřevařské fakulty (LDF) Mendelovy univerzity v Brně byl prováděn monitoring návštěvnosti území jednak specializovanou firmou pomocí pyroelektrického senzoru, jednak studenty LDF. Údaje jsou uvedeny z jedné lokality v souhrnu za první týden monitoringu v časovém intervalu 9 – 17 h. Sledovány byly dva směry – IN a OUT. Úkolem je pomocí regresní analýzy zjistit, zda pyroelektrický senzor poskytuje relevantní informace. Data jsou uvedena v tabulce 3.1. Pro zpracování dat bude využit program ADSTAT.

Tab. 3.1 Vstupní data analýzy; n = 24

Doba	Směry	Ruční sčítání	Sčítač
1. hodina	IN	36	27
	OUT	13	9
	CELK.	49	36
2. hodina	IN	22	23
	OUT	11	11
	CELK.	33	34
3. hodina	IN	56	45
	OUT	22	15
	CELK.	78	60
4. hodina	IN	23	15
	OUT	23	16
	CELK.	46	31
5. hodina	IN	14	16
	OUT	20	21
	CELK.	34	37
6. hodina	IN	18	14
	OUT	17	15
	CELK.	35	29
7. hodina	IN	16	10
	OUT	39	24
	CELK.	55	34
8. hodina	IN	16	10
	OUT	29	28
	CELK.	45	38

3.2 Návrh modelu

Pro validaci modelu je navržen přímkový regresní model. Pomocí metody nejmenších čtverců určíme parametry modelu, kde y je závislá proměnná (sčítač) a x je nezávislá proměnná (ruční sčítání). Rovnici modelu lze tedy zapsat: $y = \beta_0 + \beta_1 x$. U tohoto modelu bude testována nulová hypotéza ($H_0: \beta_0 = 0; \beta_1 = 1$), kdy předpokládáme ideální případ – přímka s nulovým úsekem a nulovou směrnici.

3.3 Předběžná analýza dat

Předběžná analýza dat je uvedena v tabulce 3.2. Párový (Pearsonův) korelační koeficient (R) ukazuje významnou korelaci proměnných y a x.

Tab. 3.2 Předběžná analýza dat

Proměnná	Průměr	Směrodatná odchylka	Párový korelační koeficient	Spočtená hladina významnosti
y	24,917	12,840	1,000	-----
x	32,208	16,395	0,886	0,000

3.4 Odhadování parametrů

Odhad parametrů byl proveden klasickou metodou nejmenších čtverců (MNC), jejíž pomocí byly stanoveny regresní parametry úsek β_0 a směrnice β_1 . Analýza probíhala na hladině významnosti alfa rovno 0,05. S využitím Studentova t-testu bylo zjištěno, že absolutní člen (úsek) lze považovat za statisticky nevýznamný, zatímco směrnice je statisticky významná, jelikož t je větší než kvantil Studentova rozdělení $t_{krit} = 2,074$. Tvzení dokládá tabulka 3.3.

Tab. 3.3 Odhady regresních parametrů

Parametr	Odhad	Směrodatná odchylka	t-kritérium	H0 je	Hladina významnosti
b0	2,5686	2,7867	0,9217	akceptována	0,367
b1	0,6939	0,0774	8,9601	zamítnuta	0,000

3.5 Základní statistické charakteristiky

Z tabulky 3.4 je zřejmé, že dle párového korelačního koeficientu R je navržený lineární model statisticky významný. Hodnota koeficientu determinace (78,49 %) uvádí podíl experimentálních bodů vyhovujících navrženému modelu.

Tab. 3.4 Základní statistické charakteristiky regrese

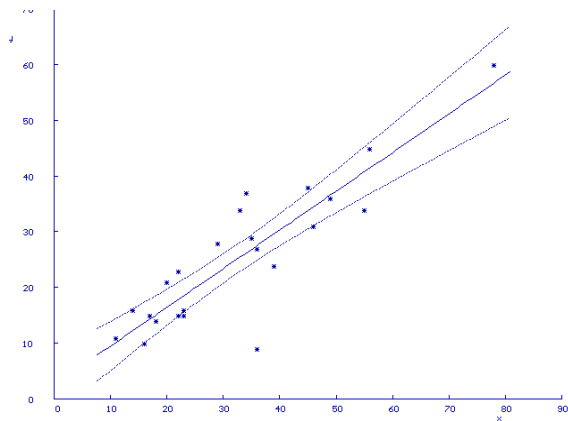
Statistické charakteristiky regrese	Hodnota
Vícenásobný korelační koeficient (R)	0,88595
Koeficient determinace (R^2)	0,78491
Predikovaný korelační koeficient (R_p)	0,86869
Střední kvadratická chyba predikce (MEP)	38,7680
Akaikeho informační kritérium (AIC)	88,6200

3.6 Regresní diagnostika

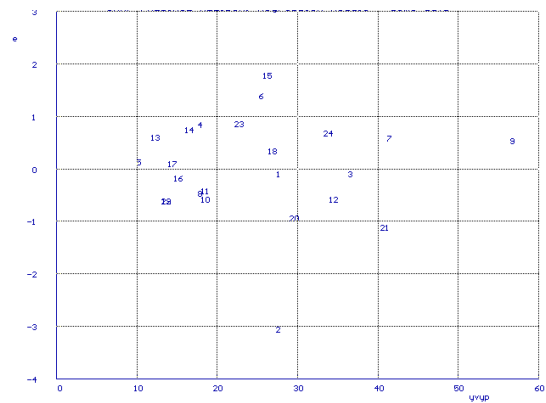
Regresní diagnostika se skládá z tzv. regresního tripletu, který obsahuje kritiku dat, modelu a metody.

3.6.1 Kritika dat

Je třeba zjistit věrohodnost odhadů parametrů β_0 , β_1 . Je třeba identifikovat vlivné body a silně odlehlé hodnoty vyloučit, aby bylo možné stanovit zpřesněný model.



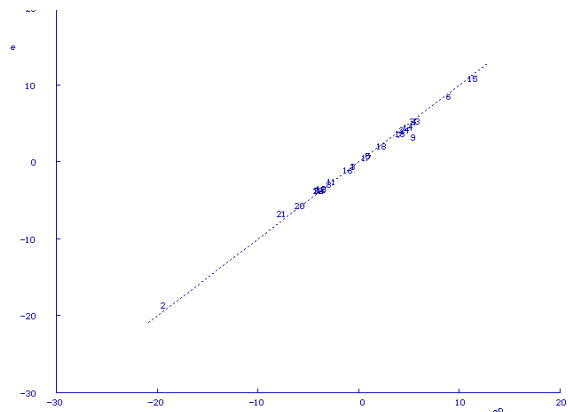
Obr. 3.1 Graf regresního modelu



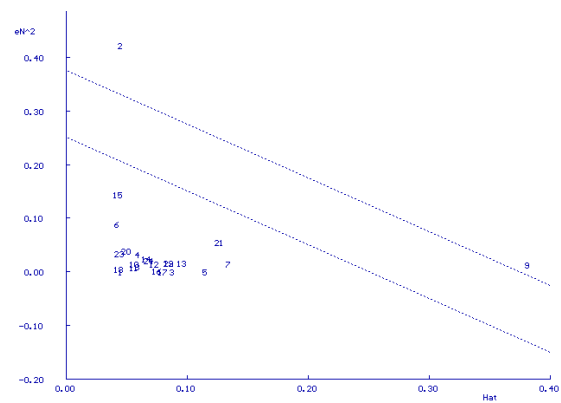
Obr. 3.2 Analýza klasických reziduí

Analýza klasických reziduí není příliš spolehlivá a nemusí indikovat odlehlé body. Analýza klasických reziduí (rezidua versus predikce) umožňuje identifikovat podezřelé body, trend a heteroskedasticitu.

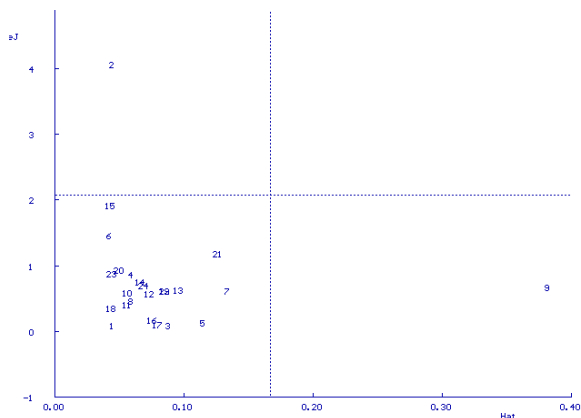
Grafy vlivných bodů jsou schopny identifikovat přítomnost odlehlých hodnot a extrémů.



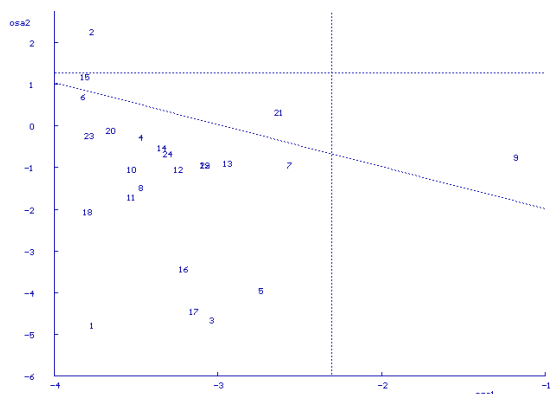
Obr. 3.3 Graf predikovaných reziduí



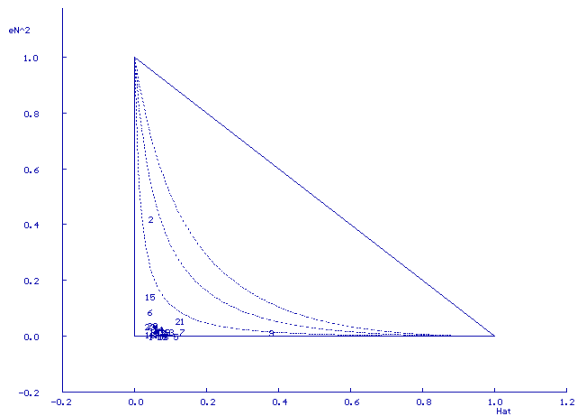
Obr. 3.4 Pregibonův graf



Obr. 3.5 Williamsův graf



Obr. 3.6 McCullohův-Meeterův graf



Obr. 3.7 L-R graf

Z grafů vlivných bodů lze usoudit, že model obsahuje odlehlé body, které je třeba odstranit. Jedná se o body 2 a 9.

3.6.2 Kritika modelu

Vhodnost modelu lze posoudit pouze z průběhu modelové funkce (obr. 3.1). Navržený lineární model lze posoudit jako vhodný. Přímková závislost byla dodržena.

3.6.3 Kritika metody

Neboli kritika splnění základních předpokladů MNČ:

1. Fisher-Snedcorův test významnosti modelu, F

- hodnota kritéria F: 80,284
- tabulkový kvantil $F(1-\alpha, m-1, n-m)$: 4,3009
- závěr: navržený model je přijat jako významný
- spočtená hladina významnosti: 0,000

2. Scottovo kritérium multikolinearity, M

- hodnota kritéria M: 0,000
- závěr: navržený model je korektní

3. Cook-Weisbergův test heteroskedasticity, Sf

- hodnota kritéria Sf: 121,000
- tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3,8415
- závěr: rezidua vykazují heteroskedasticitu
- spočtená hladina významnosti: 0,000

4. Jarque-Berruv test normality reziduí, L(e)

- hodnota kritéria L(e): 8,5603
- tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5,9915
- závěr: normalita není přijata
- spočtená hladina významnosti: 0,014

5. Waldův test autokorelace, Wa

- hodnota kritéria Wa: 1,0933

- tabulkový kvantil, χ^2 (1-alfa, 1): 3,8415
- závěr: rezidua nejsou autokorelována

6. Znaménkový test reziduí Dt

- hodnota kritéria Dt: -1,0436
- tabulkový kvantil, N (1-alfa/2): 1,6449
- závěr: rezidua nevykazují trend
- spočtená hladina významnosti: 0,148

Fisher-Snedecorův test prokázal významnost modelu. Rezidua vykazují heteroskedasticitu a nemají normální rozdělení.

3.7 Konstrukce zpřesněného modelu

Po odstranění bodů č. 2 a 9 bude sestaven zpřesněný regresní model a budou nalezeny nové odhady parametrů zpřesněného modelu – viz tabulka 3.5.

Tab. 3.5 Odhad parametrů zpřesněného modelu

Parametr	Odhad	Směrodatná odchylka	t-kritérium	H0 je	Hladina významnosti
b0	3,6162	2,4773	1,4323	akceptována	0,159
b1	0,6820	0,0753	9,0517	zamítnuta	0,000

Základní statistické charakteristiky modelu uvádí tabulka 3.6.

Tab. 3.6 Základní statistické charakteristiky regrese

Statistické charakteristiky regrese	Hodnota
Vícenásobný korelační koeficient (R)	0,89655
Koeficient determinace (R^2)	0,80379
Predikovaný korelační koeficient (R_p)	0,87401
Střední kvadratická chyba predikce (MEP)	24,5860
Akaikeho informační kritérium (AIC)	70,3760

Hodnota střední kvadratické chyby predikce (MEP) klesla z hodnoty 38,768 na hodnotu 24,586 a zároveň klesla hodnota Akaikeho informačního kritéria (AIC) z hodnoty 88,620 na hodnotu 70,376. Snížení obou těchto hodnot značí zkvalitnění modelu. Odstraněním odlehlých bodů došlo také ke zkvalitnění charakteristik modelu z pohledu normality reziduí:

Jarque-Berruv test normality reziduí, L(e)

- hodnota kritéria L(e): 1,0236
- tabulkový kvantil, χ^2 (1-alfa, 2): 5,9915
- závěr: normalita není přijata
- spočtená hladina významnosti: 0,599

3.8 Zhodnocení kvality modelu

Zpřesněný model lze zapsat:

$$y = 3,6162(2,4773) + 0,6820(0,0753)x$$

Jelikož došlo ke snížení rozhodujících kritérií MEP a AIC a zároveň ke zvýšení Pearsonova korelačního koeficientu i koeficientu determinace, lze považovat zpřesněný model za kvalitnější.

Intervalový odhad parametrů úseku β_0 :

$$b_0 - t_{1-\frac{\alpha}{2}}(n-2)\sqrt{D(b_0)} \leq \beta_0 \leq b_0 + t_{1-\frac{\alpha}{2}}(n-2)\sqrt{D(b_0)}$$
$$3,3162 - 2,074 \times 2,4773 \leq \beta_0 \leq 3,3162 + 2,074 \times 2,4773$$
$$-1,8217 \leq \beta_0 \leq 8,4541$$

Intervalový odhad parametrů směrnice β_1 :

$$b_1 - t_{1-\frac{\alpha}{2}}(n-2)\sqrt{D(b_1)} \leq \beta_1 \leq b_1 + t_{1-\frac{\alpha}{2}}(n-2)\sqrt{D(b_1)}$$
$$0,6820 - 2,074 \times 0,0753 \leq \beta_1 \leq 0,6820 + 2,074 \times 0,0753$$
$$0,5258 \leq \beta_1 \leq 0,8381$$

3.9 Závěr

Byla provedena analýza validace analytické metody, kdy bylo posuzováno, zda pyroelektrický senzor poskytuje relevantní informace a zda jím lze nahradit ruční sčítání návštěvníků území. Původní regresní model zahrnoval 24 bodů. Body 2 a 9 byly z důvodu jejich odlehlosti, což bylo zjištěno především grafickou analýzou, vyloučeny. Dále byl sestaven zpřesněný model, který vykazuje lepší vlastnosti. Intervalový odhad úseku pro výsledný model obsahoval hodnotu 0, tudíž je možné úsek považovat za nulový. Intervalový odhad směrnice neobsahoval hodnotu 1, tudíž nelze považovat směrnici za jednotkovou. Na základě výsledků lze konstatovat, že výsledky monitoringu návštěvnosti pomocí pyroelektrických sčítačů se pouze mírně odlišují od metody ručního sčítání.

4 Úloha 4 Vícerozměrný lineární regresní model

4.1 Zadání a vstupní data

U dvaceti odvozních souprav byly zjišťovány následující parametry – spotřeba paliva, množství dopraveného jehličnatého dřeva (m^3), množství dopraveného listnatého dřeva (m^3), počet najetých kilometrů v případě plné soupravy a počet najetých kilometrů v případě prázdné soupravy. Naměřená data jsou uvedena v tabulce 4.1. Cílem úlohy je zjistit vliv jednotlivých parametrů na celkovou spotřebu paliva (v litrech) a vytvořit vícerozměrný lineární regresní model. Použitým programem byl QC Expert.

Tab. 4.1 Vstupní data analýzy

Množství dopraveného dříví - jehličnaté	16 587	36 921	30 964	31 854	36 746	26 977	64 987	87 543	89 439	125 587
Množství dopraveného dříví - listnaté	1 712	5 719	454	387	6 964	167	543	6 676	10 000	136 922
Kilometry ujeté s prázdným vozidlem	19 266	35 716	48 636	68 684	69 243	66 410	99 132	144 440	145 533	206 428
Kilometry ujeté s naloženým vozidlem	16 257	32 844	59 304	65 251	75 921	103 903	103 872	143 594	141 985	200 835
Spotřeba paliva	25 436	39 721	59 134	75 376	86 924	88 026	126 982	175 521	176 692	268 208
Množství dopraveného dříví - jehličnaté	54 815	280 687	222 361	466 824	332 340	500 592	620 743	737 066	845 577	664 279
Množství dopraveného dříví - listnaté	8 902	70 092	10 885	8 620	4 610	16 649	136 862	27 556	31 982	30 034
Kilometry ujeté s prázdným vozidlem	162 924	450 117	381 229	885 291	576 334	1 057 360	1 253 788	1 328 219	1 457 620	1 427 617
Kilometry ujeté s naloženým vozidlem	364 819	435 920	402 871	679 328	850 768	1 311 130	1 497 326	1 901 003	2 003 174	1 985 227
Spotřeba paliva	292 969	446 247	473 217	813 548	817 000	1 330 002	1 562 097	1 814 832	1 857 609	1 859 244

4.2 Návrh modelu

Pro řešení závislosti spotřeby paliva na jednotlivých parametrech byl sestaven regresní model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

U stanoveného modelu bude testována významnost jednotlivých regresních parametrů.

4.3 Předběžná analýza dat

V případě předběžné analýzy dat bude sledován průměr a směrodatná odchylka nezávislých proměnných a vypočteny párové korelační koeficienty, které vypoví o síle závislosti mezi jednotlivými proměnnými. Předběžná analýza dat, která obsahuje průměr, směrodatnou odchylku, korelační koeficienty nezávislých proměnných vůči závislé proměnné a významnost je uvedena v tabulce 4.2.

Tab. 4.2 Předběžná analýza dat

Proměnná	Průměr	Směrodatná odchylka	Kor.vs.Y	Významnost
Množství dopraveného dříví - jehličnaté	263644,45	275894,4054	0,982504876	1,38E-14
Množství dopraveného dříví - listnaté	25786,8	41407,2384	0,322555781	0,165430632
Kilometry ujeté s prázdným vozidlem	494199,35	529898,1956	0,991639468	0
Kilometry ujeté s naloženým vozidlem	618766,6	713019,7875	0,996526041	0

Hodnoty korelačních koeficientů se pohybovaly v rozmezí 0,322 po 0,99. U množství dopraveného listnatého dříví nebyla závislost prokázána.

Sledovány byly i závislosti mezi proměnnými. Tabulka 4.4 uvádí párové korelace dle Pearsona pro všechny dvojice nezávislých proměnných.

Tab. 4.3 Párové korelace

Párové korelace	r	Pravděpodobnost
A – B	0,331705376	0,153079026
A – C	0,991507013	0
A – D	0,973044665	6,47E-13
B – C	0,334316549	0,149674912
B – D	0,289543984	0,215627446
C – D	0,980936788	2,95E-14

4.4 Odhadování parametrů

Odhad parametrů byl proveden klasickou metodou nejmenších čtverců (MNČ), jejíž pomocí byly stanoveny odhady regresních parametrů. Analýza probíhala na hladině významnosti alfa rovno 0,05. Výsledky jsou uvedeny v tabulce 4.4.

Tab. 4.4 Odhady regresních parametrů

Proměnná	Odhad	Směrodatná odchylka	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	9953,43536	9081,02712	Nevýznamný	0,29033	-9402,31576	29309,18648
Množství dopraveného dříví - jehličnaté	-0,14567	0,17821	Nevýznamný	0,42650	-0,52550	0,23417
Množství dopraveného dříví - listnaté	0,29464	0,16755	Nevýznamný	0,09904	-0,06249	0,65176
Kilometry ujeté s prázdným vozidlem	0,52223	0,11140	Významný	0,00029	0,28478	0,75968
Kilometry ujeté s naloženým vozidlem	0,61769	0,04720	Významný	0,00000	0,51709	0,71829

S využitím Studentova t-testu bylo zjištěno, že absolutní člen (úsek) lze považovat za statisticky nevýznamný, také jsou nevýznamné parametry množství dopraveného jehličnatého dříví a množství dopraveného listnatého dříví.

4.5 Základní statistické charakteristiky

Bude zjišťováno, zda navržený model je statisticky významný. Základní statistické charakteristiky regrese uvádí tabulka 4.5.

Tab. 4.5 Základní statistické charakteristiky regrese

Statistické charakteristiky regrese	Hodnota
Vícenásobný korelační koeficient (R)	0,999334221
Koeficient determinace (R^2)	0,998668886
Predikovaný korelační koeficient (R_p)	0,992973813
Střední kvadratická chyba predikce (MEP)	1539883592
Akaikeho informační kritérium (AIC)	413,654616700

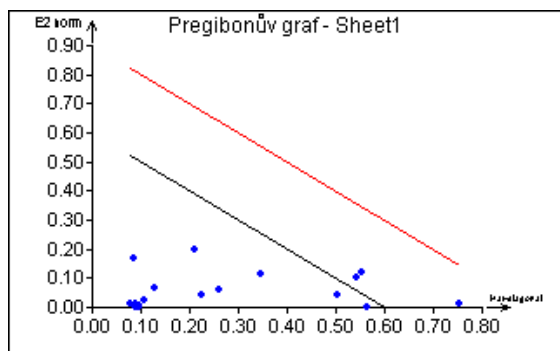
Vysoká hodnota vícenásobného korelačního koeficientu R znamená, že model je významný. Hodnota koeficientu determinace, která indikuje podíl experimentálních bodů vyhovujících modelu, je také poměrně vysoká – 99,87 % experimentálních bodů vyhovuje modelu.

4.6 Regresní diagnostika

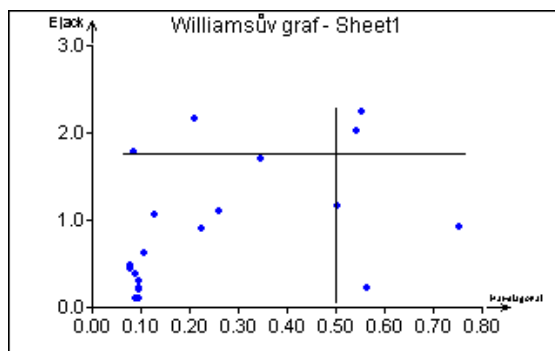
Regresní diagnostiku tvoří tzv. regresní triplet – kritika dat, modelu a metody.

4.6.1 Kritika dat

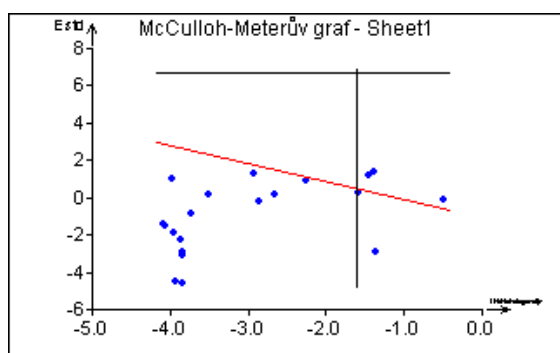
Dle grafů na obrázcích 4.1 až 4.6 budou identifikovány vlivné body, odlehlé body je možné vyloučit, aby mohl být stanoven zpřesněný model. K identifikaci vlivných bodů je dále možné využít indexové a ranktové grafy.



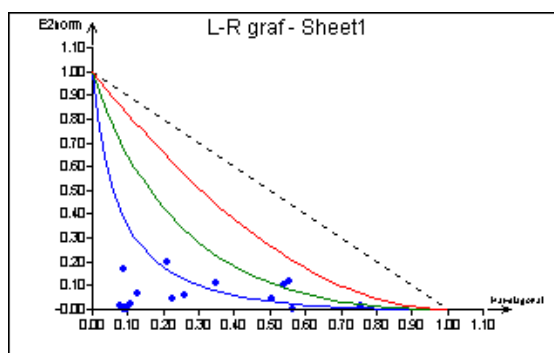
Obr. 4.1 Pregibonův graf



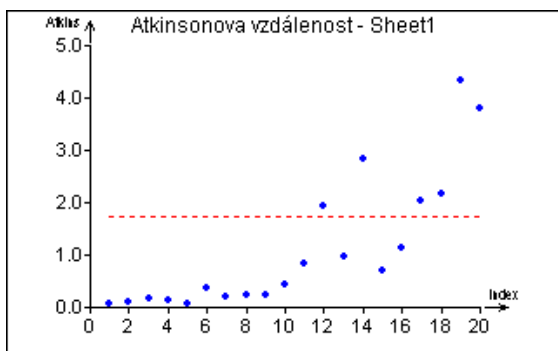
Obr. 4.2 Williamsův graf



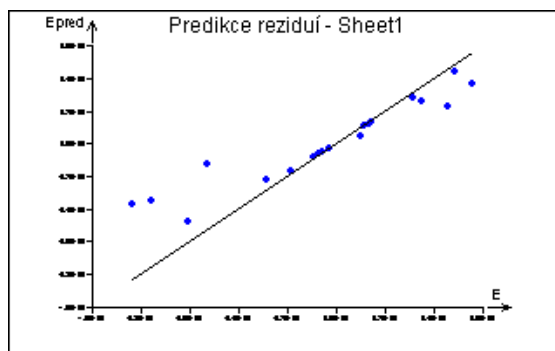
Obr. 4.3 McCullohův-Meeterův graf



Obr. 4.4 L-R graf



Obr. 4.5 Graf Atkinsonovy vzdálenosti



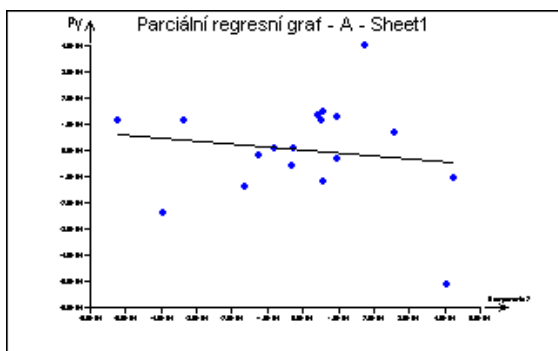
Obr. 4.6 Graf predikovaných reziduí

Pregibonův graf (obr. 4.1) odhalil 3 středně vlivné body – 14, 20 a 19. Williamsův graf (4.2) ukazuje 7 vlivných bodů, z toho body 12, 13, 19 a 20 jsou odlehlé body, zbytek spíše jsou extrémny. McCullohův-Meeterův graf (4.3) identifikoval body 19, 20 a 14 jako podezřelý a 10, 17 jako extrémny. V případě L-R grafu (4.4) jsou podezřelý body 12, 14, 17, 18, 19 a 20, stejné body uvádí Atkinsonova vzdálenost (4.5) i graf predikce reziduí (4.6).

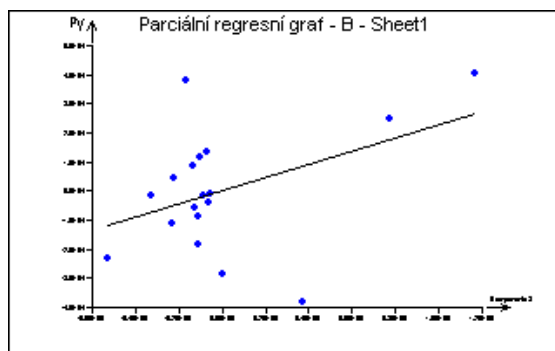
Jelikož jde o reálná data, nelze z modelu žádná data vyloučit.

4.6.2 Kritika modelu

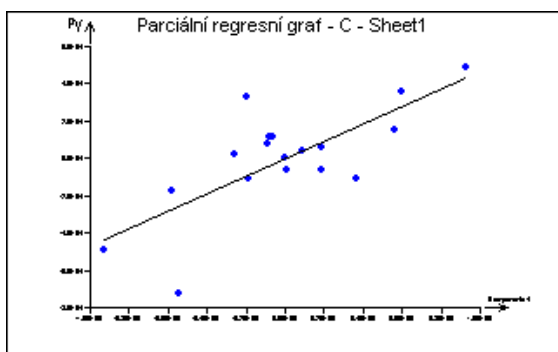
Navržený regresní model není vhodný, jelikož některé z parametrů byly identifikovány jako nevýznamné. Je třeba zjistit, které regresní parametry jsou významné a ty je třeba zahrnout do výsledného modelu. Pro posouzení významnosti regresních parametrů bude využito grafických diagnostik – parciálních regresních a reziduálních grafů (obr. 4.7 až 4.14).



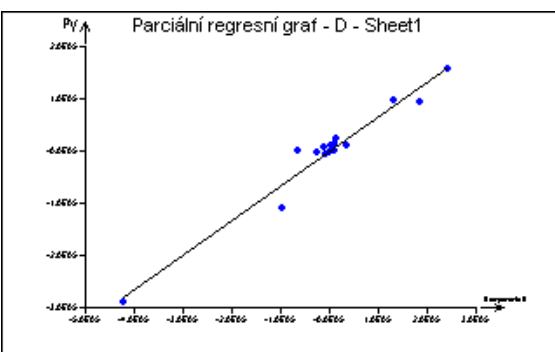
Obr. 4.7 Parciální regresní graf (x_1)



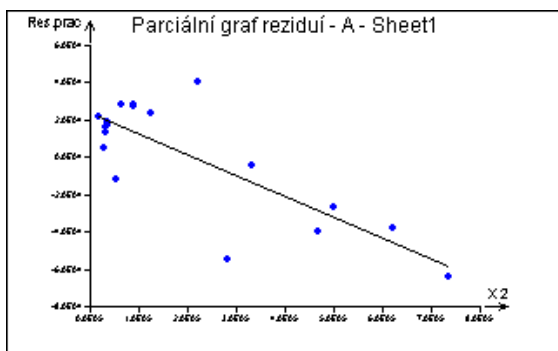
Obr. 4.8 Parciální regresní graf (x_2)



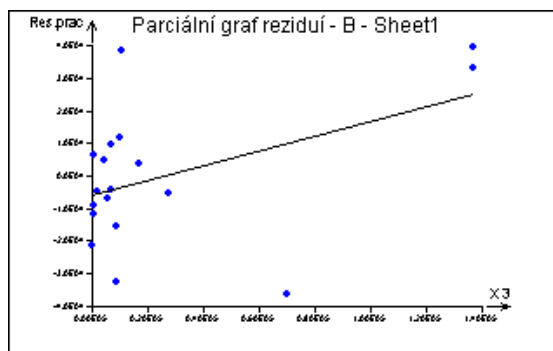
Obr. 4.9 Parciální regresní graf (x_3)



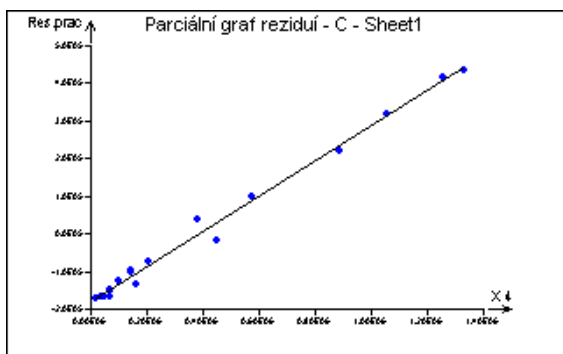
Obr. 4.10 Parciální regresní graf (x_4)



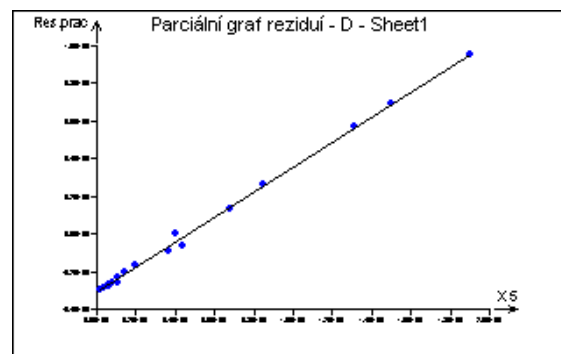
Obr. 4.11 Parciální reziduální graf (x_1)



Obr. 4.12 Parciální reziduální graf (x_2)



Obr. 4.13 Parciální reziduální graf (x_3)



Obr. 4.14 Parciální reziduální graf (x_4)

Z parciálních regresních a reziduálních grafů lze usuzovat, že statisticky nevýznamná je proměnná x_1 – množství, jelikož je v grafu na obrázku 4.7 vysoký rozptyl bodů s téměř nulovou směrnici. Největší význam dle obrázků lze usuzovat u parametru kilometry ujeté s plně naloženou odvozní soupravou.

4.6.3 Kritika metody

Bude posouzeno splnění základních předpokladů MNČ.

1. Fisher-Snedcorův test významnosti modelu, F

- hodnota kritéria F: 2825,018
- tabulkový kvantil $F(1-\alpha, m-1, n-m)$: 3,1791
- závěr: model je významný
- pravděpodobnost: 5,8477-E019

2. Cook-Weisbergův test heteroskedasticity, CW

- hodnota kritéria CW: 0,0487
- tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3,8414
- závěr: rezidua vykazují homoskedasticitu
- pravděpodobnost: 0,825

4. Jarque-Berrův test normality reziduí, JB

- hodnota kritéria JB: 1,4721
- tabulkový kvantil, $\chi^2(1-\alpha, 2)$: 5,9915
- závěr: rezidua mají normální rozdělení
- pravděpodobnost: 0,4790

5. Waldův test autokorelace, WA

- hodnota kritéria WA: 0,3513
- tabulkový kvantil, $\chi^2(1-\alpha, 1)$: 3,8415
- závěr: autokorelace je nevýznamná
- pravděpodobnost: 0,553

6. Znaménkový test reziduí Sg

- hodnota kritéria Sg: 0,3001
- tabulkový kvantil, N (1- α /2): 1,9600
- závěr: v reziduích není trend
- pravděpodobnost: 0,763

4.7 Zhodnocení kvality modelu

Sestavený regresní model má tvar:

$$y = 0,5222(0,1114)x_3 + 0,6177(0,0472)x_4$$

4.8 Závěr

Byl sestaven vícerozměrný lineární regresní model, který měl odhalit závislost množství dopraveného jehličnatého dřeva, množství dopraveného listnatého dřeva, počet najetých kilometrů v případě plné soupravy a počet najetých kilometrů v případě prázdné soupravy na spotřebě pohonných hmot u dvaceti odvozních souprav. Navržený model byl ohodnocen jako nejlépe vhodný, z důvodu relevance dat nebylo možné vynechání odlehlých bodů. Bylo zjištěno, že na spotřebu paliva má vliv počet najetých kilometrů u odvozních souprav.