

6. Lineární regresní modely

- 6.1 Jednoduchá regrese a validace
- 6.2 Testy hypotéz v lineární regresi
- 6.3 Kritika dat v regresním tripletu
- 6.4 Multikolinearita a polynomy
- 6.5 Kritika modelu v regresním tripletu
- 6.6 Kritika metody v regresním tripletu
- 6.7 Lineární a nelineární kalibrace
- 7. Korelační modely

REGRESNÍ DIAGNOSTIKA

Vyšetřuje **regresní triplet**, což představuje

- ◆ **Kritiku dat** (zkoumá kvalitu dat pro navržený model)
- ◆ **Kritiku modelu** (zkoumá kvalitu modelu pro daná data)
- ◆ **Kritiku metody odhadu** (prověřuje splnění všech předpokladů požadovaných metodou MNČ)

1

Kritika dat

Vyšetření vlivných bodů:

- (1) Odlehlé body
- (2) Extrémy

Metoda:

Grafy vlivných bodů
Vybrané diagnostiky vlivných bodů

1. KRITIKA DAT

Výskyt *vlivných bodů* (VB):

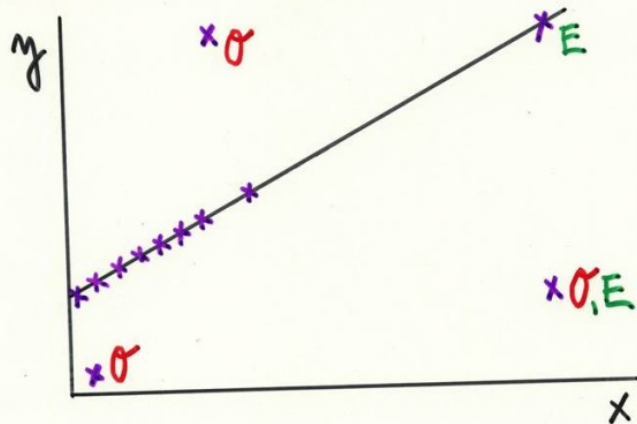
- zdrojem řady problémů,
- zkreslení odhadů a růst rozptylů odhadů parametrů.

Dělení *vlivných bodů* dle charakteru:

1. Hrubé chyby: vybočující pozorování, **extrémy**.
2. Body s vysokým vlivem (tzv. golden points): rozšiřují predikční schopnosti modelu.
3. Zdánlivě vlivné body: důsledek nesprávného regresního modelu.

Dělení vlivných bodů dle výskytu:

1. vybočující pozorování (outliers, O): na y se liší,
2. extrémy (high leverage points, E): liší se na ose x



Vlastnosti klasických reziduí:

1. Rozptyl reziduí

$$D(\hat{\epsilon}_i) = (1 - H_{ii}) \hat{\sigma}^2$$

je nekonstantní, i když rozptyl chyb σ^2 je konstantní.

2. Rezidua jsou korelovaná: existuje párový korelační koeficient r_{ij} mezi dvěma rezidui ϵ_i a ϵ_j

$$r_{ij} = \frac{-H_{ij}}{\sqrt{(1 - H_{ii})(1 - H_{jj})}}$$

i když chyby ϵ_i a ϵ_j jsou nezávislé.

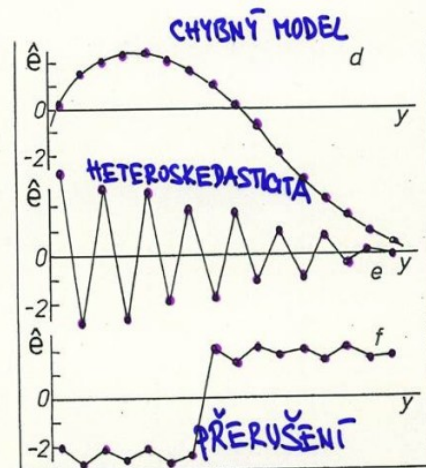
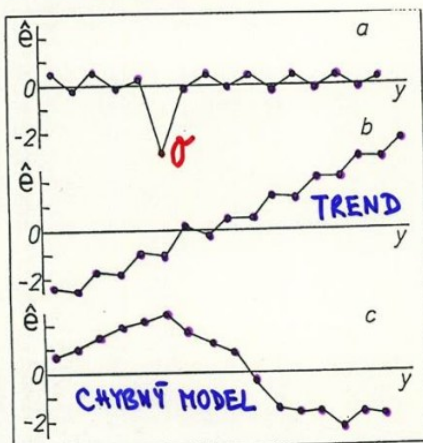
3. Rezidua neindikují silně extrémní hodnoty.

4. Rezidua jsou normálnější než chyby:
(efekt supernormality)

5. U malých výběrů nemusí správně indikovat model.

Statistická analýza klasických reziduí

1. Grafická analýza



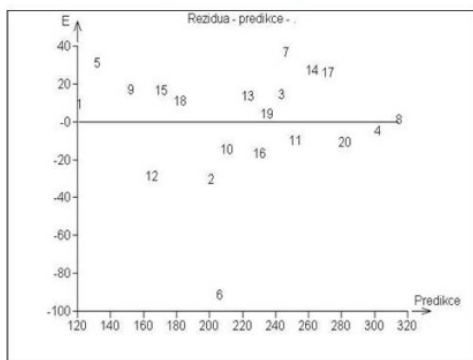
2. Numerická analýza:

- (a) Střední hodnota reziduí $E(\hat{\epsilon}) = 0$,
- (b) Průměrné residuum $|\bar{\epsilon}| = \frac{1}{n} \sum |\hat{\epsilon}_i| \approx \epsilon$,
- (c) Směrodatná odchylka střední hodnoty reziduí $s(\hat{\epsilon}) \approx \epsilon$,
- (d) Koeficient šikmosti souboru reziduí $g_1(\hat{\epsilon}) \approx 0$,
- (e) Koeficient špičatosti souboru reziduí $g_2(\hat{\epsilon}) \approx 3$,
- (f) Pearsonův χ^2 -test dobré shody: χ^2_{exp} vs. χ^2_{crit} ,
- (g) Hamiltonův R-faktor relativní těsnosti: $R \approx 0.5 \%$.

Reziduum	Reziduum [%Y]
9.324031	7.172332
-30.597716	-17.998656
14.587063	5.6539
-4.927073	-1.658947
30.906639	18.961128
-91.306412	-79.39688
36.732715	12.979758
1.228361	0.388722
16.926202	10.015504
-14.587934	-7.442823
-9.975981	-4.122306
-28.918364	-21.263503
13.5675	5.724683
27.0338	9.354256
16.372939	8.755583
-16.568371	-7.742229
26.043582	8.739457
10.955547	5.676449
4.150107	1.736447
-10.946636	-4.039349

9

Grafická analýza klasických reziduí:



Numerická analýza klasických reziduí:

Reziduální součet čtverců:	15731.026573
Průměr absolutních reziduí:	20.782849
Reziduální směr. odchylka:	29.562576
Reziduální rozptyl:	873.945921
Šikmost reziduí:	2.450042
Špičatost reziduí:	6.078201

1. vzorová úloha na výstavbu lineárního regresního modelu

J625

10

Úloha J6.25 Závislost celkového cholesterolu v krvi na denní spotřebě tuku

Zadání: U vzorku 20 Američanů byla v analýze krve sledována denní spotřeba tuku ve stravě x v gramech a obsah celkového cholesterolu y v mg na 100 ml krve. Byl navržen jednoduchý lineární regresní model $y = \beta_0 + \beta_1 x$.

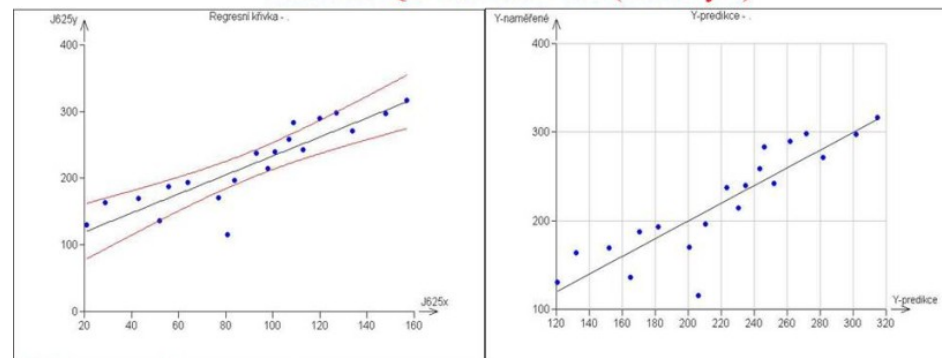
Úkoly: Ukažte

- (1) platnost navrženého regresního modelu a existenci vlivných bodů.
- (2) Testujte statistickou významnost obou parametrů, úseku β_0 a směrnice β_1 .
- (3) Sestrojte 95%ní interval spolehlivosti úseku β_0 a vysvětlete fakt, že $\beta_0 = 0$.
- (4) Sestrojte 95 % oboustranný interval spolehlivosti směrnice β_1 .
- (5) Naleznete 95 % interval spolehlivosti celkového cholesterolu u lidí, kteří denně spotřebují 50 g tuku.
- (6) Jaký je Pearsonův korelační koeficient mezi celkovým cholesterolem v krvi y a denní spotřebou tuku x u sledovaných jedinců?

Data: Denní spotřeba tuku x [g], obsah celkového cholesterolu v krvi y [mg/100 ml]:

x	y
21	130
...	...
134	271

Software QC-EXPERT 3.1 (TriloByte)



Odhady parametrů				Spodní mez		Horní mez	
Proměnná	Odhad	Směr.odch.	Závěr	Pravděpodobnost			
Abs	90.705314	17.420915	Významný	0.000059	54.10533	127.305298	
J625x	1.427174	0.177707	Významný	0.0	1.053825	1.800523	

Statistické charakteristiky regrese

Vícenásobný korelační koeficient R :	0.884201
Koeficient determinace R ² :	0.781812
Predikovaný korelační koeficient Rp :	0.555095
Střední kvadratická chyba predikce MEP :	919.087734
Akaikeho informační kritérium :	137.35316

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu	
Hodnota kritéria F :	64.497725
Kvantil F (1-alfa, m-1, n-m) :	4.413873
Pravděpodobnost :	0.0
Závěr :	Model je významný

12

Analyza klasických reziduí

Index	Yexp	Yvyp	Sm. od. Y	Reziduum	Reziduum [%Y]	Váhy
1.0	130.0	120.675969	14.039759	9.324031	7.172332	1.0
2.0	170.0	200.597716	7.044467	-30.597716	-17.998656	1.0
3.0	258.0	243.412937	7.217183	14.587063	5.6539	1.0
4.0	297.0	301.927073	12.140139	-4.927073	-1.658947	1.0
5.0	163.0	132.093361	12.803052	30.906639	18.961128	1.0
6.0	115.0	206.306412	6.831445	-91.306412	-79.39688	1.0
7.0	283.0	246.267285	7.367025	36.732715	12.979758	1.0
8.0	316.0	314.771639	13.509713	1.228361	0.388722	1.0
9.0	169.0	152.073798	10.749443	16.926202	10.015504	1.0
10.0	196.0	210.587934	6.716764	-14.587934	-7.442823	1.0
11.0	242.0	251.975981	7.707245	-9.975981	-4.122306	1.0
12.0	136.0	164.918364	9.539081	-28.918364	-21.263503	1.0
13.0	237.0	223.4325	6.623017	13.5675	5.724683	1.0
14.0	289.0	261.9662	8.414764	27.0338	9.354256	1.0
15.0	187.0	170.627061	9.040035	16.372939	8.755583	1.0
16.0	214.0	230.568371	6.736482	-16.568371	-7.742229	1.0
17.0	298.0	271.956418	9.236325	26.043582	8.739457	1.0
18.0	193.0	182.044453	8.136967	10.955547	5.676449	1.0
19.0	239.0	234.849893	6.859125	4.150107	1.736447	1.0
20.0	271.0	281.946636	10.144258	-10.946636	-4.039349	1.0

Reziduální součet čtverců: 15731.026573

Průměr absolutních reziduí: 20.782849

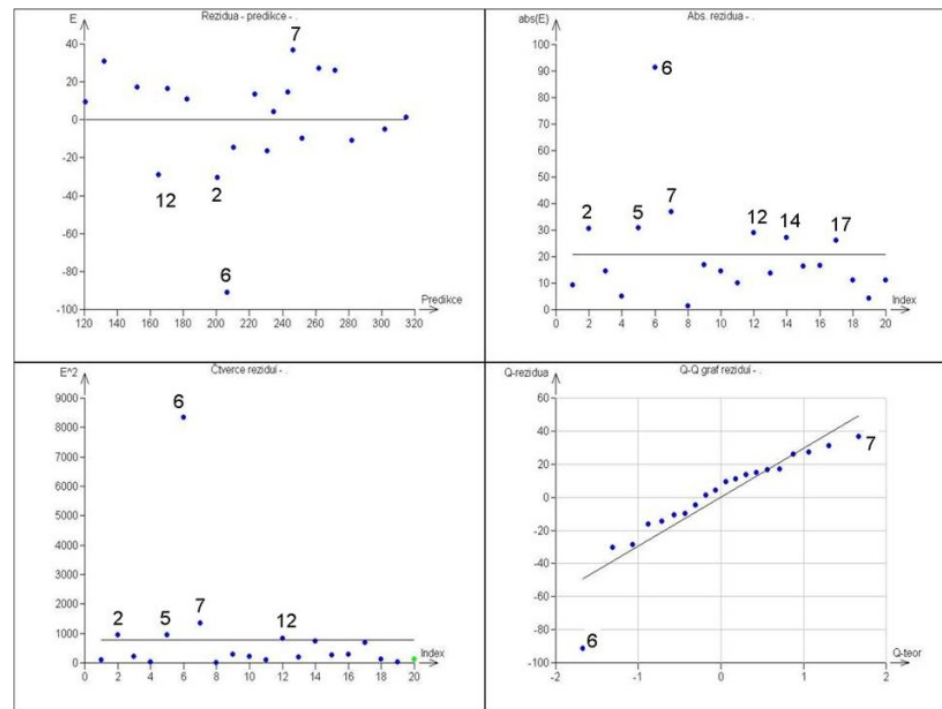
Reziduální směr. odchylka: 29.562576

Reziduální rozptyl: 873.94592

Šikmost reziduí: 2.450042

Špičatost reziduí: 6.078201

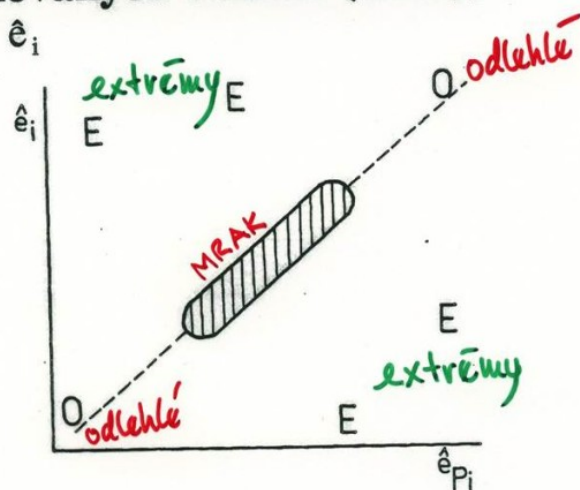
13



Grafy identifikace vlivných bodů

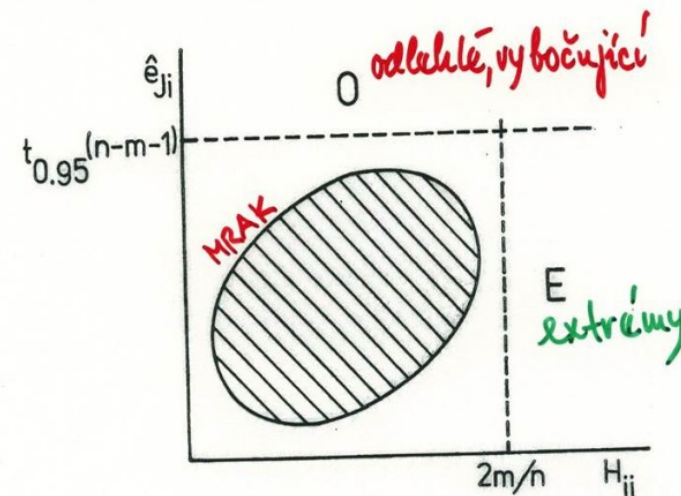
1. Graf predikovaných reziduí (GPR),

osa x: \hat{e}_{Pi} , osa y: \hat{e}_i

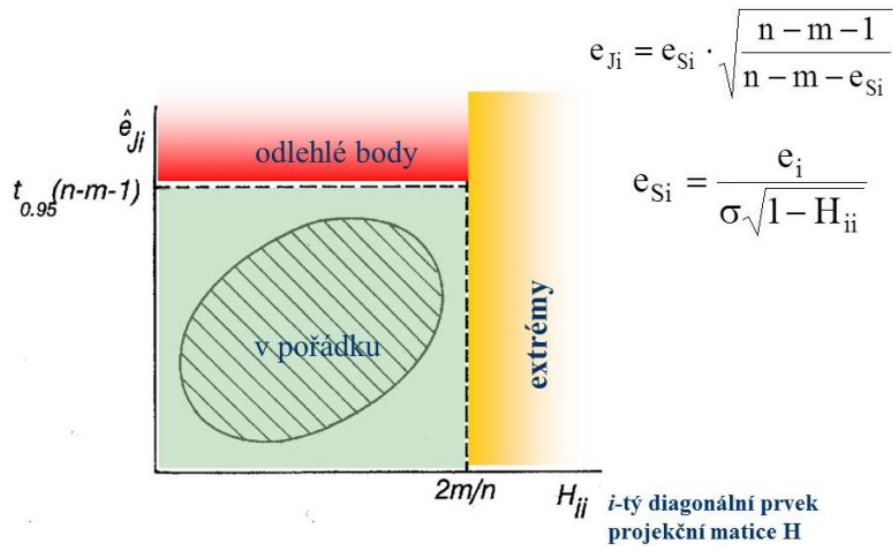


2. Williamsův graf (WG),

osa x: prvky H_{ii} , osa y: \hat{e}_{ji}

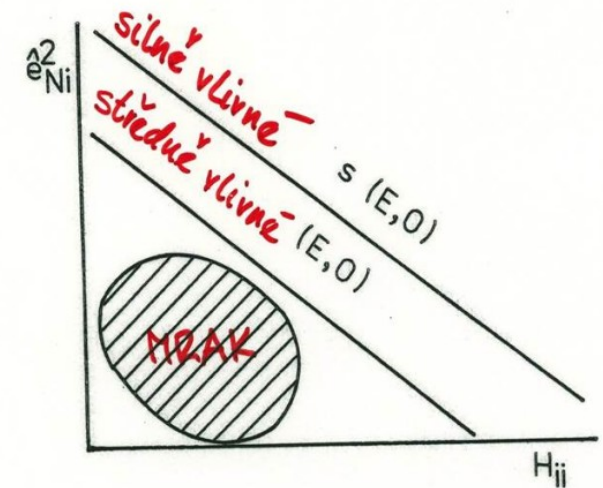


Williamsův graf vlivných bodů



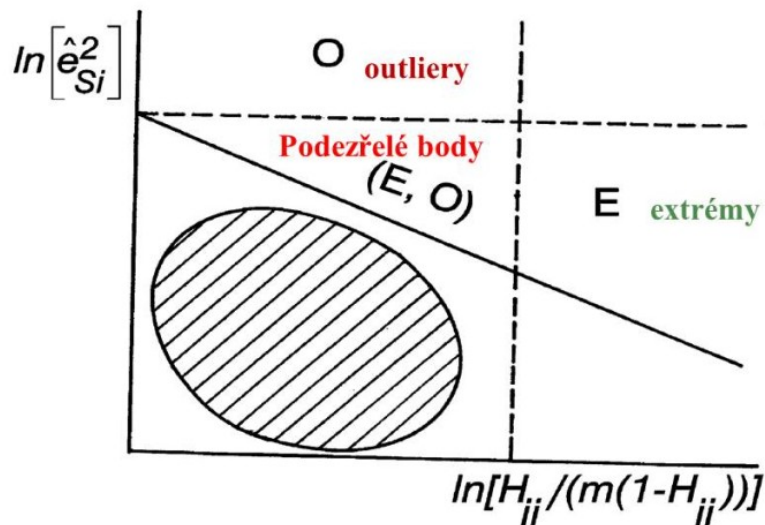
3. Pregibonův graf (PG),

osa x: prvky H_{ii} , osa y: \hat{e}_{Ni}^2

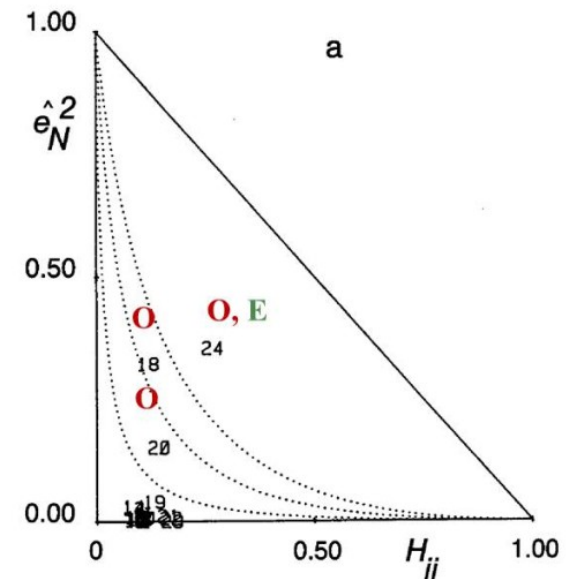


4. McCullohův-Meeterův graf (MMG)

osa x: $\ln[H_{ii}/(m(1-H_{ii}))]$, osa y: $\ln e_{Si}^2$



5. L-R graf (LR)

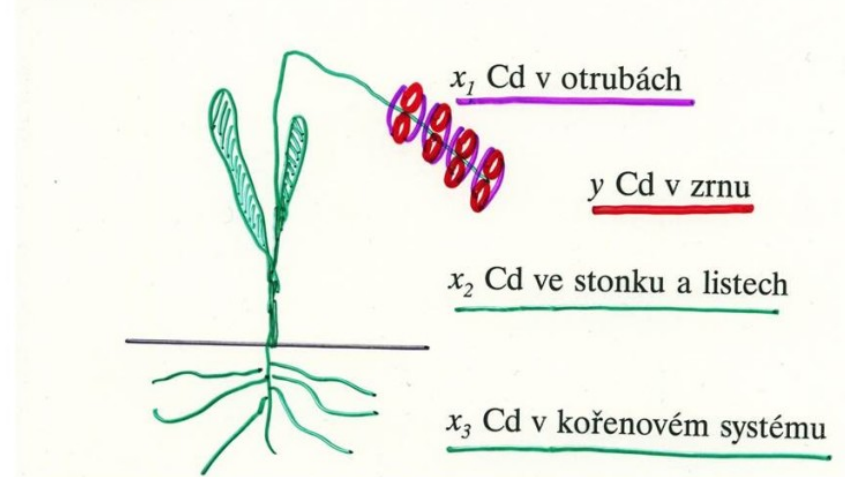


2. vzorová úloha na výstavbu lineárního regresního modelu

M619

21

Úloha M619. Vliv tří parametrů na obsah kadmia v potravinářské pšenici
Obsah kadmia v zrně y [mg/l] v závislosti na obsahu kadmia v otrubách x_1 [mg/l], ve stonku s listy x_2 [mg/l] a v kořenovém systému x_3 [mg/l].
Vyšetřete regresní triplet (data, model, metoda) a nalezněte lineární regresní model.



Výstavba lineárního regresního modelu:

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

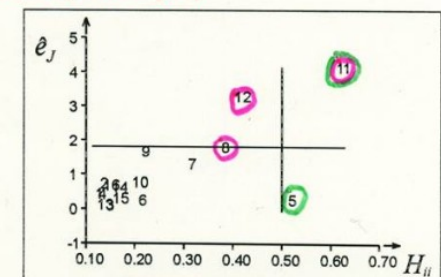
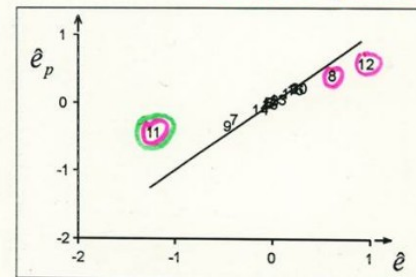
Kritika dat

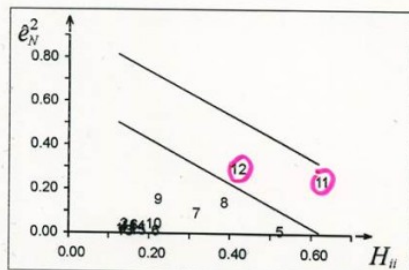
Kritika modelu

Kritika metody

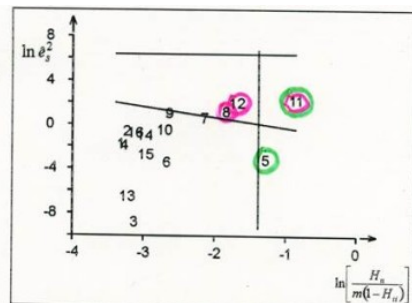
Grafy vlivných bodů (odlehých O a extrémů E)

Software QC-EXPERT 3.1 (TriloByte)

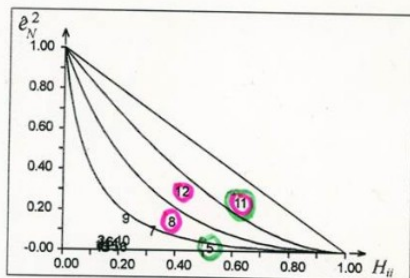




Pregibonův graf



McCulloh-Meeterův graf



L-R graf

LINEÁRNÍ REGRESE Regresní diagnostika

Název: m619

V S T U P

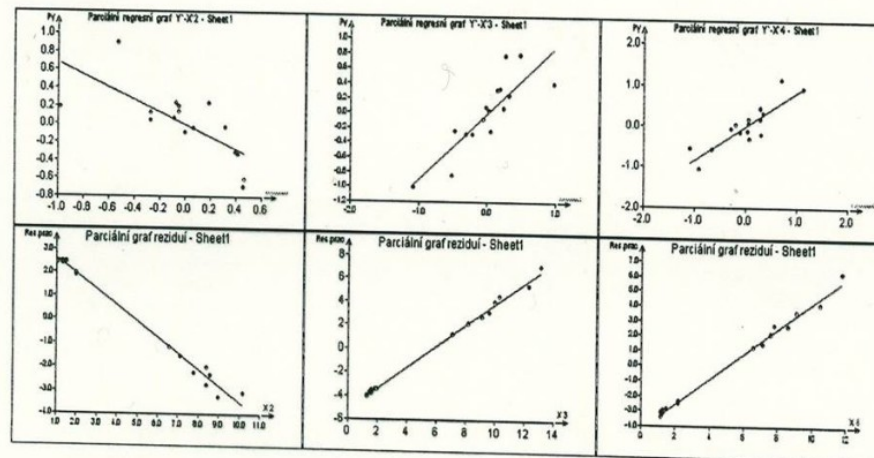
(1) ZVOLENÁ STRATEGIE REGRESNÍ ANALÝZY:

Omezení, P	: 1.0000E-34
Transformace	: Ne
Váhy	: Ne
Absolutní člen zahrnut	: Ano

(2) PODMÍNKY A KVANTILY PRO STATISTICKÉ TESTY:

Hladina významnosti, alfa	: 0.050
Počet bodů, n	: 16
Počet parametrů, m	: 3
Kvantil Studentova rozdělení t(1-alpha/2, n-m)	: 2.179
Kvantil rozd. Chí-kvadrát Chi-square(1-alpha, m)	: 9.488
Jméno výstupního souboru	: M619.txt

Parciální grafy



V Ý S T U P

(1) PŘEDBĚŽNÁ STATISTICKÁ ANALÝZA:

Proměnná	Průměr	Směrodatná odchylna	Párový korelační koeficient	Spočtená hladina výz.
y	6.0125E+00	4.8734E+00	1.0000	----
x1 OTRUBY	4.8937E+00	3.5692E+00	0.9837	0.000
x2 LISTY	5.7812E+00	4.5296E+00	0.9935	0.000
x3 KOŘEN	5.0812E+00	3.8782E+00	0.9948	0.000

Párové korelační koeficienty mezi dvojicemi vysvětlujících proměnných

x1 versus x2 :	9.9344E-01	Spočtená hladina výz.
x1 versus x3 :	9.8693E-01	0.000
x2 versus x3 :	9.8847E-01	0.000

(2) INDIKACE MULTIKOLINEARITY:

C	Vlastní čísla [j] korel. matice I[j]	Čísla podmíněnosti K[j]	Variance inflation factor VIF[j]	Vícenás. korel. koef pro X[j]
1	6.4568E-03	4.6141E+02	8.3272E+01	0.9940
2	1.4307E-02	2.0823E+02	9.4324E+01	0.9947
3	2.9792E+00	1.0000E+00	4.7508E+01	0.9894

Maximální číslo podmíněnosti K : 4.6141E+02
 (K[j], K > 1000 indikuje silnou multikolaritu)
 (VIF[j] > 10 indikuje silnou multikolaritu)

Kritika dat: původní data obsahují vlivné body 8, 11, 12

Kritika modelu $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$; technika MEP, AIC, $s(e)$ našla, že absolutní člen β_0 statisticky nevýznamný

Odhad parametru	Směrodatná odchylka	Student t-test $t_{0.95(16-4)} = 2.18$	Parameter je	Hladina významnosti
-----------------	---------------------	---	--------------	---------------------

A. Původní data a původní model vedou k závěru

$$y = -0.07(0.14) - 0.69(0.19)x_1 + 0.90(0.16)x_2 + 0.84(0.13)x_3$$

(LS: $t_{1-0.052}(16-4) = 2.179$, $D = 99.72\%$, $MEP = 0.21014$, $AIC = -36.18$, $s(e) = 0.290$)

β_0	-0.073	0.138	-0.5269	Nevýznamný	0.608
β_1	-0.685	0.192	-3.5746	Významný	0.004
β_2	0.896	0.161	5.5761	Významný	0.000
β_3	0.838	0.133	6.2879	Významný	0.000

B. Data bez 8, 11, 12 a původní model vedou k závěru

$$y = -0.02(0.10) - 0.80(0.50)x_1 + 0.92(0.34)x_2 + 0.91(0.15)x_3$$

(LS: $t_{1-0.052}(13-4) = 2.262$, $D = 99.83\%$, $MEP = 0.06118$, $AIC = -41.58$, $s(e) = 0.178$)

β_0	-0.017	0.103	-0.1637	Nevýznamný	0.874
β_1	-0.802	0.504	-1.5919	Nevýznamný	0.146
β_2	0.920	0.338	2.7193	Významný	0.024
β_3	0.906	0.152	5.9698	Významný	0.000

C. Data bez 11, 12 a opravený model vedou k závěru

$$y = -1.18(0.38)x_1 + 1.25(0.24)x_2 + 0.92(0.15)x_3$$

(LS: $t_{1-0.052}(14-3) = 2.201$, $D = 99.80\%$, $MEP = 0.06078$, $AIC = -43.47$, $s(e) = 0.193$)

β_0	0.000	---	---	---	---
β_1	-1.181	0.383	-3.0854	Významný	0.010
β_2	1.245	0.236	5.2751	Významný	0.000
β_3	0.917	0.151	6.091	Významný	0.000

D. Data bez 8, 11, 12 a opravený model vedou k závěru

$$y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$$

(LS: $t_{1-0.052}(13-3) = 2.228$, $D = 99.83\%$, $MEP = 0.05101$, $AIC = -43.55$, $s(e) = 0.170$)

β_0	0.000	---	---	---	---
β_1	-0.855	0.372	-2.3002	Významný	0.044
β_2	0.954	0.251	3.8038	Významný	0.003
β_3	0.916	0.132	6.9263	Významný	0.000

Model: $y = -0.86(0.37)x_1 + 0.95(0.25)x_2 + 0.92(0.13)x_3$

Úlohy na výstavbu

lineárního regresního modelu

Kritika dat

**Software QC-EXPERT 3.1,
ADSTAT 1.25**

Úloha M6.09 Závislost obsahu cholesterolu na věku a váze pacienta

Zadání: Výběr 25 pacientů, nemocných s hyperliproteinemií byl vyšetřován na hladinu lipidů v plasmě totálního cholesterolu y s přihlédnutím k váze x_1 a věku x_2 pacienta.

Úkoly: (1) Navrhnete lineární regresní model a testujete statistickou významnost jednotlivých regresních parametrů.

(2) Vyšetřete regresní triplet a soustředte se na odhalení vlivných bodů. (3) Jaké závěry můžete učinit z parciálních regresních a parciálních reziduálních grafů?

Data: Váha x_1 [kg] a věk pacienta x_2 [roky], obsah cholesterolu y [mg/100 ml]:

x_1	x_2	y
84	46	354
...
63	30	244

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	77,98253861	52,42963881	Nevýznamný	0,15110913	-30,74987727	186,7149545
M609x1	0,4173620988	0,7287760772	Nevýznamný	0,5726622864	-1,09402698	1,928751178
M609x2	5,216590809	0,7572444634	Významný	6,430143229E-007	3,64616191	6,787019707

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,8399704181
Koeficient determinace R ² :	0,7055503032
Predikovaný korelační koeficient Rp :	0,4190010579
Střední kvadratická chyba predikce MEP :	2050,962356
Akaického informačního kritérium :	192,139075

Reziduální součet čtverců :	42806,22535
Průměr absolutních reziduí :	33,08966795
Reziduální směr. odchylka :	44,1105148
Reziduální rozptyl :	1945,737516
Šikmost reziduí :	0,1318140961
Špičatost reziduí :	2,494872789

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu	
Hodnota kritéria F :	26,35782417
Kvantil F (1-alfa, m-1, n-m) :	3,443356779
Pravděpodobnost :	1,442521271E-006
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,2252724112
Závěr :	Model je korektní.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	0,0345357629
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,8525718025
Závěr :	Rezidua vykazují homoskedasticitu.

Jarque-Berrův test normality

Hodnota kritéria JB :	0,8150102957
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0,6653080265
Závěr :	Rezidua mají normální rozdělení.

Waldův test autokorelace

Hodnota kritéria WA :	0,9210366416
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,3372029312
Závěr :	Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW :	-1
Kritické hodnoty DW :	1,12
Závěr :	Positivní autokorelace reziduí není prokázána.

Znaménkový test reziduí

Hodnota kritéria Sg :	0,6396021491
Kvantil N(1-alfa/2) :	1,959963999
Pravděpodobnost :	0,522431285
Závěr :	V reziduích není trend.

33

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	0,2800623685	0,07576827835	Významný	0,0004308755135	0,1289473023	0,4311774348
M615x1	2,180116568	0,6005895517	Významný	0,0005353606507	0,9822784771	3,377954659
M615x2	-1,741947387	0,752351329	Významný	0,02353072646	-3,242464799	-0,2414299753
M615x3	7,178149993	0,3830484857	Významný	0	6,414183878	7,942116109

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9587042693
Koeficient determinace R ² :	0,919113876
Predikovaný korelační koeficient Rp :	0,2411434818
Střední kvadratická chyba predikce MEP :	0,6122282744
Akaického informačního kritérium :	-164,4148787

Reziduální součet čtverců :	7,200383586
Průměr absolutních reziduí :	0,190713291
Reziduální směr. odchylka :	0,3207220333
Reziduální rozptyl :	0,1028626227
Šikmost reziduí :	3,754928183
Špičatost reziduí :	14,22533139

Testování regresního tripletu

Fisher-Snedecorův test významnosti modelu	
Hodnota kritéria F :	265,1380656
Kvantil F (1-alfa, m-1, n-m) :	2,735541451
Pravděpodobnost :	3,862813994E-038
Závěr :	Model je významný

Scottovo kritérium multikolinearity

Hodnota kritéria SC :	0,4689878624
Závěr :	Model vyžaduje multikolinearit!

Jarque-Berrův test normality

Hodnota kritéria JB :	434,8356476
Kvantil Chi ² (1-alfa,2) :	5,991464547
Pravděpodobnost :	0
Závěr :	Rezidua nemají normální rozdělení!

Waldův test autokorelace

Hodnota kritéria WA :	0,3860176122
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	0,5344001773
Závěr :	Autokorelace je nevýznamná

Durbin-Watsonův test autokorelace

Hodnota kritéria DW :	-1
Kritické hodnoty DW :	1,49
Závěr :	Rezidua nejsou autokorelována

Znaménkový test reziduí

Hodnota kritéria Sg :	1,710237181
Kvantil N(1-alfa/2) :	1,959963999
Pravděpodobnost :	0,08722202288
Závěr :	V reziduích není trend.

Cook-Weisbergův test heteroskedasticity

Hodnota kritéria CW :	23,57564012
Kvantil Chi ² (1-alfa,1) :	3,841458829
Pravděpodobnost :	1,20098341E-006
Závěr :	Rezidua vykazují heteroskedasticitu!

35

Úloha M6.15 Vliv obsahu PCB v kongenerech na PCB v mateřském mléce

Zadání: V tuce mateřského mléka 74 matek byl stanoven obsah polychlorovaných bifenylů, a to celkový obsah suma PCB y, a jednak jako obsah PCB jednotlivých kongenerů x₁, x₂, x₃.

Úkoly: (1) Vyšetřete lineární regresní model mezi sumou PCB y a obsahem kongenerů PCB#138 x₁, PCB#153 x₂ a PCB#180 x₃

(2) Posuďte predikční schopnost regresního modelu a spolehlivost predikce celkového obsahu PCB na základě znalosti tří kongenerů.

(3) Vyšetřete regresní triplet a testujte statistickou významnost jednotlivých kongenerů.

Data: Obsah PCB#138 x₁ [mg/kg], obsah PCB#153 x₂ [mg/kg], obsah PCB#180 x₃ [mg/kg], suma PCB y [mg/kg].:

x ₁	x ₂	x ₃	y
0.184	0.218	0.228	1.737
...
0.107	0.186	0.081	1.031

34

Úloha M6.28 Vliv čtyř parametrů na práci levé srdeční komory

Zadání: Byla sledována závislost práce levé srdeční komory LVSWI y na čtyřech parametrech, a to srdečním indexu CI x₁, pulmonární vaskulární rezistenci PVRI x₂, systémové vaskulární rezistenci SVRI x₃ a spotřebě kyslíku VO₂ x₄.

Úkoly: (1) Vyšetřete statistickou významnost jednotlivých regresních parametrů a regresního tripletu.

(2) Jsou v datech vlivné body?

(3) Který z vyšetřovaných regresních parametrů vyšel jako statisticky významnější? (4) Popište pět nejdůležitějších grafů identifikace vlivných bodů a uveďte obecné tvary diagnostických grafů při této analýze.

Data: Srdeční index CI x₁, pulmonární vaskulární rezistence PVRI x₂, systémová vaskulární rezistence SVRI x₃, spotřeba kyslíku VO₂ x₄, práce levé srdeční komory LVSWI y:

x ₁	x ₂	x ₃	x ₄	y
4.6	211	880	367	20
...
6.2	206	852	555	43

36

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-62,56460697	16,15861044	Významný	0,003777768158	-99,11792333	-26,01129061
M628x1	14,04140648	2,543061089	Významný	0,0003697866064	8,288602621	19,79421034
M628x2	-0,05322916434	0,036538729	Nevýznamný	0,1791605309	-0,135885514	0,02942718536
M628x3	0,03119674024	0,009953979	Významný	0,01204195425	0,008679275055	0,05371420542
M628x4	0,004647488755	0,01304358971	Nevýznamný	0,7298224832	-0,02485916114	0,03415413865

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9431418891
Koeficient determinace R ² :	0,8895166231
Predikovaný korelační koeficient Rp :	0,4276571259
Střední kvadratická chyba predikce MEP :	16,89264631
Akaikeho informační kritérium :	33,59244458

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-68,69128087	15,10384788	Významný	0,0008329552804	-101,9346259	-35,44793583
M628x1	14,09921966	1,717918901	Významný	5,118446609E-006	10,31810566	17,88033367
M628x3	0,02768396296	0,009477917984	Významný	0,006823206129	0,01391529687	0,006823206129
	0,04854471979					

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9285532055
Koeficient determinace R ² :	0,8622110554
Predikovaný korelační koeficient Rp :	0,4700587757
Střední kvadratická chyba predikce MEP :	15,34744665
Akaikeho informační kritérium :	32,68445731

37

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-193,6253449	68,65424579	Významný	0,01545143086	-343,2100964	-44,04059335
M636x1	1,402213808	0,5296908562	Významný	0,02128764253	0,2481165743	2,556311041
M636x2	0,7723109038	0,2027625676	Významný	0,002489006667	0,3305292202	1,214092587
M636x3	1,047758688	0,1360194526	Významný	5,526372972E-006	0,7513977598	1,344119616
M636x4	-0,1244881671	0,1732033206	Nevýznamný	0,4860614488	-0,5018657841	0,25288945
M636x5	0,07188017742	0,1305649581	Nevýznamný	0,5920557687	-0,2125964283	0,3563567832
M636x6	0,09177697575	0,1626338519	Nevýznamný	0,58293291	-0,2625717474	0,4461256989
M636x7	-0,1067234527	0,156238526	Nevýznamný	0,5075301603	-0,4471379577	0,2336910523

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9568918648
Koeficient determinace R ² :	0,915642041
Predikovaný korelační koeficient Rp :	0,6108492984
Střední kvadratická chyba predikce MEP :	10,30682204
Akaikeho informační kritérium :	43,62805288

Odhady parametrů

Proměnná	Odhad	Směr.Odch.	Závěr	Pravděpodobnost	Spodní mez	Horní mez
Abs	-199,7222351	33,79003067	Významný	2,19399367E-005	-271,3539002	-128,0905701
M636x1	1,372797085	0,4510266936	Významný	0,007739757285	0,4166632076	2,328930963
M636x2	0,688085446	0,1615609509	Významný	0,0005998377246	0,3455915299	1,030579362
M636x3	1,10591789	0,09942008578	Významný	6,122320428E-009	0,8951567232	1,316679057

Statistické charakteristiky regrese

Vicenasobný korelační koeficient R :	0,9511133377
Koeficient determinace R ² :	0,9046165812
Predikovaný korelační koeficient Rp :	0,7185205855
Střední kvadratická chyba predikce MEP :	7,188446076
Akaikeho informační kritérium :	38,08476477

39

Úloha M6.36 Dědičné vlivy na výšku 18letých mladíků

Zadáni: Byla vyšetřována závislost výšky dvaceti 18letých mladíků y na výškách $x_1 \dots$ až x_7 jejich rodičů a obou prarodičů, žijících izolovaně v horské vesnici po několik generací.

Úkoly:

- (1) Postavte lineární regresní model a testujte statistickou významnost jednotlivých regresních parametrů β_0, \dots, β_7 .
- (2) Rozhodněte mezi dvěma navrženými modely: Model A: $y = f(\beta_0, \dots, \beta_7)$ a model B: $y = f(\beta_0, \dots, \beta_3)$.
- (3) Vypočítejte oboustranný 95% interval spolehlivosti všech odhadnutých regresních parametrů a vysvětlete význam intervalu spolehlivosti pro β_1 .
- (4) Predikujte výšku 18letého mladíka z dat jeho rodičů a prarodičů: $x_1 = 50.8$ cm, $x_2 = 152.4$ cm, $x_3 = 182.9$ cm, $x_4 = 154.9$ cm, $x_5 = 180.3$ cm, $x_6 = 157.7$ cm, $x_7 = 177.8$ cm.

Data: x_1 porodní délka chlapce, x_2 výška matky v jejím věku 18 let, x_3 výška otce v jeho věku 18 let, x_4 výška babičky z matčiny strany v jejím věku 18 let, x_5 výška dědečka z matčiny strany v jeho věku 18 let, x_6 výška babičky z otcovy strany v jejím věku 18 let, x_7 výška dědečka z otcovy strany v jeho věku 18 let, výška 18letého chlapce y [cm]:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
	50.0	153.7	178.6	166.9	176	166.9	170.9	170.7
...
	48.8	158.0	170.9	161.5	180.1	161.5	169.4	167.9

38

40