Lineární regrese

Menu:	QCExpert	Lineární regrese

Modul lineární regrese slouží pro tvorbu a analýzu lineárních regresních modelů v obecném tvaru

$$G(y) = a_1F_1(\mathbf{x}) + a_2F_2(\mathbf{x}) + \ldots + a_mF_m(\mathbf{x}) + a_0, \qquad \textbf{(1-1)}$$

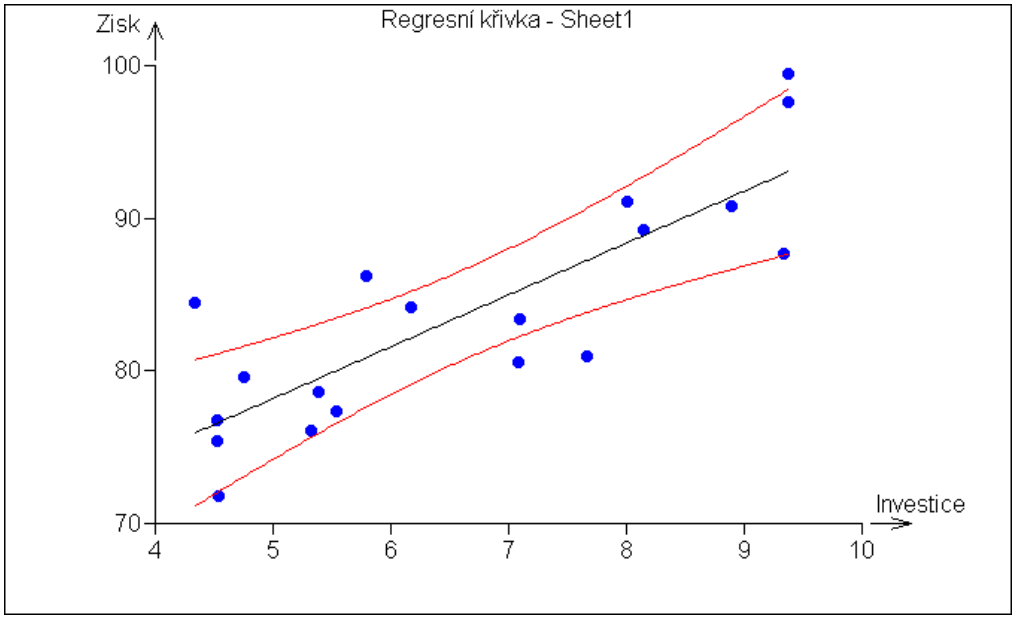
kde y je nezávisle proměnná, $\mathbf{x} = (x_1, x_2, \ldots x_p)$ jsou nezávisle proměnné, p je počet proměnných, $\mathbf{a} = (a_1, a_2, \ldots, a_m)$ jsou parametry, m je počet parametrů, a_0 je absolutní člen, $F_i(\mathbf{x})$ je libovolná funkce nezávisle proměnné (nikoli parametrů) a funkce $G(y)$ je libovolná funkce závisle proměnné. Výraz $F_i(\mathbf{x})$ se také nazývá člen lineárního modelu. Předpokládá se, že **x** je pokud možno deterministická (nenáhodná) nezávisle nastavená nebo jinak zjištěná veličina. Veličina y na **x** závisí, ale její hodnota je zatížena náhodnou chybou ε. Parametry a se odhadují na základě dat a daného modelu zvolenou robustní nebo nerobustní metodou. Tato trojice okolností (data, model, metoda) se někdy označuje jako regresní triplet a každé je třeba věnovat stejnou pozornost, chceme-li dosáhnout korektních výsledků. K tomu slouží uživateli bohatá regresní diagnostika a bohatý výběr metod a dalších nástrojů. Uživatel může zvolit tři základní možnosti modelů: prostý lineární model bez transformace, polynom, nebo obecný uživatelem definovaný model. V základním dialogovém panelu lineární regrese (Obrázek 3) lze zvolit tři tvary modelu v poli Transformace:

Bez transformace: regresní model je ve tvaru

$$y = a_1x_1 + a_2x_2 + \ldots + a_mx_m + a_0, \qquad \textbf{(1-2)}$$

kde počet parametrů m je daný počtem nezávisle proměnných vybraných v okénku *Nezávisle proměnná*. Nejjednodušším příkladem takového modelu je regresní přímka, např.
[zisk] = a_1 . [investice] + a_0,
nebo vícerozměrná závislost typu
[pevnost_oceli] = a_1 . [obsah_Cr] + a_2 . [doba_žíhání] + a_3 . [obsah_uhlíku] + a_0,

Například:



Obrázek 1 Regresní přímka

Polynom je model ve tvaru

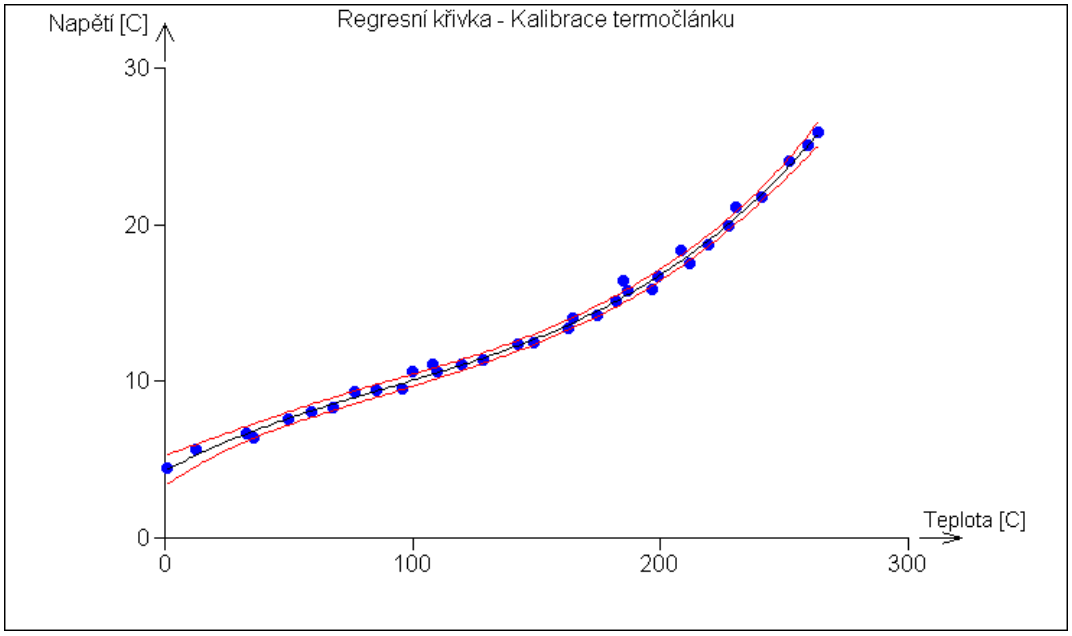
$$y = a_1 x + a_2 x^2 + \ldots + a_m x^m + a_0, \qquad \textbf{(1-3)}$$

kde m je současně stupeň polynomu i počet parametrů. Tento model obsahuje pouze jednu nezávisle proměnnou, která se v něm však vyskytuje v různých mocninách, jsou zde vždy všechny mocniny od 1 do m. Příkladem je regresní parabola vyjadřující nelineární závislost

$$[\text{obrat}] = a_0 + a_1 \,.\, [\text{náklad\_na\_reklamu}] + a_2 \,.\, [\text{náklad\_na\_reklamu}]^2 + a_0.$$

Pokud chceme do modelu zahrnout pouze třeba první a třetí mocninu, nebo vytvořit jiný obecnější model, musíme použít uživatelskou transformaci.

Například:



Obrázek 2 Regresní polynom 3. stupně

QCExpert™ umožňuje rovněž polynomickou transformaci pro více nezávisle proměnných, viz odst. 0.

Uživatelská transformace: zde můžeme vytvořit obecný lineární model obecného tvaru (1-1), který zahrnuje i oba předchozí (*bez transformace* a *polynom*). Byly-li uživatelem nějaké modely již dříve vytvořeny, lze je vybrat v okénku pod volbou. Jinak se po stisknutí tlačítka *Model…* (toto tlačítko je aktivní jen při vybrané položce *Uživatelská…*) otevře dialogový panel pro tvorbu modelu (Obrázek 4, viz dále). Zde lze zadat jednotlivé transformační funkce z rovnice (1-1) F_1, F_2, …, případně G. Příkladem uživatelské transformace je linearizace exponenciálního modelu $y = A \,.\, \exp(B\,x)$ na tvar $\ln(y) = a + b\,x$, kde $a = \ln A$, $b = B$. V tomto případě je $G = \ln(y)$, $F_1 = x$. Jiným příkladem je třeba smíšený model

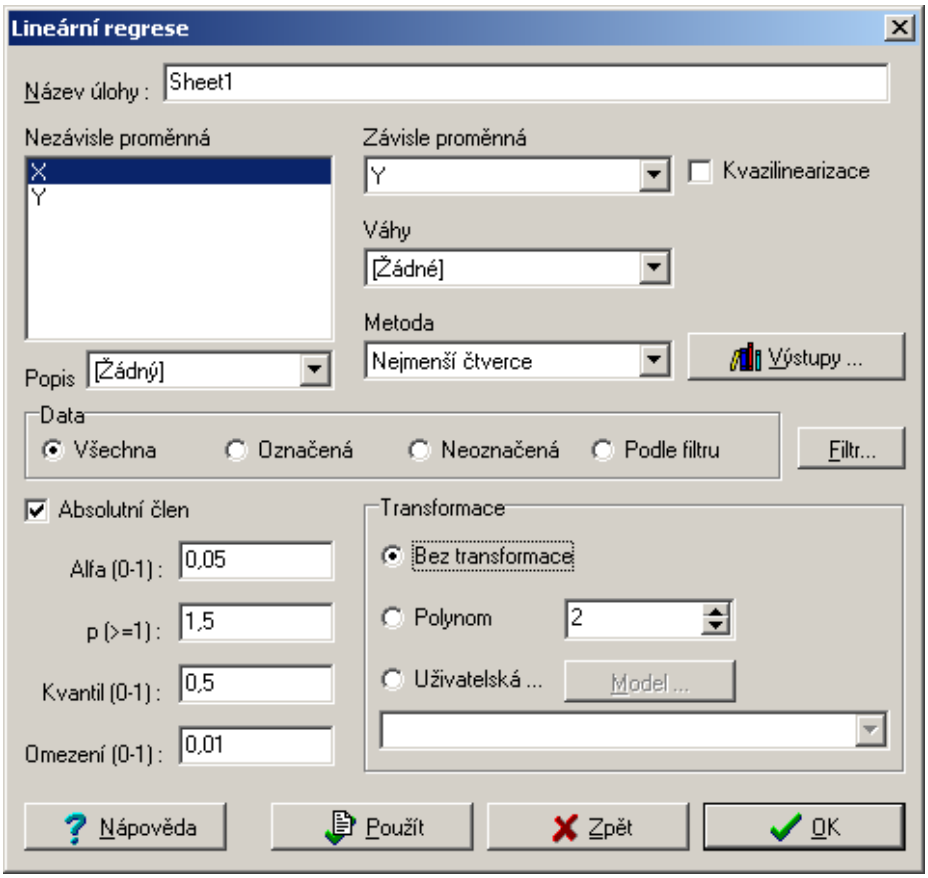
$$1 / [\text{spotřeba}] = a_1\,[\text{X1}] + a_2.[\text{X1}]^{1/2} + a_3\,[\text{X1}].[\text{X2}] + a_4 \ln[\text{X2}] + a_0,$$

kde je
$G = 1/[\text{spotřeba}]$,
$F_1 = [\text{X1}]$,
$F_2 = \text{sqrt}[\text{X2}]$,
$F_3 = [\text{X1}][\text{X2}]$,
$F_4 = \ln [\text{X2}]$,
[spotřeba] je závisle proměnná, [X1] a [X2] jsou závisle proměnné.
Transformační funkce nemohou obsahovat parametry a_i, pouze proměnné, nelze tedy vytvořit například model $y = a_1 \,.\, x^{a_2} + a_0$, nebo $y = a_0 + a_1 \,.\, \exp(a_2\,x)$.

Podle povahy dat, chyb, nebo jiných požadavků lze zvolit vhodnou robustní nebo nerobustní metodu regrese, popřípadě zadat váhy pro jednotlivá data. Rovněž lze využít metodu krokové regrese (*stepwise*) pro výběr nejvhodnějších proměnných nebo nejlepší transformace.

Data a parametry

Výpočet hodnot parametrů $\mathbf{a} = (a_1, a_2, \ldots, a_m)$ probíhá na základě dat. Ta jsou uspořádána ve sloupcích datové tabulky. Sloupec reprezentuje hodnoty jedné proměnné. K identifikaci proměnných se používá záhlaví sloupce. Výběr proměnných závisí na druhu transformace.



Obrázek 3 Základní dialogový panel pro Lineární regresi

Bez transformace: po prvním otevření dialogového panelu pro lineární regresi (Obrázek 3) se automaticky zvolí všechny sloupce s daty kromě posledního jako nezávisle proměnná, poslední sloupec jako závisle proměnná. Pokud chce uživatel jiný výběr proměnných, vybere je myší s případným použitím kláves Shift a Ctrl. Počet nezávisle proměnných není omezen, závisle proměnná je vždy jedna. Takto definovaný model má pak obecný tvar (1-2), kde data ve sloupci závisle proměnné představují y a sloupce vybrané jako nezávisle proměnné odpovídají proměnným x_1, x_2, …

Polynom: polynomická transformace ve tvaru (1-3) nabízí proložení dat polynomickou křivkou. Tato transformace je určena pouze pro jednu nezávisle a jednu závisle proměnnou. Zadává se stupeň polynomu, p_P. Doporučuje se používat spíše nižších stupňů polynomu, polynomy vyššího stupně mají sklon k numerické nestabilitě, která se projevuje silným „rozkmitáním“ křivky a špatnou predikční schopností. K orientačnímu určení rozumného stupně polynomu lze použít metody stepwise, viz dále. Polynomická transformace zahrne do modelu vždy všechny mocniny od 1. až do zvoleného stupně.

Chceme-li zahrnout do modelu pouze například 1., 3. a 5. mocninu, musíme zvolit uživatelskou transformaci a příslušný model zadat ručně.

Je-li vybráno více nezávisle proměnných, je políčko *Stupeň polynomu* neaktivní a provede se úplný rozvoj do druhého stupně Taylorova polynomu s proměnnými

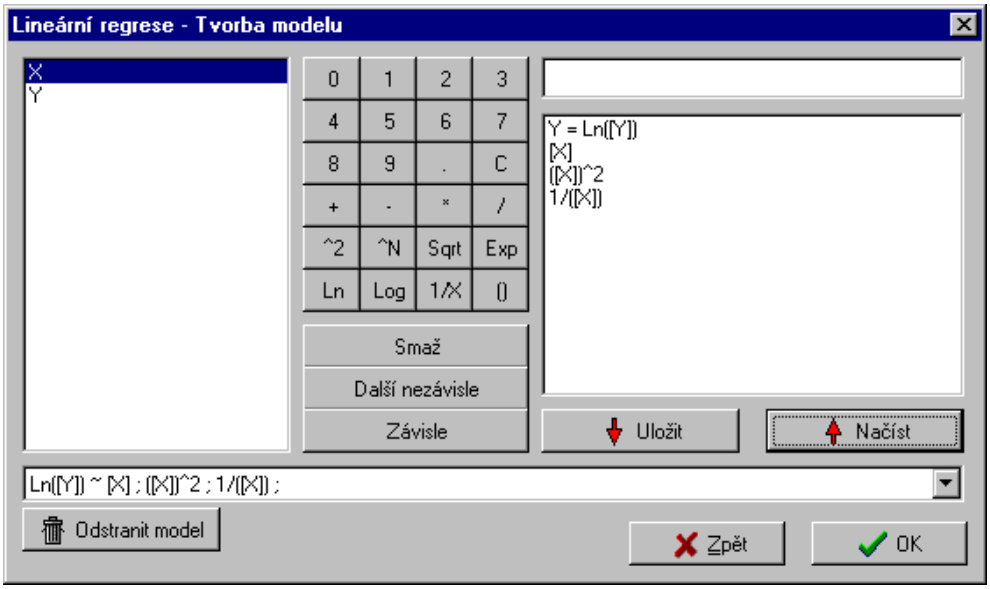
$$x_1, x_2, \ldots, x_m, x_1x_2, x_1x_3, \ldots x_ix_j, \ldots, x_mx_{m-1}, x_1^2, x_2^2, \ldots, x_m^2 \qquad \textbf{(1-4)}$$

Máme-li tedy označeny proměnné A, B, C, bude mít model s absolutním členem tvar

$$y = a_0+a_1.A+a_2.B+ a_3.C+ a_4.AB+ a_5.AC+ a_6.BC+ a_7.A^2+ a_8.B^2+ a_9.C^2$$

Takový model představuje kvadratický $m-1$ rozměrný povrch, kterým se snažíme popsat data. Tento povrch může mít jeden extrém (minimum nebo maximum), který odpovídá předpokládané nejvyšší nebo nejnižší hodnoty závisle proměnné. Tato problematika je ve zkrácené podobě (bez další diagnostiky) řešena i v modulu Optimalizace. Je třeba mít na paměti, že počet dat musí být (pokud možno mnohem) větší než počet proměnných a že při použití Taylorova rozvoje je pro m původních proměnných výsledný počet proměnných v modelu s absolutním členem roven $1.5m+m^2/2+1$, což pro m=10 je 66. V případě volby Taylorova polynomu jsou výstupu specifické výsledky uvedeny v odstavci *Analýza responsního povrchu*.

Uživatelská: po volbě uživatelské transformace lze tlačítkem *Model…* otevřít panel pro tvorbu modelu. Pokud jsme již dříve nějaké modely vytvořili, je možné z nich jeden pouze vybrat bez otevření panelu *Tvorba modelu*, je ale nutno dbát, aby se shodovaly názvy proměnných v modelu a v tabulce s daty. Panel *Tvorba modelu* (Obrázek 4) nabízí v levé části seznam proměnných v aktuálním listu tabulky s daty, z nichž se tvoří jednotlivé členy regresního modelu. V pravé části nahoře je editační řádek, kde se člen sestavuje a pod ním okénku, v němž je seznam členů právě tvořeného modelu a také závisle proměnná.



Obrázek 4 Dialogový panel pro tvorbu modelu

Členem může být libovolná funkce obsahující jednu nebo více nezávisle proměnných. Závisle proměnná je jediná uživatelem zvolená proměnná nebo její funkce. Závisle proměnná se v seznamu označí symbolem Y=. Příklad (Obrázek 4) definuje model $\ln(y) = A.x + B.x^2 + C.x^{-1}$. Absolutní člen lze do modelu přidat buď zatrhnutím okénka *Absolutní člen* (Obrázek 3), nebo přidáním 1 (jedničky) jako nezávisle proměnné v modelu. (Nelze zvolit obě možnosti zároveň, neboť by došlo k chybě způsobené přeurčeností modelu.)

Pokyny pro sestavování modelu:
Dvojitým kliknutím na proměnnou v seznamu proměnných opíšeme tuto proměnnou do editačního řádku. Název proměnné se uvádí vždy v hranatých závorkách. Při psaní složitějších výrazů je možno výhodně použít pomocných tlačítek s funkcemi. Je-li v editačním řádku označena část výrazu, stisknutím tlačítka funkce se tato funkce aplikuje na označenou část. Například výraz ln([*x*]+1) sestavíme takto: dvojitým kliknutím přepíšeme proměnnou *x* (v datech musí být sloupec tohoto jména): **[x]**; připíšeme + 1; celý výraz označíme: **[x]+1**; a klikneme na tlačítko *Ln*, výsledkem bude: **ln([x]+1)**. Podobně použijeme tlačítka *^2, ^A, Sqrt, Exp, Log, 1/X, ( )*. Tlačítko *C* smaže editační řádek. Další funkce je nutno psát ručně, seznam funkcí uvádí **Tabulka** 1. Po napsání členu klikneme na tlačítko *Další nezávisle*, nebo *Závisle* a tím přidáme člen do seznamu členů modelu v pravé části panelu. Tlačítkem *Smaž* vymažeme označený člen modelu. Model musí mít vždy jednu závisle proměnnou. Když je model sestaven, tlačítkem *Uložit* jej uložíme do seznamu modelů ve spodní části panelu. Tlačítkem *Načíst* načteme aktuální model ze seznamu modelů a můžeme jeho členy modifikovat. Tlačítkem *Odstranit model* vymažeme aktuální model v seznamu modelů: pozor, tuto operaci nelze vrátit zpět! Tlačítkem *OK* sestavení modelu ukončíme.

Hotové modely můžeme ze seznamu modelů vybírat přímo v hlavním panelu *Lineární regrese* bez otevření panelu *Tvorba modelu*, pozor na souhlas názvů proměnných.

Tabulka **1** Přehled funkcí

Funkce	Hodnota, popis, omezení	Syntaxe
Základní binární operátory		
+	Sčítání	x+y
–	Odčítání	x–y
*	Násobení	x*y
/	Dělení; y ≠ 0	x/y
^	Umocnění; pro záporné x je třeba použít funkci INTPOWER	x^y
DIV	Celočíselné dělení; y ≠ 0	x DIV y
MOD	Zbytek po dělení; y ≠ 0	x MOD y
Funkce		
TAN	Tangens; x ≠ *n*π+π/2	tan(x)
SIN	Sinus	sin(x)
COS	Kosinus	cos(x)
SINH	Hyperbolický sinus	sinh(x)
COSH	Hyperbolický kosinus	cosh(x)
ARCTAN	Arcus tangens	arctan(x)
COTAN	Kotangens; x ≠ *n*π	cotan(x)
EXP	Exponenciální funkce se základem e	exp(x)
LN	Přirozený logaritmus; x > 0	ln(x)
LOG	Dekadický logaritmus; x > 0	log(x)
LOG2	Logaritmus se základem 2; x > 0	log2(x)
SQR	Druhá mocnina	sqr(x)
SQRT	Druhá odmocnina; x ≥ 0	sqrt(x)
ABS	Absolutní hodnota (abs(0) = 0)	abs(x)
TRUNC	Uříznutí desetinné části čísla	trunc(x)
INT	Uříznutí desetinné části čísla	int(x)
CEIL	Nejmenší celé číslo větší než argument	ceil(x)
FLOOR	Největší celé číslo menší než argument	floor(x)
HEAV	Heavisidův operátor (0 pro záporný argument, 1 jinak)	heav(x)
SIGN	Znaménko (-1 pro záporný argument, 0 pro 0, 1 pro kladný	sign(x)

	argument)	
ZERO	Pro nulový argument 1, jinak 0	zero(x)
RND	Náhodné číslo z rovnoměrného rozdělení od 0 do x; x > 0	rnd(100)
RANDOM	Náhodné číslo z rovnoměrného rozdělení od 0 do 1 nezávisle na argumentu, který však musí být formálně uveden.	random(0)
–	(Unární) minus před výrazem	–x

Binární funkce (se dvěma argumenty)		
MAX	Větší ze dvou čísel	MAX(x,y)
MIN	Menší ze dvou čísel	MIN(x,0)
INTPOWER	První argument umocněný na druhý celočíselný argument; lze použít i pro záporné x	INTPOWER(x, –2)
LOGN	Logaritmus prvního argumentu se základem druhého argumentu; x > 0, y > 1	logn(x,3)

Relační funkce		
GT	Větší (*greater than*); Je-li x > y 1, jinak 0	GT(x,y)
LT	Menší (*less than*); Je-li x < y 1, jinak 0	LT(x,y)
EQ	Rovno (*equal*); Je-li x = y 1, jinak 0	EQ(x,y)
NE	Nerovno (*not equal*) ; Je-li x ≠ y 1, jinak 0	NE(x,y)
GE	Větší nebo rovno (*greater or equal*); Je-li x ≥ y 1, jinak 0	GE(x,y)
LE	Menší nebo rovno (*less or equal*); Je-li x ≤ y 1, jinak 0	LE(x,y)

Funkce lze psát malými nebo velkými písmeny. Výsledkem relačních funkcí je 0 nebo 1, čehož lze využít například k zápisu skokových funkcí jako le(x,0)*1+gt(x,0)*5, viz též Nelineární regresi.

Další popis dialogového panelu Lineární regrese (Obrázek 3):

Název úlohy: Jednořádková identifikace úlohy, která se tiskne v hlavičce protokolu a všech grafů.

Nezávisle proměnná: Vyberte jednu nebo více nezávisle proměnných, pro výběr více proměnných použijte tažení myší, Shift-klik, nebo Ctrl-klik. Při vybrané uživatelské transformaci je tato položka neaktivní, závisle proměnná se definuje přímo v okně Tvorba modelu.

Závisle proměnná: Vyberte jednu závisle proměnnou. Při vybrané uživatelské transformaci je tato položka neaktivní, závisle proměnná se definuje přímo v okně Tvorba modelu.

Absolutní člen: Zatrhnutím okénka přidáte k modelu absolutní člen. Pokud je v uživatelském modelu již ručně zadán absolutní člen v podobě jednotkové nezávisle proměnné, toto okénko nezatrhujte!

Alfa (0 – 1): Hladina významnosti α pro všechny testy a intervaly spolehlivosti. Musí být větší než 0 a menší než 1. Implicitně α=0.05.

p (p ≥ 1): Koeficient *p* pro L_p regresi. Tato hodnota se použije je-li vybrána metoda *Lp-regrese* (viz níže). Hodnota *p*=1 odpovídá metodě nejmenších absolutních odchylek, *p*=2 odpovídá metodě nejmenších čtverců, $p \to \infty$ (v praxi stačí $p \approx 10$) odpovídá metodě nejmenší maximální chyby (minimax), hodnoty *p* od 1 do 2 (přesněji: $1 \le p < 2$) vykazují robustnost vůči odlehlým hodnotám. Implicitně je *p*=1.5

Kvantil (0 – 1): Hodnota pravděpodobnosti pro kvantilovou regresi. Používá se v metodě *Kvantilová regrese* (viz níže). Musí být větší než 0 a menší než 1. Implicitní hodnota je 0.5, což odpovídá metodě nejmenších absolutních odchylek.

Omezení (0 – 1): Parametr omezení na vlastní čísla se vztahuje k metodě Korekce hodnosti. Nulová hodnota tohoto parametru odpovídá obyčejné metodě nejmenších čtverců, hodnoty větší než nula potlačí komponenty vzniklé rozkladem na vlastní čísla a vlastní vektory odpovídající nejmenším vlastním číslům. Výsledkem jsou (vychýlené) odhady parametrů s nižším rozptylem, méně citlivé na špatnou podmíněnost $\mathbf{X}^T\mathbf{X}$, typickou např. pro polynomy vyššího stupně, viz níže. Doporučuje se hodnota nejvýše kolem 0.1.

Kvazilinearizace: Je-li toto políčko zatrhnuté, provádí se kvazilinearizace, která má význam při uživatelské transformaci, je-li závisle proměnná nelineární funkcí jedné původní proměnné z datové tabulky, tedy například model **`ln(y) ~ [x]; [x]^2`**. Nelineární transformací $G(y)$ dochází k deformaci rozdělení chyb a zkreslení odhadu parametrů. Technikou kvazilinearizace se toto zkreslení z velké části eliminuje. Podstata kvazilinearizace je v zavedení vah $w_i=[\partial G(y)/\partial y]^{-1}$.

Váhy: Vyberte sloupec vah w_i v tabulce s daty nebo zadejte typ vah [Žádné], [Y], popř. [1/Y]. Posledně jmenované váhy se používají tehdy, má-li závisle proměnná konstantní relativní chybu. Váhy nesmějí být záporné. Nulová hodnota váhy znamená, že se příslušný řádek nebere při výpočtu v úvahu. V základním nastavení jsou všechny váhy jednotkové (*Žádné váhy*). Známe-li rozptyly jednotlivých hodnot závisle proměnné, měly by být váhy rovny odmocnině z převrácené hodnoty těchto rozptylů, diagonální kovarianční matice závisle proměnné by pak byla $\mathbf{S} = \mathrm{diag}(w_1^{-2}, w_2^{-2}, \ldots, w_n^{-2})$.

Metoda: Zvolte metodu výpočtu. Volba vhodné metody závisí především na povaze dat.

Nejmenší čtverce: Základní metoda založená na předpokladu normality chyb, nepřítomnosti hrubých chyb (vybočujících hodnot závisle proměnné), nepřítomnosti odlehlých měření (vybočujících hodnot nezávisle proměnné) a dobré podmíněnosti dat. Tato metoda je nevhodná není-li kterýkoliv ze jmenovaných předpokladů splněn.

Korekce hodnosti: Metoda vhodná pro polynomy vyššího stupně, Taylorovy polynomy a data se silnou kolinearitou („korelací“) mezi sloupci nezávisle proměnné, která je indikována jako multikolinearita v odstavci *Indikace multikolinearity* v protokolu. Míra korekce hodnosti je dána parametrem *Omezení* na vlastní čísla (Doporučuje se zadat hodnotu maximálně 0.1). Metoda potlačí komponenty vzniklé rozkladem na vlastní čísla a vlastní vektory odpovídající nejmenším vlastním číslům. Výsledkem jsou (vychýlené) odhady parametrů s nižším rozptylem, méně citlivé na špatnou podmíněnost $\mathbf{X}^T\mathbf{X}$.

Kvantilová regrese: Odhadne regresní model odpovídající zadanému kvantilu α v políčku *Kvantil*. Pro tento model Y je pravděpodobnost výskytu hodnoty $\mathbf{x}<Y$ rovna α. Tuto robustní metodu lze použít tam, kde nás nezajímá průběh střední hodnoty, ale průběh „krajní“ hodnoty definované zvoleným kvantilem, například k odhadu minimální pevnosti (α=0.05), maximálního znečištění (α=0.95) a podobně. Používá se iterativní techniky, doba výpočtu je závislá na počtu dat. Počet dat n by měl být větší pro kvantily blízké 1, resp. 0. Platí, že n by mělo být větší než $5/\min(\alpha,1-\alpha)$. Pro α=0.5 se jedná o robustní mediánovou regresi odpovídající Lp-regresi pro p=1, tedy metodě nejmenších absolutních odchylek. Pro velké a malé α je řešení obecně méně přesné a v některých případech může být nejednoznačné. K výpočtu se používá iterativní metoda nejmenších vážených čtverců.

Lp-regrese: Metoda minimalizující součet $\Sigma|e_i|^p$ na rozdíl od metody nejmenších čtverců, která minimalizuje Σe_i^2. Parametr p se zadává v políčku *p (p ≥ 1)*. Pro p=1 se jedná o robustní mediánovou regresi, tedy metodu nejmenších absolutních odchylek, p=2 odpovídá metodě nejmenších čtverců, $p \to \infty$ (v praxi stačí $p \approx 5 - 10$) odpovídá metodě nejmenší maximální chyby (minimax), hodnoty p od 1 do 2 (přesněji: $1 \le p < 2$) vykazují robustnost vůči odlehlým hodnotám. Implicitně je p=1.5. Metoda nejmenších absolutních odchylek při p=1 je vhodná pro data s často se vyskytujícími odlehlými hodnotami na obou stranách nebo s rozdělením podobným Lapaceovu. Metoda minimax je nerobustní

(silně citlivá na vybočující hodnoty) a je vhodná pouze mají-li chyby rovnoměrné rozdělení. Lp-regrese může v některých případech nejednoznačné řešení. K výpočtu se používá iterativní znáhodněné simplexové optimalizační metody.

Nejmenší medián: Moderní robustní regresní metoda minimalizující medián čtverců odchylek. K výpočtu se používá iterativní znáhodněné simplexové optimalizační metody.

IRWLS exp(-e): Robustní regresní metoda ze třídy M-odhadů, při níž se minimalizuje čtverec vážených normovaných reziduí $w(e_{ni})$ s vahami $w(e) = \exp(-e)$. K výpočtu se používá iterativně vážená metoda nejmenších čtverců (angl. *Iteratively Re-Weighted Least Squares*).

M-odhad, Welsch: Robustní regresní metoda ze třídy M-odhadů, při níž se minimalizuje čtverec vážených normovaných reziduí $w(e_{ni})$ s vahami $w(e) = \exp(-e^2)$. K výpočtu se používá iterativně vážená metoda nejmenších čtverců (angl. *Iteratively Re-Weighted Least Squares*).
Ohraničený vliv, BIR: Rezistentní regrese, která je robustní jak k vybočujícím hodnotám závisle proměnné, tak i k silně vlivným datům odlehlých ve smyslu nezávisle proměnné. V této druhé vlastnosti se metoda BIR (angl. *Bounded Influence Regression*) liší od předešlých robustních metod. Její použití může být výhodné například pro polynomické modely k eliminaci určujícího vlivu krajních bodů. K výpočtu se používá iterativně vážená metoda nejmenších čtverců.

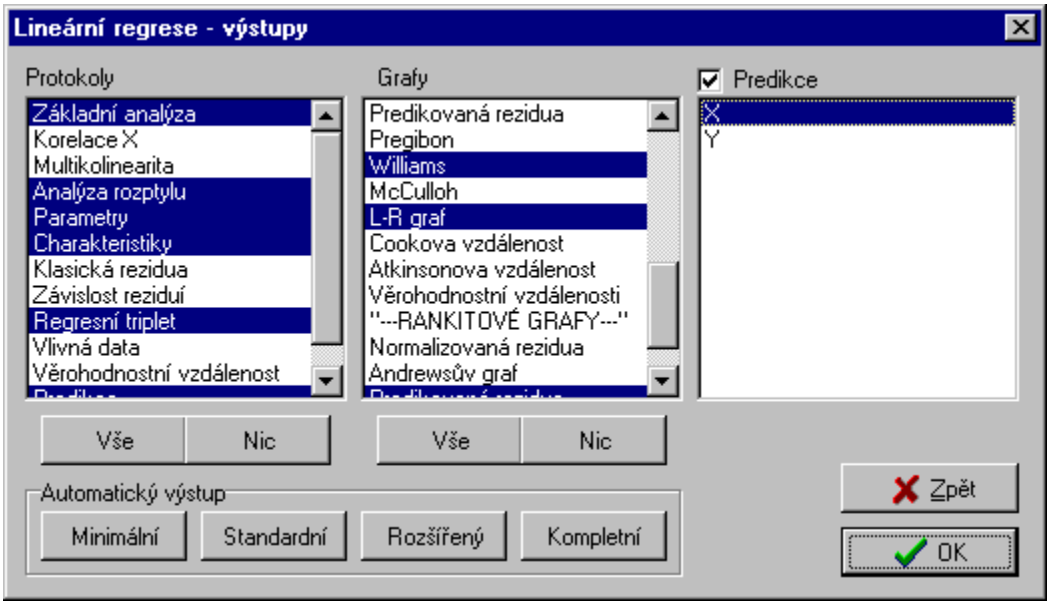
Stepwise, All: Tato metoda, slouží jako pomůcka k sestavení dobrého modelu na základě dat i bez předběžné informace o možných vztazích mezi proměnnými. Vypočítá regrese se všemi možnými kombinacemi vybraných nezávisle proměnných v regresním modelu. Pro každou regresi vypočítá tři kritéria kvality regrese: F-kritérium (FIS), Akaikeho informační kritérium (AIC) a střední kvadratickou chybu predikce (MEP). Na základě nejlepší hodnoty těchto kritérií lze pak vybrat optimální model. Výstup metody *Stepwise All* se ukládá do protokolu a rovněž do zvláštního datového listu s názvem *StepAll*, který se při výpočtu vytvoří. Tento datový list slouží ke snadné identifikaci nejlepších modelů. Pro posouzení kvality jednotlivých modelů lze použít datového listu, nebo tří grafů v grafickém okně. Pozor! Maximální počet proměnných je omezen na 12 s absolutním členem, resp. 13 bez absolutního členu, což platí i pro polynomickou , Taylorovu a obecnou transformaci. Toto omezení plyne z maximálního počtu řádků v datovém listu, kam se ukládají výsledky jednotlivých regresí. Počet regresí pro m nezávisle proměnných (včetně absolutního členu) je $2^m - 1$. K výpočtu jednotlivých regresí je použita vždy klasická metoda nejmenších čtverců. Podrobnější postup je uveden níže, na konci odstavce Protokol a Grafy.

Data: Tato položka určuje, zda se k výpočtu použijí všechny řádky, pouze označené řádky, nebo pouze neoznačené řádky.

Transformace: Definice transformace dat, resp. uživatelského modelu, viz výše.
Výstupy: Po stisknutí tohoto tlačítka se objeví panel se specifikací výstupů, viz další odstavec.
Nápověda: Vyvolá nápovědu.
Zpět: Zrušení akce.
OK: Spuštění regrese.

Výstupy

Tento dialogový panel se vyvolá stiskem tlačítka *Výstupy* v panelu Lineární regrese. Slouží ke specifikaci grafických a textových položek, které požadujeme ve výstupu. V okně lze vybrat ze tří seznamů: *Protokoly* (položky uváděné v protokolu), *Grafy* (grafy nebo jejich skupiny), *Predikce* (výběr nezávisle proměnných pro tabulku predikce na konci protokolu). K rychlému výběru slouží tlačítka *Minimální*, *Standardní*, *Rozšířený*, *Kompletní*, popř. *Vše* a *Nic*. Vybrané proměnné v seznamu *Predikce* se berou v úvahu jen je-li zaškrtnuto políčko *Predikce*.



Obrázek 5 Dialogový panel pro výběr výstupů

Při výběru položek do protokolu je třeba brát v úvahu, že délka některých položek závisí na počtu dat, což může při rozsáhlých datových souborech činit protokol značně nepřehledným. Dále uvádíme obsah jednotlivých položek protokolu a grafů.

Pole ***Protokol***
Základní analýza: Základní charakteristiky proměnných: průměr, směrodatná odchylka, korelace se závisle proměnnou, významnost korelačního koeficientu;
Korelace X: Párové korelační koeficienty mezi nezávisle proměnnými a jejich významnost;
Multikolinearita: Vlastní čísla korelační matice nezávisle proměnných, index podmíněnosti κ, faktor variance inflation, vícenásobné korelační koeficienty;
Analýza rozptylu: aritmetický průměr závisle proměnné, součet čtverců, průměrný čtverec a rozptyl pro jednotlivé složky variability: celková variabilita, variabilita vysvětlená modelem, reziduální variabilita, dále hodnota F-kritéria, kvantil $F(1-\alpha, m-1, n-m)$ a významnost modelu;
Parametry: odhady regresních koeficientů včetně jejich směrodatné odchylky, testu významnosti a intervalu spolehlivosti;
Charakteristiky: Vícenásobný korelační koeficient R, koeficient determinace R^2, predikovaný korelační koeficient R_p, střední kvadratická chyba predikce *MEP*, Akaikeho informační kritérium *AIC*;
Klasická rezidua: Y naměřené, Y vypočítané, směrodatná odchylka Y, reziduum, relativní reziduum, váhy, reziduální součet čtverců, průměr absolutních reziduí, reziduální směrodatná odchylka, reziduální rozptyl, šikmost a špičatost reziduí;
Závislost reziduí: Waldův test autokorelace, Durbin-Watsonův test autokorelace a znaménkový test závislosti reziduí;
Regresní triplet: Fisher-Snedecorův test významnosti modelu, Scottovo kritérium multikolinearity, Cook-Weisbergův test heteroskedasticity, Jarque-Berrův test normality, testy závislosti;
Vlivná data: standardní rezidua, jackknife rezidua, predikovaná rezidua, diagonální prvky projekční matice $\mathbf{H}$ a rozšířené projekční matice $\mathbf{H}^*$, Cookova vzdálenost, Atkinsonova vzdálenost, Andrews-Pregibonona statistika, vliv na predikci, vliv na parametry $LD(b)$, vliv na rozptyl $LD(s)$, celkový vliv $LD(b,s)$;
Věrohodnostní vzdálenost: vliv na parametry $LD(b)$, vliv na rozptyl $LD(s)$, celkový vliv $LD(b,s)$;
Predikce: hodnoty prediktorů, predikovaná hodnota a její interval spolehlivosti.

Pole ***Grafy***
Pole obsahuje následující položky rozdělené do 5 podskupin:
Regresní křivka;

Rezidua: *Y-predikce, Rezidua vs. Predikce, Abs. rezidua, Čtverec reziduí, QQ-graf reziduí, Autokorelace, Heteroskedasticita, Jackknife rezidua, Predikovaná rezidua;*
Parciální grafy: *Parciální regresní grafy Parciální reziduální grafy;*
Vlivná data: *Projekční matice, Predikovaná rezidua, Pregibon, Williams, McCulloh, L-R Graf, Cookova vzdálenost, Atkinsonova vzdálenost;*
Rankitové grafy: *Normalizovaná rezidua, Andrewsův graf, Predikovaná rezidua, Jackknife rezidua.*

Pole ***Predikce***

V tomto poli se vyberou proměnné (prediktory), které se použijí jako nezávisle proměnné pro výpočet predikce. Názvy prediktorů mohou být libovolné, jejich počet musí být shodný s počtem nezávisle proměnných v regresním modelu, pořadí musí být shodné s pořadím nezávisle proměnných v regresním modelu s výjimkou uživatelské transformace. V případě uživatelské transformace se po spuštění výpočtu objeví panel *Asociace proměnných* (Obrázek 6), kde je třeba přiřadit vybrané prediktory (vpravo) původním proměnným (vlevo). Počet řádků prediktoru (tedy bodů, v nichž se počítá predikce) je libovolný. Prediktorem mohou být i tytéž nezávisle proměnné (táž data), které jsou použity v regresním modelu.

Obrázek 6 Asociace proměnných pro predikci

Vše: Označí všechny položky
Nic: Zruší označení všech položek
Minimální, Standardní, Rozšířený, Kompletní: Označí položky protokolu a grafu podle následujících tabulek.

Tabulka **2** Automatický výběr položek protokolu

Položka	Minimální	Standardní	Rozšířený	Kompletní
Základní analýza		o	o	o
Korelace X			o	o
Multikolinearita			o	o
Analýza rozptylu		o	o	o
Parametry	o	o	o	o
Charakteristiky	o	o	o	o
Klasická rezidua			o*	o*
Závislost reziduí				o
Regresní triplet		o	o	o
Vlivná data			o*	o*
Věrohodnostní vzdálenost				o*
Predikce	o**	o**	o**	o**

* Velikost této položky ve výstupní sestavě závisí na počtu dat!
** Podle nastavení položky *Predikce*

Tabulka **3** Automatický výběr položek grafů

Položka	Minimální	Standardní	Rozšířený	Kompletní
Regresní křivka	o	o	o	o
Y-predikce		o	o	o
Rezidua vs. Predikce	o	o	o	o
Abs. rezidua			o	o
Čtverec reziduí				o
QQ-graf reziduí		o	o	o
Autokorelace			o	o
Heteroskedasticita			o	o
Jackknife rezidua				o
Predikovaná rezidua				o
Parciální regresní grafy			o	o
Parciální reziduální grafy				o
Projekční matice	o	o	o	o
Predikovaná rezidua				o
Pregibon				o
Williams		o	o	o
McCulloh				o
L-R Graf		o	o	o
Cookova vzdálenost				o
Atkinsonova vzdálenost				o
Normalizovaná rezidua				o
Andrewsův graf			o	o
Predikovaná rezidua		o	o	o
Jackknife rezidua				o

Protokol

Název úlohy	Název úlohy z dialogového panelu.
Hladina významnosti	Hodnota a zadaná v dialogovém panelu, která se používá pro výpočet intervalů spolehlivosti a všechny testy.
Kvantil t(1-alfa/2,n-m)	Kvantil t-rozdělení.
Kvantil F(1-alfa,m,n-m)	Kvantil F-rozdělení.
Absolutní člen	Obsahuje model absolutní člen?
Počet platných řádků	Počet řádků s platnými hodnotami všech proměnných.
Počet parametrů	Počet nezávisle proměnných v modelu včetně absolutního členu a transformovaných proměnných, např. počet parametrů polynomu 3. stupně je 4.
Metoda	Zvolená metoda výpočtu.
Sloupce pro výpočet	Seznam proměnných použitých v regresi.
Transformace	Zvolený typ transformace.
Základní analýza	
Charakteristiky proměnných	
Proměnná	Název proměnné.
Průměr	Aritmetický průměr proměnné.
Směr.Odch.	Směrodatná odchylka proměnné.
Kor.vs.Y	Párový korelační koeficient mezi nezávisle a závisle proměnnou.
Významnost	p-hodnota testu významnosti korelačního koeficientu.

Párové korelace (Xi, Xj)	Párové korelační koeficienty mezi všemi dvojicemi nezávisle proměnných.
Indikace multikolinearity	
Proměnná	Název proměnné, zde má význam pouze pro poslední sloupec (Vícenás. kor.), vlastní čísla nelze jednoznačně přiřadit k jednotlivým proměnným.
Vlas. čísla kor. m.	Vlastní čísla korelační matice nezávisle proměnné.
Podmíněnost kappa	Index (číslo) podmíněnosti κ je poměr největšího a nejmenšího vlastního čísla. Maximální hodnota $\kappa_{max} > 1000$ se považuje za indikaci silné multikolinearity.
VI faktor	Faktor vzrůstu rozptylu v důsledku multikolinearity, hodnoty VIF > 10 se považují za indikaci silné multikolinearity.
Vícenás. kor.	Vícenásobný korelační koeficient mezi danou proměnnou a všemi ostatními nezávisle proměnnými.
Analýza rozptylu	
Průměr Y	Aritmetický průměr nezávisle proměnné.
Zdroj	Zdroj variability, která je vyjádřena jako součet čtverců, průměrný čtverec a rozptyl.
Celková variabilita	Variabilita závisle proměnné pro model Y = průměr(Y).
Variabilita vysvětlená modelem	[Celková variabilita] – [reziduální variabilita].
Reziduální variabilita	Variabilita reziduí, která není vysvětlená modelem.
Hodnota kritéria F	Vypočítaná testační statistika pro daný model. Je-li větší než kvantil F, lze model považovat za statisticky významný, tedy lepší než model $Y = y$.
Kvantil F (1-alfa, m-1, n-m)	Kvantil F-rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model považován za významný.
Závěr	Verbálně vyjádřená významnost modelu.
Odhady parametrů	
Proměnná	Název proměnné.
Odhad	Odhad regresního koeficientu příslušejícího dané proměnné.
Směr.Odch.	Směrodatná odchylka regresního koeficientu.
Závěr	Verbální závěr testu statistické významnosti regresního koeficientu.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je koeficient považován za významný.
Spodní mez	Spodní mez intervalu spolehlivosti regresního koeficientu na dané hladině významnosti.
Horní mez	Horní mez intervalu spolehlivosti regresního koeficientu na dané hladině významnosti. Obsahuje-li interval spolehlivosti nulu, je koeficient statisticky nevýznamný.
Statistické charakteristiky regrese	
Vícenásobný korelační koeficient R	Vícenásobný korelační koeficient vyjadřuje relativní těsnost proložení (nikoli kvalitu modelu). Korelační koeficient vždy roste (resp. neklesá) s počtem proměnných!

Koeficient determinace R^2	Čtverec vícenásobného korelačního koeficientu.
Predikovaný korelační koeficient Rp	Predikovaný korelační koeficient je citlivější na vybočující hodnoty než klasický koeficient.
Střední kvadratická chyba predikce MEP	Chyba predikce *i*-té hodnoty závisle proměnné spočítaná regresí s vyloučením *i*-tého bodu. Citlivá na vybočující hodnoty a multikolinearitu, důležitá míra kvality regrese.
Akaikeho informační kritérium	AIC, kritérium kvality regrese vycházející z reziduálního součtu čtverců penalizovaného počtem proměnných.
Analýza klasických reziduí	
Index	
Y naměřené	Zadaná hodnota závisle proměnné.
Y vypočítané	Predikovaná hodnota závisle proměnné.
Směr. odch. Y	Odhad směrodatné odchylky predikce v *i*-tém bodě.
Reziduum	Rozdíl zadané a predikované hodnoty závisle proměnné v *i*-tém bodě.
Reziduum [%Y]	Relativní reziduum, reziduum dělené hodnotou závisle proměnné.
Váhy	Váha i-tého měření zadaná uživatelem.
Reziduální součet čtverců	Součet čtverců reziduí. S rostoucím počtem proměnných vždy klesá (resp. neroste).
Průměr absolutních reziduí	Průměr absolutních hodnot reziduí
Reziduální směr. odchylka	Směrodatná odchylka reziduí.
Reziduální rozptyl	Rozptyl reziduí
Šikmost reziduí	Šikmost reziduí
Špičatost reziduí	Špičatost reziduí
Testování regresního tripletu	
Fisher-Snedecorův test významnosti modelu	Testuje, zda použitý model je lepší, než prostý průměr závisle proměnné (tedy "žádný model").
Hodnota kritéria F	Vypočítaná testační statistika.
Kvantil F (1-alfa, m-1, n-m)	Příslušný kvantil F-rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model statistiky významný.
Závěr	Verbální závěr testu.
Scottovo kritérium multikolinearity	Testuje, zda mezi nezávisle proměnnými není příliš veliká kolinearita ("závislost"), která může velmi výrazně zvýšit rozptyly parametrů.
Hodnota kritéria SC	Vypočítaná testační statistika.
Závěr	Verbální závěr testu.
Cook-Weisbergův test heteroskedasticity	Testuje konstantnost rozptylu chyb. Je-li přítomna heteroskedasticita, je nutno uvažovat o použití vhodných vah.
Hodnota kritéria CW	Vypočítaná testační statistika.
Kvantil Chi^2(1-alfa,1)	Příslušný kvantil χ^2-rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model statistiky významný.
Závěr	Verbální závěr testu.

Jarque-Berrův test normality	Testuje normalitu rozdělení chyb pomocí rozdělení reziduí.
Hodnota kritéria JB	Vypočítaná testační statistika.
Kvantil Chi^2(1-alfa,2)	Příslušný kvantil χ^2-rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model statistiky významný.
Závěr	Verbální závěr testu.
Waldův test autokorelace	Testuje přítomnost autokorelace chyb na základě vypočítaných reziduí.
Hodnota kritéria WA	Vypočítaná testační statistika.
Kvantil Chi^2(1-alfa,1)	Příslušný kvantil χ^2-rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model statistiky významný.
Závěr	Verbální závěr testu.
Durbin-Watsonův test autokorelace	Testuje přítomnost autokorelace chyb na základě vypočítaných reziduí.
Hodnota kritéria DW	Vypočítaná testační statistika.
Závěr	Verbální závěr testu.
Znaménkový test reziduí	Neparametricky ověřuje přítomnost závislostí, které nejsou postihnuty modelem.
Hodnota kritéria Sg	Vypočítaná testační statistika.
Kvantil N(1-alfa/2)	Příslušný kvantil normálního rozdělení.
Pravděpodobnost	p-hodnota testu, je-li menší než zadaná hladina významnosti, je model statistiky významný.
Závěr	Verbální závěr testu.
Indikace vlivných dat	
A. Analýza reziduí	
Index	
Standardní	Klasické reziduum dělené svojí směrodatnou odchylkou $1/s_r$.sqrt(1-H_{ii}), někdy nazýváno studentizované, s_r je reziduální směrodatná odchylka.
Jackknife	Jackknife reziduum, jako Standardní, místo s_r je pro i-tý bod použita směrodatná odchylka získaná vynecháním i-tého bodu. Toto reziduum citlivěji indikuje vybočující body.
Predikované	Predikované reziduum, rozdíl i-té hodnoty nezávisle proměnné od modelu získaného po vynechání i-tého bodu. Toto reziduum citlivěji indikuje vybočující body.
Diag(Hii)	Diagonální prvky projekční matice, velké hodnoty naznačují velký vliv daného bodu na regresi. Součet H_{ii} je roven počtu parametrů. Příliš vlivné body jsou zvýrazněny červeně.
Diag(H*ii)	Diagonální prvky projekční matice rozšířené o závisle proměnnou, velké hodnoty naznačují velký vliv daného bodu na regresi. Součet H^*_{ii} je roven počtu parametrů + 1. Příliš vlivné body jsou zvýrazněny červeně.
Cookova vzdál.	Cookova vzdálenost je mírou vlivu i-tého bodu na hodnoty regresních koeficientů. Příliš vlivné body jsou zvýrazněny červeně.
B. Analýza vlivu	
Index	

Atkinsonova vzdál.	Atkinsonova modifikace Cookovy vzdálenosti (1985), ve většině případů poskytuje podobné výsledky. Příliš vlivné body jsou zvýrazněny červeně.
Andrews-Pregibon st.	Andrews-Pregibon statistika je mírou vlivu jednotlivých bodů na rozptyl parametrů (objem konfidenčního elipsoidu parametrů). Příliš vlivné body jsou zvýrazněny červeně.
Vliv na Y^	Relativní vliv jednotlivých bodů na predikci. Příliš vlivné body jsou zvýrazněny červeně.
Vliv na parametry LD(b)	Relativní vliv na hodnoty parametrů. Příliš vlivné body jsou zvýrazněny červeně.
Vliv na rozptyl LD(s)	Relativní vliv na rozptyl reziduí. Příliš vlivné body jsou zvýrazněny červeně.
Celkový vliv LD(b,s)	Souhrnný vliv na parametry a rozptyl. Příliš vlivné body jsou zvýrazněny červeně.

Predikce	
Hodnota prediktoru	Hodnoty všech nezávisle proměnných, při polynomické, Taylorově, nebo uživatelské transformaci hodnoty všech transformovaných proměnných. Absolutní člen je reprezentován sloupcem jedniček.
Predikce	Hodnota závisle proměnné vypočítaná z modelu.
Spodní mez	Spodní mez intervalu spolehlivosti predikce pro zadanou hladinu významnosti α.
Horní mez	Horní mez intervalu spolehlivosti predikce pro zadanou hladinu významnosti α.

Protokol pro Stepwise regresi

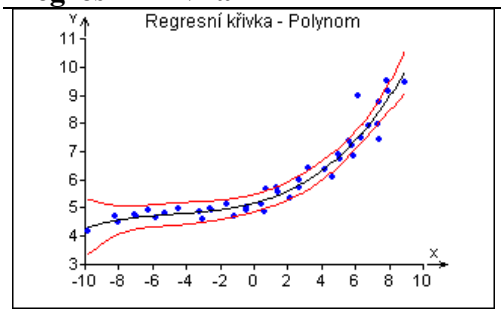
Metoda *Stepwise-All* (angl. *all possible subsets regression*) je určena k nalezení nejvhodnějšího modelu na základě dat pomocí tří kritérií: F-statistiky, Akaikeho kritéria, nebo *MEP* (střední kvadratická chyba predikce). Tato metoda počítá regrese pro všechny možné kombinace daných nezávisle proměnných a výsledky shrnuje do protokolu a do zvláštního datového listu označeného *StepAll*, který se automaticky přidá do datového okna a slouží k nalezení optimálního modelu. Protokol obsahuje odstavec s použitými proměnnými a odstavec s hodnotami kritérií pro každý model. V tabulce použitých proměnných je každé proměnné přiřazeno jedno písmeno abecedy a číslo, které se pak použijí při identifikaci proměnných v tabulce hodnocení modelu. Nejlepší modely lze pak nalézt buď setříděním řádků podle zvoleného kritéria (před tříděním je nutné označit všechny sloupce v datovém listu *StepAll*), nebo označením bodů s nejlepší hodnotou kritéria ve zvoleném grafu, viz konec následujícího odstavce Grafy. Nejlepším modelům odpovídají *nejvyšší* hodnoty *fis*, *nejnižší* hodnoty *AIC*, *nejnižší* hodnoty *MEP*. Doporučuje se vždy posuzovat několik nejlepších modelů, nikoli pouze jeden s absolutně nejlepší hodnotou kritéria. Každé ze tří kritérií dává obvykle mírně odlišné výsledky, je třeba uvážit povahu kritéria: *fis* je klasické F-kritérium, které porovnává statistickou významnost modelů, Akaikeho kritérium $AIC = n.\ln(RSČ/n) + 2.m$ porovnává reziduální součet čtverců penalizovaný počtem parametrů a lze jej chápat jako míru informačního zisku, střední predikovaná chyba *MEP* porovnává predikční schopnost modelu. Univerzální pojem „nejlepšího“ modelu neexistuje. Při volbě modelu je vždy výhodné využít dalších znalostí a vědomostí o povaze dat a vztahů mezi nimi.

Vybrané sloupce	Proměnné uvažované jako kandidáty pro regresní model. Každé proměnné je přiřazeno písmeno a číslo pro snadnou orientaci ve druhé části tabulky.
Hodnocení modelů	Kopie této tabulky se ukládá do přidaného datového listu, kde lze provádět třídění a označování řádků ve spojení s grafy. V okně

	Protokol tyto operace provádět nelze. Ve sloupcích jsou uvedeny hodnoty jednotlivých kritérií *fis*, *AIC* a *MEP* a pro orientaci rovněž reziduální součet čtverců *RSC*. K nalezení nejlepších hodnot zvoleného kritéria lze využít funkce třídění (*Menu – QCExpert – Třídění* nebo označení bodů s pomocí grafů. Pozor: *RSC* **není** kritérium kvality modelu! Je vždy nejmenší pro maximální počet proměnných.

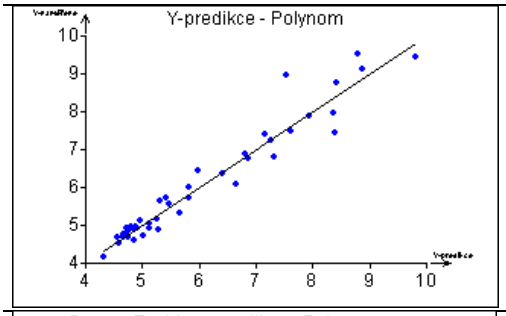
Grafy

Regresní křivka

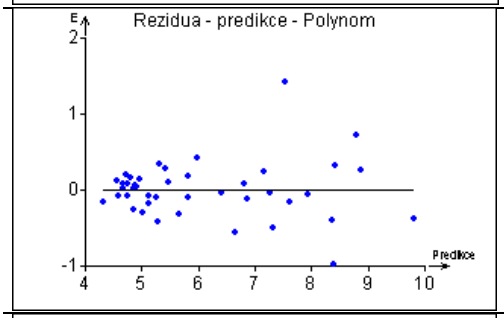


Pokud je vybráno více nezávisle proměnných, tento graf se nekreslí. Je-li v datech pouze jedna nezávisle proměnná, představuje graf průběh regresního modelu. Červeně je vyznačen pás spolehlivosti modelu na zadané hladině významnosti. Je nutné mít na paměti, že pás spolehlivosti predikce, zvláště mimo interval dat, je reálný pouze pokud zvolený model odpovídá skutečnosti. Vhodným zmenšením měřítka (zoom) lze získat detail, nebo naopak průběh i mimo interval měřených dat.

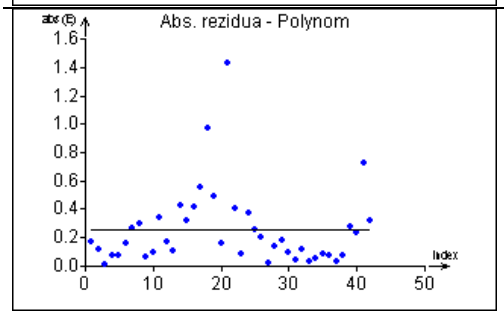
Rezidua



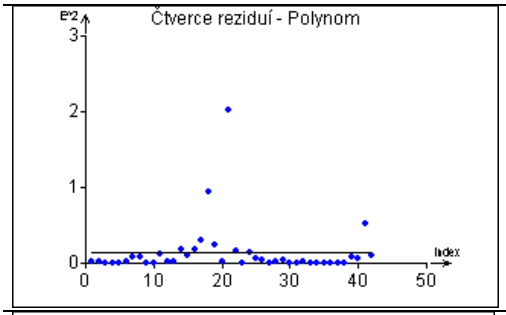
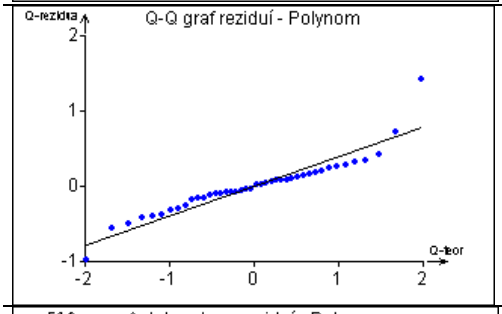
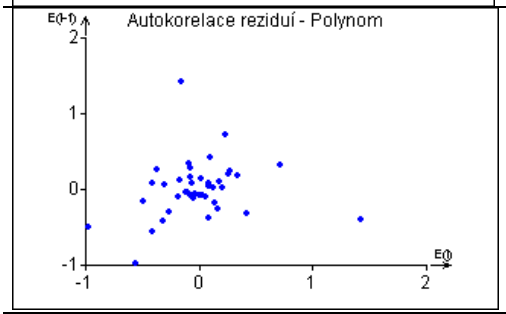
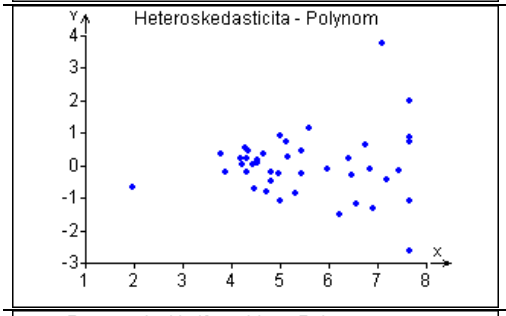
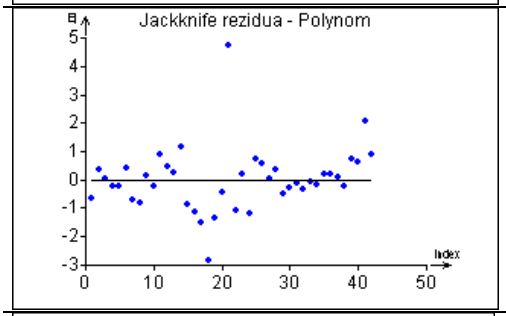
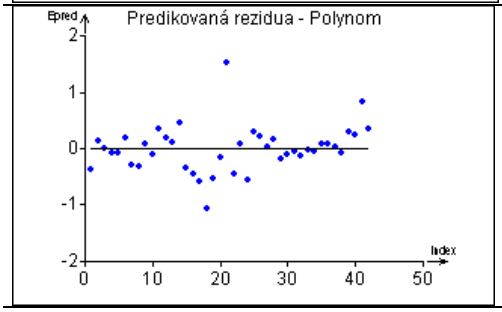
Graf vyjadřující těsnost proložení. Na ose *X* jsou vypočítané hodnoty závisle proměnné, na ose *Y* jsou naměřené hodnoty. Svislé vzdálenost bodu od přímky odpovídá reziduu.



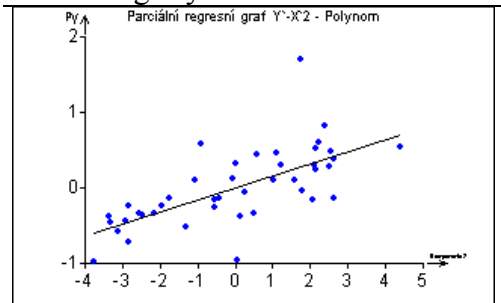
Graf normovaných reziduí, na ose *X* je hodnota závisle proměnné. vodorovná přímka odpovídá průměru reziduí. V případě nevážené metody nejmenších čtverců je průměr reziduí roven nule.



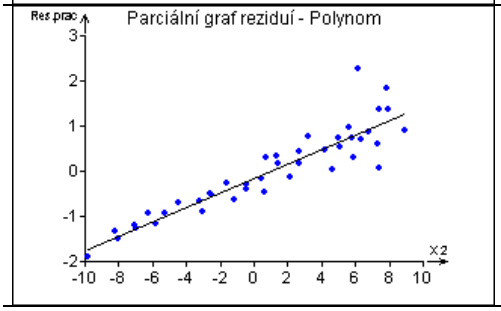
Absolutní hodnoty reziduí, na ose *X* je pořadí bodu. Vodorovná přímka odpovídá průměrné absolutní chybě.

	Druhá mocnina hodnoty reziduí, na ose *X* je pořadí bodu. Vodorovná přímka odpovídá průměrné (střední) kvadratické chybě.
	Q-Q graf pro posouzení normality reziduí. Přímka odpovídá normálnímu (Gaussovu) rozdělení reziduí. Je nutno brát v úvahu, že metoda nejmenších čtverců uměle zvyšuje normalitu (tzv. supernormalita). V případě pochybností se doporučuje vyhodnotit tento graf i pro některou robustní metodu.
	Grafické posouzení autokorelace reziduí prvního řádu, na ose *X* je *i*-té reziduum, na ose *Y* je (*i*–1) reziduum. "Mrak" bodů s kladnou směrnicí, naznačuje pozitivní autokorelaci, klesající trend negativní autokorelaci. Autokorelace reziduí nemusí nutně dokazovat autokorelaci chyb, neboť autokorelace vypočítaných reziduí je vždy nenulová.
	Grafické posouzení heteroskedasticity (nekonstantnosti rozptylu). Tvar výseče, resp. trojúhelníku naznačuje přítomnost heteroskedasticity.
	Jackknife rezidua (viz Protokol) daleko citlivěji indikují vybočující měření, než klasická rezidua. V přítomnosti většího počtu blízkých vybočujících hodnot však mohou selhat.
	Predikovaná rezidua jsou rovněž velmi citlivým indikátorem vybočujících bodů. V přítomnosti většího počtu blízkých vybočujících hodnot však mohou selhat.

Parciální grafy

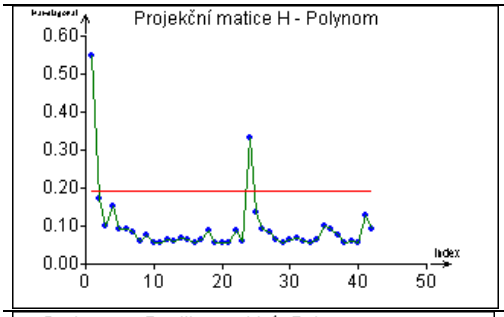


Parciální regresní graf vyjadřuje závislost závisle proměnné na zvolené jediné nezávisle proměnné s eliminací vlivu ostatních nezávisle proměnných. Směrnice přímky odpovídá příslušnému regresnímu koeficientu. Těsnost proložení souvisí s významností daného koeficientu.

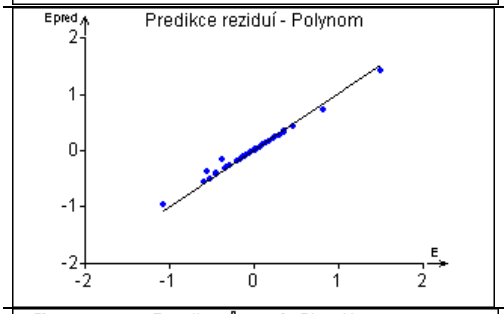


Parciální reziduální graf, modifikace parciálního regresního grafu, nelineární tvar bodů indikuje přítomnost nelineární závislosti, kterou je možné popsat například přidáním vyšší mocniny příslušné proměnné.

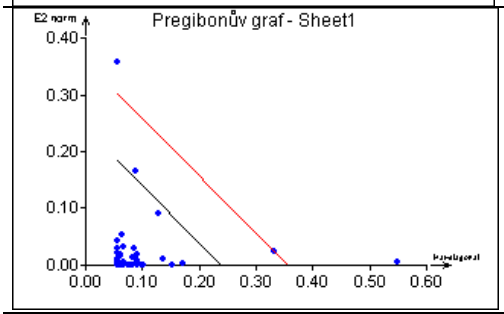
Vlivná data



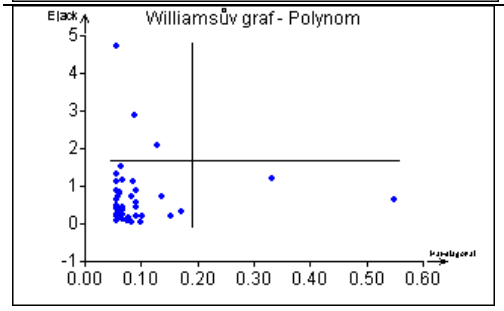
Diagonální prvky projekční matice $\mathbf{H}=\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, které vyjadřují míru vlivu jednotlivých dat na regresi ($\mathbf{X}$ je matice nezávisle proměnných). Body nad vodorovnou přímkou se považují za silně vlivné.



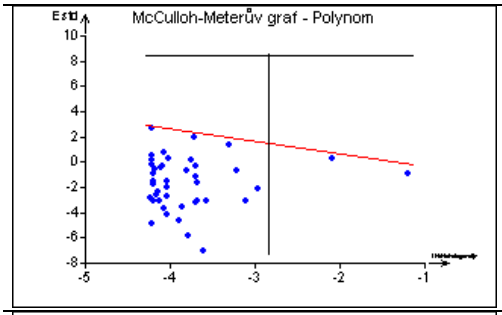
Graf predikce reziduí Grafické srovnání skutečných a predikovaných reziduí. Výraznější odchylka od přímky indikuje vybočující hodnotu. Tento graf je velmi citlivý na jednotlivé vybočující hodnoty, špatně indikuje skupiny vybočujících hodnot.



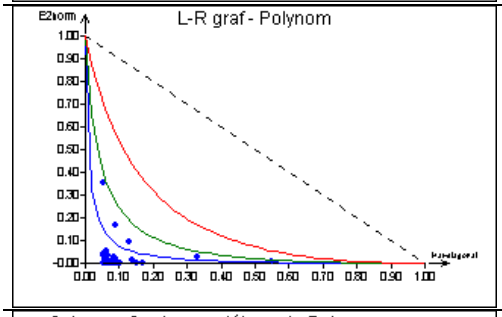
Graf pro společné posouzení vybočujících bodů a vlivných bodů. Body nad nižší (černou) přímkou se považují za vlivné, nad vyšší (červenou) přímkou za silně vlivné nebo vybočující a je třeba jim věnovat pozornost.



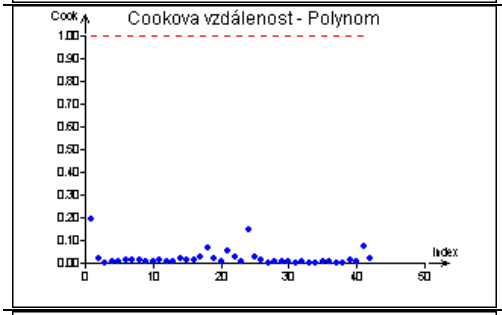
Williamsův graf slouží k indikaci vlivných i vybočujících bodů. Body vpravo od svislé přímky jsou silně vlivné, body nad vodorovnou přímkou jsou silně vybočující.



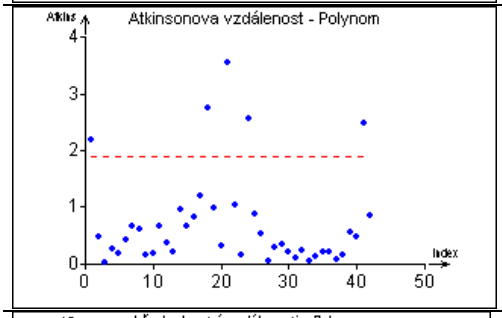
McCulloh-Meterův graf je další alternativou k indikaci vlivných a vybočujících bodů. Body vpravo od svislé přímky jsou silně vlivné, body nad vodorovnou přímkou jsou silně vybočující. Body nad šikmou (červenou) přímkou jsou podezřelé vybočující nebo vlivné.



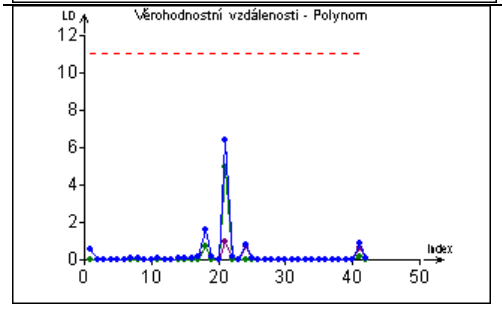
L-R graf je další alternativou k indikaci vlivných bodů. Hyperbolické křivky jsou linie stejného vlivu. Podle polohy bodů vůči třem křivkám lze data rozdělit na slabě vlivná, vlivná a silně vlivná. Tento graf je vhodný pro menší rozsahy dat.



Cookova vzdálenost vyjadřuje vliv dat na velikost (nikoli rozptyl) odhadovaných parametrů.

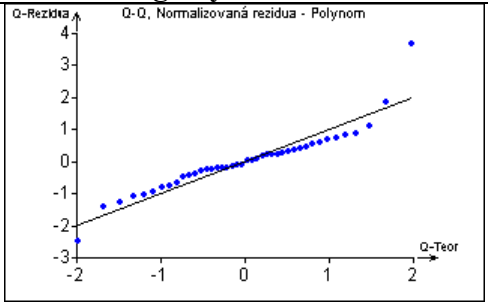


Atkinsonova vzdálenost je další diagnostikou k posouzení vlivu jednotlivých dat. Je modifikací Cookovy vzdálenosti, ve většině případů poskytuje podobné výsledky. Data nad vodorovnou přímkou se považují za silně vlivná.

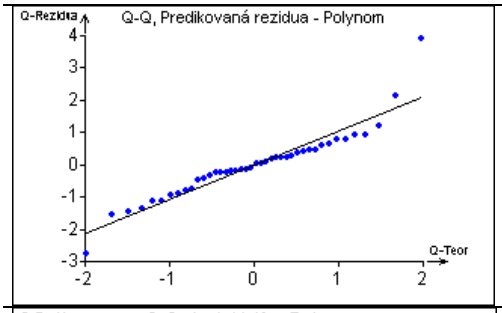
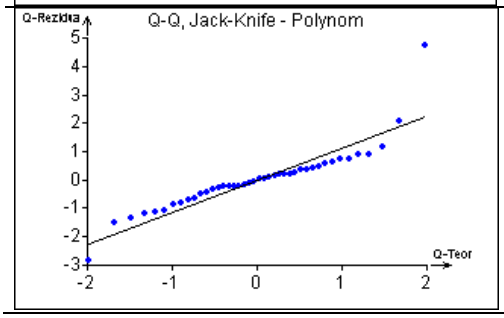


Graf věrohodnostní vzdálenosti vyjadřuje vliv jednotlivých dat na parametry (fialová), predikci (zelená) a parametry i predikci (modrá).

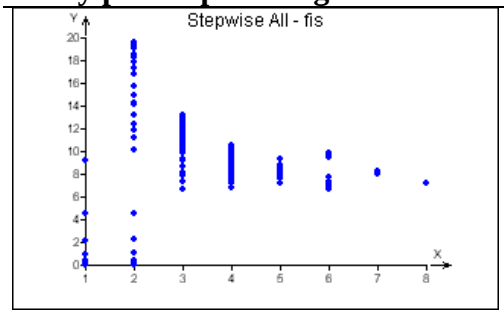
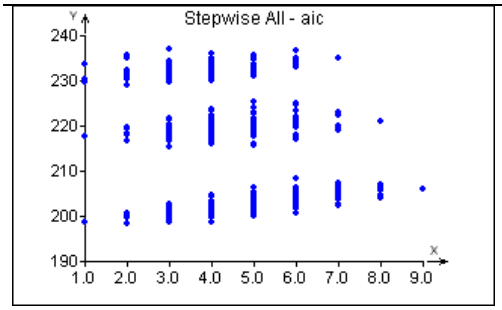
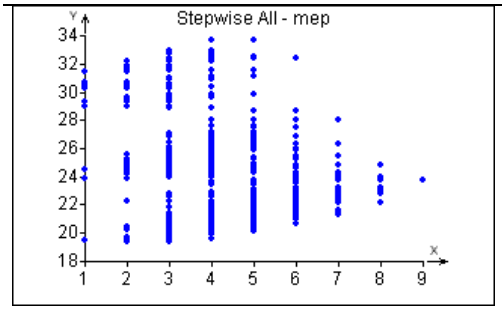
Rankitové grafy



Q-Q graf normovaných reziduí pro posouzení normality reziduí, přímka odpovídá normálnímu rozdělení reziduí.

	Q-Q graf predikovaných reziduí pro posouzení normality, přímka odpovídá normálnímu rozdělení predikovaných reziduí.
	Q-Q graf jackknife reziduí pro posouzení normality, přímka odpovídá normálnímu rozdělení jackknife reziduí.

Grafy pro Stepwise regresi:

	Tento graf je generován v metodě Stepwise a slouží k identifikaci nejlepších modelů pro daná data. Na ose *X* je počet proměnných v modelu, na ose *Y* je hodnota F-kritéria významnosti modelu. Nejlepší modely z hlediska F-kritéria mají nejvyšší hodnoty F. K identifikaci se použije možnosti interaktivního označení nejvýše ležících bodů a nalezení odpovídajících modelů v datovém listu *Stepwise-All*. Doporučuje se označit více bodů a pak zvolit modely na základě dalších znalostí.
	Tento graf je generován v metodě Stepwise a slouží k identifikaci nejlepších modelů pro daná data. Na ose *X* je počet proměnných v modelu, na ose *Y* je hodnota Akaikeho kritéria *AIC*. Nejlepší modely z hlediska *AIC* mají nejnižší hodnoty. K identifikaci se použije možnosti interaktivního označení nejníže ležících bodů a nalezení odpovídajících modelů v datovém listu *Stepwise-All*. Doporučuje se označit více bodů a pak zvolit modely na základě dalších znalostí. Přítomnost oddělených pásů je způsobena přítomností proměnné, nebo kombinace proměnných s velkou významností.
	Tento graf je generován v metodě Stepwise a slouží k identifikaci nejlepších modelů pro daná data. Na ose *X* je počet proměnných v modelu, na ose *Y* je hodnota střední kvadratické chyby predikce (*MEP*). Nejlepší modely z hlediska *MEP* mají nejnižší hodnoty. K identifikaci se použije možnosti interaktivního označení nejníže ležících bodů a nalezení odpovídajících modelů v datovém listu *Stepwise-All*. Doporučuje se označit více bodů a pak zvolit modely na základě dalších znalostí.