UNIVERZITA PARDUBICE

FAKULTA EKONOMICKO-SPRÁVNÍ

KVALITA REGRESNÍHO MODELU

Radek Fajfr

Bakalářská práce

2010

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: **2009/2010**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Radek FAJFR, DiS.**
Studijní program:	**B6209 Systémové inženýrství a informatika**
Studijní obor:	**Regionální a informační management**
Název tématu:	**Kvalita regresního modelu**
Zadávající katedra:	**Ústav matematiky**

Zásady pro vypracování:

Úvod
1. Problematika regresní analýzy
2. Metody posuzování kvality zvoleného regresního modelu
3. Ukázka volby nejvhodnějšího regresního modelu pro konkrétní data z ekonomické praxe
Závěr
Literatura

Rozsah grafických prací:	–
Rozsah pracovní zprávy:	**cca 30 stran**
Forma zpracování bakalářské práce:	**tištěná/elektronická**

Seznam odborné literatury:

ANDĚL, J. Základy matematické statistiky. Praha: Matfyzpress, 2005. 360 s. ISBN 80-86732-40-1.
KUBANOVÁ, J. Statistické metody pro ekonomickou a technickou praxi. Bratislava: Statis, 2004. 256 s. ISBN 80-85659-37-9.
HINDLS, R.; HRONOVÁ, S.; SEGER, J. Statistika pro ekonomy. Praha: Professional Publishing, 2003. 418 s. ISBN 80-86419-52-5.

Vedoucí bakalářské práce:	**Mgr. Pavla Jindrová** Ústav matematiky

Datum zadání bakalářské práce:	**15. června 2009**
Termín odevzdání bakalářské práce:	**30. dubna 2010**

doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.

doc. RNDr. Bohdan Linda, CSc.
vedoucí ústavu

V Pardubicích dne 14. července 2009

Prohlášení

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury. Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 21.4.2010

Fajfr Radek

ANOTACE

Bakalářská práce je vypracována na téma Kvalita regresního modelu. Cílem je vymezení základních pojmů regresní analýzy a využití těchto znalostí při praktickém řešení příkladů.

V praktické části jsou příklady řešené metodami regresní analýzy a následuje vytvoření regresních funkcí. Dále následuje výběr optimálního regresního modelu spolu se zdůvodněním tohoto výběru.

KLÍČOVÁ SLOVA

Regresní analýza, Náhodná veličina, Metoda nejmenších čtverců, Index determinace, Index korelace, Rezidua, Významnost F.

TITLE

Quality of Regression Model

ANOTATION

This bachelor work is created for the topic Quality of Regression Model. The aim of this work is to explain of basic expressions related to regression analysis and application of this knowledge in practicular solutions.

In practical part of the bachelor work there are examples solved by methods of regression analysis and then are regression functions created. The following topic is the choice of the optimal regression model together with the explanation of this choice.

KEYWORDS

Regression analysis, Radom variable, Method of the least squares, Index determination, Index corelation, Residua, Meaning F.

Poděkování

Zde bych rád poděkoval Mgr. Haně Boháčové a Mgr. Pavle Jindrové za ochotu a odborné vedení, cenné rady a náměty během tvorby práce. Dále bych rád poděkoval celé své rodině, zejména otci Fantišku Fajfrovi a bratrovi Martinovi Fajfrovi za projevenou podporu.

OBSAH:

Úvod

Úvodem lze říci, že regresní analýza si klade za cíl vniknout do podstaty sledovaného jevu a procesů určité oblasti. Tím se snaží přiblížit příčinným souvislostem. Příčinnou souvislostí je situace, kdy existence určitého jevu souvisí s existencí jiného jevu. Důvodem, proč jsem si vybral toto téma je skutečnost, že s regresí se setkáváme v mnoha situacích reálného světa právě v podobě vztahu příčiny a následku. Například ekonomická krize vyvolává nárůst nezaměstnanosti, což je typická ukázka vztahu příčiny a následku. Dalším důvodem, proč jsem si dané téma zvolil byl fakt, že tato matematická metoda je velmi důležitá nejen pro statistiku jako vědní obor, ale i pro jiná odvětví a tato univerzálnost mě rovněž zaujala.

Jedním z hlavních cílů této práce je výběr optimálního regresního modelu pro vybraná data. Dříve než však můžeme vytvořit, analyzovat jednotlivé regresní modely a nakonec vybrat optimální regresní model, je potřeba seznámit se s teorií tohoto problému. Proto práce zahrnuje oddělenou teoretickou část, kde jsou vysvětleny jednotlivé pojmy regresní analýzy i hlavní matematické vztahy s tímto problémem související. Všechny popsané vztahy a definice jsou zpracované pomocí relevantní matematické literatury uvedené v použitých zdrojích. Teoretická část obsahuje a vysvětluje základní pojmy lineární regrese (např. volná, pevná závislost) a také popisuje základní regresní modely, které jsou používány při řešení příkladů. Dále teoretická část mimo jiné popisuje používanou metodu nejmenších čtverců a definuje základní parametry (např. index determinace), podle kterých se vybírá optimální regresní model v řešených příkladech.

V další části práce se pak v souladu se zadáním zaměřuji na samotné řešení vybraných příkladů. Zpracoval jsem dva odlišné příklady, ale průběh analýzy je v obou případech stejný. Nejprve je pomocí nástroje MS Excel vždy vytvořen regresní model. Dále se zhodnotí proložení regrese daty jednak vizuálně z vytvořeného grafu, ale hlavně podle ukazatelů kvality regresního modelu. V některých případech je provedena předběžná predikce porovnáním vstupních závislých proměnných y a očekávaných Y vytvořených analýzou. Na závěr každého z příkladů je vytvořena hodnotící tabulka, která obsahuje všechny vytvořené modely s jejich parametry. Vzájemným porovnáním parametrů se vybere nejoptimálnější model.

1 Pojmy regresní analýzy

1.1 Pevná a volná závislost

Je vhodné rozlišit tzv. **pevné** a **volné** závislosti. Závislostí pevnou je případ, kdy výskytu jednoho jevu nutně odpovídá výskyt druhého jevu. Pro pevnou závislost je charakteristické to, že se opakuje ve všech jednotlivých případech, v nichž je pozorována. To znamená, že ji charakterizuje i jediné pozorování [1]. Známým příkladem, který mohu uvést je proslulé Descartovo "Myslím, tedy jsem". Jde o vztah, který se projeví s jistotou, tedy pravděpodobností rovné jedné. Průběh závislosti lze přesně charakterizovat matematickou funkcí. O závislosti volné hovoříme v těch případech, kdy výskyt jednoho jevu ovlivňuje výskyt druhého jevu. Podmínkou je nastoupení prvního jevu. Je to tedy závislost, při níž jeden jev podmiňuje jev jiný jen s určitou pravděpodobností a v různé intenzitě. Jsou-li tedy jevy spojeny volně, neříká nám jediné pozorování o jejich závislosti vůbec nic. Takto pozorovaná závislost může být nahodilá [1].

Při rozšíření těchto úvah na statistické znaky, které bude vhodnější nazývat proměnnými, znamená pevná závislost vztah, kdy hodnotě jedné proměnné odpovídá jedna a jen jedna hodnota jiných proměnných a podobně i naopak. Jedná se o pevnou závislost, protože je v podstatě kauzální (když nastane jedna událost, tak s pravděpodobností 1 nastane další určitá událost). Volnou závislostí pak dle Hindlse rozumíme vztah, kdy hodnotám (např. jedné proměnné) odpovídají sice různé hodnoty jiné proměnné, ale kdy lze hovořit o jakési "obecné tendenci", která se projevuje při změnách hodnot těchto proměnných. Z toho tedy plyne, že hodnoty se mění plynule s určitou pravděpodobností v závislosti na proměnných. Označme jeden statistický znak (jednu proměnnou) jako x, druhý statistický znak jako y. Pak můžeme při volné závislosti mezi x a y očekávat změny hodnot y při změnách hodnot x a naopak tendenci změn hodnot x při změnách hodnot y [5].

1.2 Funkční závislost

Pokud hodnota jedné proměnné závisí na hodnotách druhé proměnné, pak říkáme, že je tato závislost určena funkčním vztahem $\boldsymbol{y = f(x)}$. Pokud tedy známe konkrétní hodnoty x, pak dokážeme přesně určit, jaké hodnoty nabude proměnná y. Kdekoliv je takový pevný, čili funkční stav mezi kvantitativními statistickými znaky, říkáme, že závislost je *úplná* [2]. V praktických úlohách není situace zdaleka tak jednoznačná. Na sledovanou veličinu nepůsobí obvykle pouze jedna náhodná veličina *X*, ale většinou je jich více. V takovémto případě není mezi veličinami *X* a *Y* funkční závislost, ale přesto se jedná o veličiny závislé. Nemluvíme pak o závislosti funkční, ale stochastické.

1.3 Statistická závislost (stochastická)

Definice [7]**:** Nechť *X*, *Y* jsou dvě náhodné veličiny. Jestliže změna hodnoty jedné náhodné veličiny vyvolá změnu rozdělení pravděpodobnosti druhé náhodné veličiny, říkáme, že náhodné veličiny *X*, *Y* jsou **stochasticky závislé**.

Proměnnou *y* považujeme za stochasticky závislou na proměnné *x* tehdy, jestliže se při změnách proměnné *x*, mění podmíněná rozdělení četností proměnné *y*[5].

Znaky stochastické závislosti [7]:

a) Změny závislé proměnné jsou vysvětleny pouze některými (ne všemi) činiteli těchto změn,

b) bere se v úvahu působení náhodných vlivů,

c) připouštíme možnost chyb.

Příkladem může být počet členů domácnosti a výdaje domácnosti na nákup potravin. Je možné tvrdit, že určitému počtu členů domácnosti odpovídá určité rozdělení výdajů na potraviny. Výdaje na potraviny jsou ovlivněny i různými nekontrolovatelnými vlivy, které nazýváme náhodnými vlivy (oslava jubilea, návštěva, nemoc, dieta..)

Závislostí statistickou nazýváme volnou závislostí (definováno v podkapitole 1.1), která se týká kvantitativních znaků. Se závislostmi pevnými se většinou setkáváme v teoretické oblasti. Takovým způsobem byl zformulován Newtonův gravitační zákon nebo Ohmův fyzikální zákon a z oblasti ekonomické je možné sem zařadit různé teoretické zákony. Jedná se například o závislosti množství peněz v ekonomice na úrocích. V reálných situacích se setkáváme většinou pouze s volnými závislostmi. Za obecnými tendencemi, projevujícími se v souboru statistických údajů se mohou ukrývat hlubší zákonitosti vztahů mezi veličinami. K poznání a matematickému popisu statistických závislostí slouží metody **regresní analýzy.**

2 Regresní funkce

Definice [7]**:** Nechť existují X a Y, které jsou náhodnými veličinami. Podmíněnou střední hodnotu $\boldsymbol{E(Y|x)}$, považovanou za funkci proměnné x, budeme nazývat regresní funkcí náhodné veličiny Y vzhledem k X. Regresní funkce vyjadřuje změny podmíněné střední hodnoty jedné náhodné veličiny při změně hodnot druhé náhodné veličiny. Graf regresní funkce nazýváme regresní křivka.

Jak již bylo napsáno výše, pomocí regresní funkce můžeme předpovídat, jaké hodnoty nabude jedna náhodná veličina, když známe hodnotu druhé náhodné veličiny. Protože Y je náhodná veličina, nemusí vždy při dané hodnotě x náhodné veličiny X nabýt hodnoty $E(Y|x)$, ale bude nabývat hodnoty rozptýlené okolo ní, což vyplývá z vlastností náhodných veličin. Hlavním úkolem regresní analýzy je zjištění tvaru stochastické závislosti a parametrů regresní funkce. V regresní analýze se budeme zabývat závislostí náhodné veličiny Y na veličině X (nezávisle proměnné), která může být obecně m – rozměrná. [7].

2.1 Jednoduchý model lineární regrese

Jednoduchý model lineární regrese je takovým regresním modelem, kdy grafem regresní funkce je přímka. Pro větší srozumitelnost u parametrů β_0 a β_1 použijeme označení α a β. Předpokládejme, že Y_n je n-tice nekorelovaných (nezávislých) náhodných veličin s vlastnostmi střední hodnoty $EY_i = \alpha + \beta x_i$, $DY_i = \sigma^2$, $i = 1,2,..,n$, kde α, β, σ^2 jsou neznámé parametry a $x_1, x_2,\ldots, x_n$ je n-tice známých hodnot. Pak jednoduchým modelem lineární regrese budeme nazývat model $Y_i = \alpha + \beta x_i + \varepsilon_i$, kde složky ε_i jsou nezávislé náhodné veličiny, pro které platí $E\varepsilon_i = 0, D\varepsilon_i = \sigma^2$, $i = 1,2,..,n$. Tyto složky obsahují působení náhodných vlivů, které nejsou zahrnuty do modelu [7].

Přímka $y = \alpha + \beta x$ se nazývá regresní přímka. β je její směrnice, která udává sklon regresní přímky. Pokud je směrnice kladná, je regresní přímka rostoucí, v opačném případě je klesající. α je konstanta matematického modelu, která nám určuje, v jaké vzdálenosti od počátku přímka vede, přičemž může být i záporná. Abychom dostali platný regresní model, musíme odhadnout neznámé parametry α, β, σ^2 modelu. Tyto odhady budeme po řadě značit *A, B,* S^2. **Bodové odhady parametrů α, β získáme metodou nejmenších čtverců**, která je popsána podrobně v kapitole 3**.**

2.2 Modely lineární vzhledem k parametrům[1]

Modely lineárními vzhledem k parametrům rozumíme modely, kde je závislost popsána regresní funkcí $g(x, \beta_0, \beta_1, \beta_2, ..., \beta_k) = \sum_{i=0}^{k} \beta_i g_i(x)$, kde g_i jsou funkce proměnných $x = (x_1, x_2, ..., x_m)$. Ukázky lineárních modelů jsou uvedeny v následujících rovnicích [7]:

Přímková regrese $$g(x, \beta_0, \beta_1) = \beta_0 + \beta_1 x_1 . \tag{2.0}$$

Parabolická regrese $$g(x, \beta_0, \beta_1, \beta_2) = \beta_0 + \beta_1 x + \beta_2 x^2 . \tag{2.1}$$

Hyperbolická regrese $$g(x, \beta_0, \beta_1) = \beta_0 + \beta_1 \frac{1}{x} . \tag{2.2}$$

Logaritmická regrese $$g(x, \beta_0, \beta_1) = \beta_0 + \beta_1 \log x . \tag{2.3}$$

Regresní rovina $$g(x, \beta_0, \beta_1, \beta_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 . \tag{2.4}$$

2.2.1 Přímková regrese

Nejpoužívanějším typem regresní funkce je přímková regrese uvedená ve tvaru $y = \beta_0 + \beta_1 x$. Stanovíme odhady parametrů β_0 a β_1. K odhadům parametrů používáme metodu nejmenších čtverců, formulovanou podmínkou $\sum_{\mathrm{i}=1}^{\mathrm{n}} (y_i - \hat{y}_i)^2$, aby byl součet čtverců chyb minimální. Dosadíme-li do této podmínky rovnici regresní, dostaneme:

$$Q = \sum_{i=1}^{n} \varepsilon^2{}_i = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2 . \tag{2.5}$$

Součet čtverců Q je funkcí neznámých parametrů. Pro určení minima je nutné vypočítat první parciální derivaci podle β_j, kde *j=0,1*. V dalším kroku tyto derivace potom

---

[1] Podkapitoly upraveny ze zdroje Hindls, R., Hronová, S., Seger, J. *Statistika pro ekonomy*.

položit rovny nule. Parametry β_j nahradíme jejich odhady b_j, $j=0,1$. Zdůrazňuji, že parciální derivace jsou v tomto případě podle odhadů b_0 v případě první rovnice a b_1 v případě druhé rovnice (2.6). Ostatní členy jsou konstanty.

$$2\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)(-1) = 0\,, \tag{2.6}$$

$$2\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)(-x_i) = 0\,.$$

Po následném počítání se sumami a po úpravě dostaneme dvě normální rovnice:

$$\sum_{i=1}^{n} y_i = nb_0 + b_1\sum_{i=1}^{n} x_i\,, \tag{2.7}$$

$$\sum_{i=1}^{n} y_i x_i = b_0\sum_{i=1}^{n} x_i + b_1\sum_{i=1}^{n} x_i^{\,2}\,.$$

Veličiny n, $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} y_i$, $\sum_{i=1}^{n} x_i^{\,2}$ a $\sum_{i=1}^{n} y_i x_i$ můžeme vypočítat z empirických pozorování. Proto je nutné určit pouze odhady parametrů β_0 a β_1 řešením soustavy rovnic (2.7). Pro výpočet odhadů b_0 a b_1 použijeme Cramerovo pravidlo, které se používá pro řešení soustavy lineárních algebraických rovnic. Použitím **Cramerova pravidla** dostáváme :

$$b_0 = \frac{\begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i x_i & \sum_{i=1}^{n} x_i^{\,2} \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^{\,2} \end{vmatrix}} = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^{\,2} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}\,, \tag{2.8}$$

$$b_1 = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i x_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}} = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \quad .$$

2.2.2 Parabolická regrese

Parabolická regrese má tvar $y = \beta_0 + \beta_1 x + \beta_2 x^2$. Pokud tuto rovnici opět dosadíme do podmínky nejmenších čtverců $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ a budeme aplikovat postup použitý u přímkové regrese, dostaneme formulaci znázorněnou vztahem (2.9).

$$Q = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i - \beta_2 x^2{}_i \right)^2 . \tag{2.9}$$

Po výpočtu prvních parciálních derivací výrazu podle β_0, β_1 a β_2 nahradíme obecně β_j jejich odhady b_j, $j=0,1,2$ a pak parciální derivace položíme rovny nule. Bude tedy platit:

$$2\sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i - b_2 x^2{}_i \right) \left(-1 \right) = 0 , \tag{2.10}$$

$$2\sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i - b_2 x^2{}_i \right) \left(-x_i \right) = 0 ,$$

$$2\sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i - b_2 x^2{}_i \right) \left(-x_i^2 \right) = 0 .$$

Po úpravě máme tři normální rovnice, jejichž řešením získáme odhady parametrů β_0, β_1 a β_2.

Rovnice mají tvar:

$$\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} x_i + b_2 \sum_{i=1}^{n} x_i^2 , \tag{2.11}$$

$$\sum_{i=1}^{n} y_i x_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^{\,2} + b_2 \sum_{i=1}^{n} x_i^{\,3} \, ,$$

$$\sum_{i=1}^{n} y_i x_i^{\,2} = b_0 \sum_{i=1}^{n} x_i^{\,2} + b_1 \sum_{i=1}^{n} x_i^{\,3} + b_2 \sum_{i=1}^{n} x_i^{\,4} \, .$$

Odhady b_0, b_1, b_2 parametrů β_0, β_1 a β_2 získáme vyřešením soustavy rovnic (2. 11).

2.2.3 Polynomická regrese

Zobecněním předcházejících typů regresních funkcí je polynomická regrese ve tvaru

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p \, .$$

Postupujeme jako u paraboly a dostaneme:

$$2\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i \dots - b_p x_i^{\,p}\right)\left(-1\right) = 0 \, , \tag{2.12}$$

$$2\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i \dots - b_p x_i^{\,p}\right)\left(-x_i\right) = 0 \, ,$$

…………………………………………..

$$2\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i \dots - b_p x_i^{\,p}\right)\left(-x_i^{\,p}\right) = 0 \, .$$

Po úpravě dostaneme soustavu rovnic, které mají následující tvar:

$$\sum_{i=1}^{n} y_i = n b_0 + b_1 \sum_{i=1}^{n} x_i + \dots + b_p \sum_{i=1}^{n} x_i^{\,p} \, , \tag{2.13}$$

$$\sum_{i=1}^{n} y_i x_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^{\,2} + \dots + b_p \sum_{i=1}^{n} x_i^{\,p+1} \, ,$$

…………………………………………………..

$$\sum_{i=1}^{n} y_i x_i^{\,p} = b_0 \sum_{i=1}^{n} x_i^{\,p} + b_1 \sum_{i=1}^{n} x_i^{\,p+1} + \dots + b_p \sum_{i=1}^{n} x_i^{\,2p} \, .$$

V praxi se setkáváme nejvýše s polynomy 3. až 4. stupně. I v tomto případě dostaneme odhady parametrů $\beta_0, \beta_{1,} ..., \beta_p$ vyřešením předchozí soustavy rovnic (2.13).

2.2.4 Hyperbolická regrese

Hyperbolická regrese má tvar $y = \beta_0 + \frac{\beta_1}{x}$. Při použití stejného postupu, sloužícího k odhadu parametrů jako v předcházejících případech, získáme metodou nejmenších čtverců soustavu normálních rovnic.

$$\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} \frac{1}{x_i}, \tag{2.14}$$

$$\sum_{i=1}^{n} \frac{y_i}{x_i} = b_0 \sum_{i=1}^{n} \frac{1}{x_i} + b_1 \sum_{i=1}^{n} \frac{1}{x_i^2}.$$

Řešením těchto rovnic dostaneme odhady parametrů:

$$b_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} \frac{1}{x_i^2} - \sum_{i=1}^{n} \frac{1}{x_i} \sum_{i=1}^{n} \frac{y_i}{x_i}}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^2}, \tag{2.15}$$

$$b_1 = \frac{n \sum_{i=1}^{n} \frac{y_i}{x_i} - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \frac{1}{x_i^2}}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^2}.$$

2.2.5 Logaritmická regrese

Poslední z funkcí lineárních v parametrech, o které se chci stručně zmínit pro její použitelnost je funkce $y = \beta_0 + \beta_1 \log x$. Metodou nejmenších čtverců dostaneme normální rovnice:

$$\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} \log x_i, \tag{2.16}$$

$$\sum_{i=1}^{n} y_i \log x_i = b_0 \sum_{i=1}^{n} \log x_i + b_1 \sum_{i=1}^{n} \log^2 x_i \, .$$

2.3 Modely nelineární vzhledem k parametrům

Tyto modely je možné vhodnou transformací upravit na lineární tvar vzhledem k parametrům. Odhady parametrů modelů lineárních a modelů, které je možno transformovat na lineární tvar se provádějí nejčastěji metodou nejmenších čtverců [6], která bude popsána v kapitole 3. Ukázky nelineárních modelů jsou uvedeny v následujících rovnicích [7]:

Regresní mocninná funkce $$g(x,\ \beta_0,\ \beta_1) = \beta_0 \cdot x^{\beta_1} \, . \tag{2.17}$$

Regresní exponenciální funkce $$g(x,\ \beta_0,\ \beta_1) = \beta_0 \cdot \beta_1^{x} \, . \tag{2.18}$$

2.3.1 Exponenciální regrese[2]

Parametry funkcí, které nejsou lineární neodhadujeme metodou nejmenších čtverců přímo, protože její použití vede k soustavě nelineárních rovnic. Proto najdeme jejich vhodný počáteční odhad a postupným zlepšováním řešení nalezneme odhad s požadovanou přesností. Metod počátečního odhadu je celá řada. Jejich numerické řešení však bývá někdy zdlouhavé a následné nalezení vhodného počátečního odhadu nemusí být jednoznačné. Ukážeme si jednoduchý způsob, kdy určitou regresní funkci, která není lineární z hlediska parametrů, můžeme pomocí **linearizující transformace** na funkci lineární v parametrech převést.

Transformace spočívá v tom, že pomocí logaritmů, převrácením hodnot a dalšími úpravami dojdeme k takovému tvaru regresní funkce, že její parametry bude už možné odhadnout metodou nejmenších čtverců. Exponenciální funkce má tvar $y = \beta_0 \cdot \beta_1^{x}$

Provedeme **logaritmickou transformaci:**

$$\log y = \log \beta_0 + x \log \beta_1 \tag{2.19}$$

Po linearizaci (2.19) již můžeme postupovat stejně jako v případě lineární regrese s tím rozdílem, že podmínka metody nejmenších čtverců bude v logaritmickém tvaru (2.20, 2.21):

[2] Upraveno ze zdroje Hindls, R., Hronová, S., Seger, J. *Statistika pro ekonomy.*

$$Q = \sum_{i=1}^{n} \left(\log y - \log \hat{y}_i \right)^2 , \tag{2.20}$$

$$Q = \sum_{i=1}^{n} \left(\log y_i - \log \beta_0 - x_i \log \beta_1 \right)^2 . \tag{2.21}$$

Stejným způsobem dostaneme normální rovnice:

$$\sum_{i=1}^{n} \log y_i = n \log b_0 + \log b_1 \sum_{i=1}^{n} x_i , \tag{2.22}$$

$$\sum_{i=1}^{n} x_i \log y_i = \log b_0 \sum_{i=1}^{n} x_i + \log b_1 \sum_{i=1}^{n} x_i^2 .$$

Jejich řešením dostaneme:

$$\log b_0 = \frac{\sum_{i=1}^{n} \log y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \log y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} , \tag{2.23}$$

$$\log b_1 = \frac{n \sum_{i=1}^{n} x_i \log y_i - \sum_{i=1}^{n} \log y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} .$$

Odhady parametrů β_0 a β_1 jsou potom $b_0 = 10^{\log b_0}$ a $b_1 = 10^{\log b_1}$.

2.4 Modely obtížně linearizovatelné vzhledem k parametrům

Tyto nelineární modely se nedají jednoduše transformovat na lineární tvar

$$g(x,\ \beta_0,\ \beta_{1,}\ \beta_2) = \beta_0 \cdot \beta_1^{x} + \beta_2 \tag{2.24}$$

Není proto vhodné použít metodu nejmenších čtverců pro odhady parametrů. Používáme jiné metody, např. metodou částečných součtů, metodou dílčích průměrů, nebo metodou vybraných bodů [7].

3 Metoda nejmenších čtverců

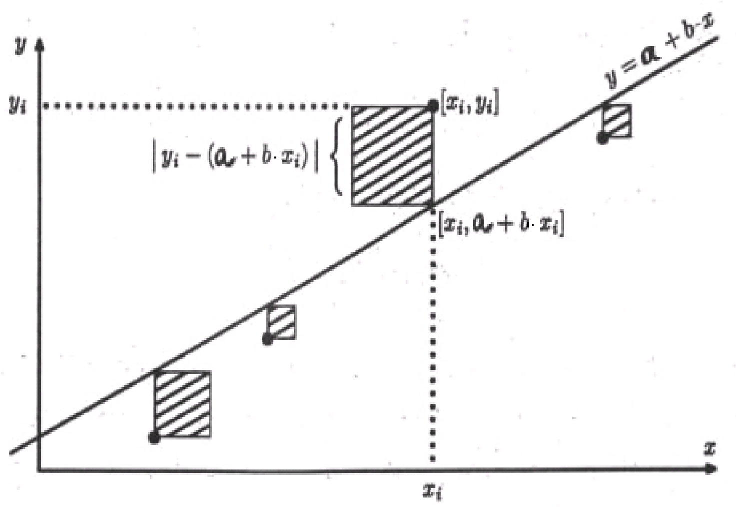
Až dosud jsme se zabývali odhadem neznámých parametrů pro průběh lineární regrese, aniž bychom si o metodě odhadu těchto parametrů řekli něco podrobnějšího. Proto je potřeba si říci o principech této metody více.

Parametry empirických regresních funkcí se nejčastěji určují **metodou nejmenších čtverců.** Metodu nejmenších čtverců lze použít ke stanovení parametrů i jiných funkcí, než je přímka. Tato metoda je použitelná ke stanovení parametrů všech funkcí, jež jsme nazvali lineární regresní funkce [3].

Předpokládejme, že máme konkrétní dvojice naměřených hodnot $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Hledáme takovou funkci (**odhad**) $\hat{y} = a + bx$, aby v jistém smyslu co nejvíce „přiléhala“ k bodům $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, kde „přiléhání“ měříme součtem rozdílů $\hat{y}$ – y_i (tzv. reziduí). Jde tedy v podstatě o to, že chceme odhadnout reálné souřadnicové body, aby rozdíly mezi skutečnými hodnotami a těmi odhadnutými byly co nejmenší. Aby se ovšem nestalo, že při značných odchylkách mezi $\hat{y}$ a y_i se kladné a záporné rozdíly navzájem odečtou, vezmeme jako míru přiléhání ne prostý součet reziduí, ale **součet jejich čtverců** [7]**.** Můžeme tedy říci, že dvojice $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ jsou počáteční naměřená vstupní data, která budou pomocí nejmenších čtverců proložena například regresní přímkou $\hat{y} = a + bx$. Nám jde o to, aby rozdíl mezi prokládanými skutečnými daty a jejich odhadem daným regresní funkcí byl co možná nejmenší. Tento vztah vyjadřuje vzorec (3.0) [7].

$$\sum_{\mathrm{i=1}}^{\mathrm{n}} (y_i - \hat{y}_i)^2 = \min . \qquad (3.0)$$

Na následujícím obrázku (Obrázek 1) je metoda zachycena graficky. Stručně jej lze popsat tak, že dvojice (x_i, y_i) je *i*-tá hodnota skutečně naměřených bodů. Body regresní přímky $\hat{y} = a + bx$ jsou odhadem skutečných hodnot. Čím menší jsou čtverce vzdáleností (vztah 3.0), tím lepší je odhad a proložení regresní funkce.



Obrázek 1: Metoda nejmenších čtverců [13]

Budeme tedy v jednoduchém lineárním modelu hledat minimum funkce, kde *A*, *B* jsou odhady konstanty *a* a směrnice *b* dané přímky, *Y* je náhodnou veličinou. Následující matematické vztahy jsou převzaty ze zdroje [7].

$$S(A,B) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - A - Bx_i)^2 , \tag{3.1}$$

Hledáme parciální derivaci (směrovou dle jedné a následně druhé proměnné) prvního řádu:

$$\frac{\partial S}{\partial A} = -2\sum_{i=1}^{n} (Y_i - A - Bx_i) \quad ; \quad \frac{\partial S}{\partial B} = -2\sum_{i=1}^{n} (Y_i - A - Bx_i) \cdot x_i ,$$

Vypočítáme extrém funkce dvou proměnných. Nutnou podmínkou je nulovost obou parciálních derivací prvního řádu:

$$-2\sum_{i=1}^{n} (Y_i - A - Bx_i) = 0 \quad ; \quad -2\sum_{i=1}^{n} (Y_i - A - Bx_i) \cdot x_i = 0 .$$

Po úpravě (jedná se o algebraické úpravy se sumami) dostaneme soustavu normálních rovnic:

$$nA + B\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i , \tag{3.2}$$

$$A\sum_{i=1}^{n} x_i \;+ B\sum_{i=1}^{n} x_i^{\,2} = \;\sum_{i=1}^{n} x_i Y_i \,.$$

Z rovnic (3.2) vypočteme odhady *A*, *B*:

$$B = \frac{n\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2} \;\;;\;\; A = \frac{\sum_{i=1}^{n} x_i^{\,2}\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i Y_i}{n\sum_{i=1}^{n} x_i^{\,2} - (\sum_{i=1}^{n} x_i)^2}\,. \qquad (3.3)$$

K ověření, že funkce S (3.1) nabývá minima musíme určit parciální derivace druhého řádu:

$$\frac{\partial^2 S}{\partial A^2} = 2n\;;\;\; \frac{\partial^2 S}{\partial A \partial B} \;= 2\sum_{i=1}^{n} x_i \;\;;\;\; \frac{\partial^2 S}{\partial B^2} \;= 2\sum_{i=1}^{n} x_i^{\,2}\,.$$

Výsledkem výrazu $\frac{\partial^2 S}{\partial A^2}\cdot\frac{\partial^2 S}{\partial B^2} - \left(\frac{\partial^2 S}{\partial A \partial B}\right)^2$ dostaneme:

$$4n\sum_{i=1}^{n} x_i^{\,2} - 4\left(\sum_{i=1}^{n} x_i\right)^2 = 4n\left[\sum_{i=1}^{n} x_i^{\,2} - 2\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = 4n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \Rightarrow$$

$$4n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 > 0.$$

Podle věty z diferenciálního počtu víme, že pokud výraz $\frac{\partial^2 S}{\partial A^2}\cdot\frac{\partial^2 S}{\partial B^2} - \left(\frac{\partial^2 S}{\partial A \partial B}\right)^2$ nabývá kladné hodnoty a také parciální derivace $\frac{\partial^2 S}{\partial A^2} = 2n$ a $\frac{\partial^2 S}{\partial B^2} = 2\sum_{i=1}^{n} x_i^{\,2}$ nabývají kladných hodnot (což platí), má funkce *S* ostré lokální minimum.

Regresní přímka, získaná takto metodou nejmenších čtverců, má tvar $\hat{y} = A + Bx$. Uvedenou rovnici lze upravit na tvar $\hat{Y} = \overline{Y} + B\,(x - \overline{x})$[7].

Dále dokážeme, že odhady *A*, *B* parametrů α, β jsou nevychýlené. Pro toto tvrzení musí platit $EB = \beta$ a $EA = \alpha$.

Nejprve upravíme tvar odhadu parametru β (vzorec 3.3):

$$B = \frac{n\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} \overline{x} Y_i}{\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + n\overline{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) Y_i}{\sum_{i=1}^{n} (x_i - \overline{x})^2} =$$

$$\sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \cdot Y_i .$$

Nyní ověřujeme $EB = \beta$

$$EB = E\sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \cdot Y_i = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \cdot EY_i = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \cdot (\alpha + \beta x_i) =$$

$$\alpha \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} + \beta \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right) x_i}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} = \beta \text{ protože } \sum_{i=1}^{n}\left(x_i - \overline{x}\right) = 0 ; \sum_{i=1}^{n}\left(x_i - \overline{x}\right) x_i = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 .$$

Dále ověříme $EA = \alpha$

$$EA = E\,\frac{\sum_{i=1}^{n} Y_i - B\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\sum_{i=1}^{n} EY_i - \frac{1}{n}\sum_{i=1}^{n} x_i EB = \frac{1}{n}\left[\sum_{i=1}^{n}(\alpha + \beta x_i) - \sum_{i=1}^{n}\beta x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\alpha = \alpha .$$

Dokázali jsme, že odhady *A*, *B* parametrů α, β pomocí metody nejmenších čtverců jsou nevychýlené. Funkce $\hat{Y} = A + Bx$ je tedy nevychýleným odhadem regresní přímky $y = \alpha + \beta x$ [7].

4 Volba regresní funkce

4.1 Věcně ekonomická kritéria

Vhodná regresní funkce by měla být zvolena na základě věcného rozboru vztahů mezi veličinami. Základem pro rozhodování o vhodném typu regresní funkce by měla být věcná kritéria. Při věcné analýze založené na platné teorii lze v některých případech posoudit, zda jde o funkci rostoucí či klesající, jaký je smysl zakřivení, přichází-li v úvahu inflexní bod či nikoliv. Zda jde o funkci nekonečně rostoucí nebo naopak o funkci s růstem ke konečné limitě. K tomu nám samozřejmě může dobře posloužit vyšetření průběhu funkce dle obecných postupů. Lze získat i předběžné informace o parametrech modelu apod. Jindy lze použít při volbě regresní funkce zkušenosti získané s použitím určitého typu regresní funkce již v minulosti. Jde-li o závislost, která byla již jednou popsána, stačí ověřit, zda nedošlo k takové změně podmínek nebo zkoumaného jevu, který by měl vliv na výběr regresní funkce [5]. Toto všechno jsou postupy označované v odborné literatuře jako věcně ekonomické nástroje, které nám mohou usnadnit rozmýšlení, který regresní model prokládá danou závislost mezi proměnnými nejlépe.

4.2 Empirický způsob volby

Nestanovíme-li vhodný typ regresní funkce na základě ekonomických kritérií, uchylujeme se k empirickému způsobu volby. Základní metodou je *metoda grafická*, kdy průběh závislosti znázorňujeme ve formě bodového diagramu. Každá dvojice x a y zde tvoří jeden bod tohoto grafu. Podle průběhu proložení bodového grafu rozhodujeme, jaký typ konkrétní regresní funkce (přímka, parabola...) je pro popis sledované závislosti nejvhodnější. K tomu, abychom zhodnotili kvalitu získané regresní funkce a eventuálně i posoudili oprávněnost některých předpokladů, které souvisejí s uplatněním požadovaných metod odhadu používáme různá matematicko – statistická kritéria [6].

Ekonomická i matematicko-statistická kritéria mají své výhody i nevýhody. Podle zastánců ekonomických kritérií dobrý ekonomický rozbor situace umožňuje nalézt vhodný typ funkce. Zastánci používání matematicko-statistických kritérií se naopak přiklání k názoru, že kvalifikovaný rozbor číselných údajů je schopen jednoznačně určit tvar „nejlepší“ regresní funkce bez znalosti zkoumaných ekonomických veličin[6].

5 Kvalita regresní funkce a intenzita závislosti

Jedním z úkolů regresní analýzy je posouzení kvality regresní funkce a zjištění intenzity (síly, těsnosti) závislosti. Regresní funkce je tím lepší, čím je posuzovaný vztah silnější, a čím více jsou empirické hodnoty vysvětlované proměnné soustředěné kolem odhadu regresní funkce. Naopak vztah je tím slabší, čím více jsou empirické hodnoty vzdáleny hodnotám vyrovnaných *Y*. Míra intenzity závislosti úzce souvisí s hodnocením účinnosti odhadnuté regresní funkce a tedy s kvalitou regresního odhadu. Pro kvalitu regresní funkce používáme zejména charakteristiky jako index korelace, index determinace, rozptyl a analýzu reziduí. Tyto charakteristiky jsou v jednotlivých podkapitolách vysvětleny.

5.1 Rozptyly empirických, vyrovnaných a skutečných hodnot

Můžeme zkonstruovat tři rozptyly se zcela odlišnou vypovídající schopností [5]:

a) Rozptyl empirických hodnot: $s_y^2 = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2$, (5.0)

b) rozptyl vyrovnaných hodnot: $s_Y^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \overline{y} \right)^2$,

c) rozptyl skutečných hodnot (**reziduální rozptyl):** $s^2_{(y-Y)} = \frac{1}{n} \sum_{i=1}^{n} (y_i - Y_i)^2$.

Při použití metody nejmenších čtverců mezi uvedenými rozptyly platí: $s_y^2 = s_Y^2 + s^2_{(y-Y)}$

Rozptyl empirických hodnot můžeme rozložit na rozptyl vyrovnaných hodnot a rozptyl reziduálních hodnot. Všechny empirické hodnoty by byly zároveň hodnotami vyrovnanými, kdyby mezi závisle proměnnou *y* a vysvětlující proměnnou *x* existovala funkční závislost. Potom by se rozptyl empirických hodnot rovnal rozptylu vyrovnaných hodnot a reziduální rozptyl by byl nulový. Platilo by $s_y^2 = s_Y^2$. Pokud by existovala úplná nezávislost mezi oběma proměnnými, pak by všechny vyrovnané hodnoty byly stejné a jejich rozptyl nulový. **Vhodná je regresní funkce s menším rozptylem** [5]**.**

5.2 Index determinace

Z uvedeného vyplývá, že intenzita závislosti bude zřejmě tím silnější, čím větší bude podíl rozptylu vyrovnaných hodnot na celkovém rozptylu. Naopak intenzita bude tím slabší, čím bude podíl tohoto rozptylu menší. Toto určuje, jakou část variability sledovaných hodnot je možné vysvětlit daným modelem. Parametr může nabývat hodnot v intervalu <0,1> [6]. Sílu závislosti je tedy možné měřit poměrem rozptylu vyrovnaných a empirických hodnot.

$$I^2_{yx} = \frac{s^2_Y}{s^2_y} \tag{5.1}$$

Tento poměr se nazývá **index determinace** [5] a v případě lineární regrese se značí jako koeficient R^2. U funkční závislosti bude jeho hodnota rovna 1, v případě nezávislosti nabude hodnoty nula. Čím více se bude blížit jedné, tím je závislost silnější a dobře vystihuje regresní funkci. Čím více se bude blížit nule, tím považujeme danou závislost za slabší a regresní funkci za méně výstižnou [6]. Během hodnocení na základě indexu determinace je také třeba uvažovat, že jeho velikost je ovlivněna tím, zda se nám podařilo nalézt vhodný typ regresní funkce pro popis dané závislosti. Vyjde-li potom nízká hodnota indexu determinace, nemusí to ještě znamenat nízký stupeň závislosti mezi proměnnými, ale může to signalizovat chybnou volbu regresní funkce.

Vhodnější je model s vyšším indexem determinace. Je třeba vzít v úvahu, že hodnota indexu determinace bývá vyšší pro regresní funkce s větším počtem parametrů. Proto je vhodné ověřit volbu vhodného modelu dalšími testy.

5.3 Index korelace

K měření těsnosti závislosti se v praxi obyčejně nepoužívá pouze samotného indexu determinace, ale také jeho odmocniny, kterou nazýváme **index korelace** [5]**.**

$$I_{yx} = \sqrt{\frac{s^2_Y}{s^2_y}} \tag{5.2}$$

Index korelace poskytuje stejné informace o těsnosti závislosti jako index determinace. Tento index však má menší vypovídací schopnost [5].

5.4 Analýza reziduí

Umožňuje posoudit vhodnost zvolené regresní funkce podle průběhu reziduí $e_i = y_i - Y_i$ [6]. Rezidua zobrazená v závislosti na hodnotách proměnné x_i umožňují ověřit vhodnost tvaru regresní funkce a splnění předpokladu konstantnosti rozptylu. Rezidua zobrazená v závislosti na pořadí pozorování umožňují odhalit porušení předpokladu nezávislosti. Rezidua zobrazená v závislosti na hodnotách v modelu dosud nezařazených proměnných ukazují, zda je vhodné zařadit příslušnou proměnnou do modelu. **Vhodný je model s nízkými hodnotami reziduí** [6]**.**

6 Cíl regresní analýzy

Cílem regresní analýzy je přispět k poznání příčinných vztahů mezi statistickými znaky. Úkolem regresní analýzy je také matematický popis systematických okolností, které provázejí statistické závislosti. Je zde snaha nalézt „idealizující" matematickou funkci tak, aby co nejlépe vyjadřovala charakter závislostí a co nejlépe zobrazovala průběh změn podmíněných průměrů závislé proměnné. Tato hypotetická matematická funkce se nazývá regresní funkce. Záměrem analýzy je co nejlepší přiblížení empirické regresní funkce k hypotetické regresní funkci. Pro hlavní cíle regresní analýzy je nutno splnit řadu dílčích úkolů. Některé z nich jsou například [5]:

a) Shromáždit a matematicky formulovat apriorní představy o charakteru regresní funkce,

b) formulovat naše představy o souhrnném působení neuvažovaných statistických znaků,

c) odhadnout empirickou regresní funkci na základě statistických pozorování,

d) posoudit kvalitu empirické regresní funkce z hlediska důvodů a cílů statistického zjišťování.

Zvolený typ regresní funkce by měl respektovat zákonitosti i souvislosti jednotlivých náhodných jevů. Při volbě typu regresní funkce se přihlíží k tomu, aby zvolený model byl nejjednodušší a zároveň aby odchylky teoretických a empirických hodnot byly minimální. Rozhodování často usnadní sestavení bodového diagramu, kterým se příslušná regresní funkce proloží.

7 Příklad

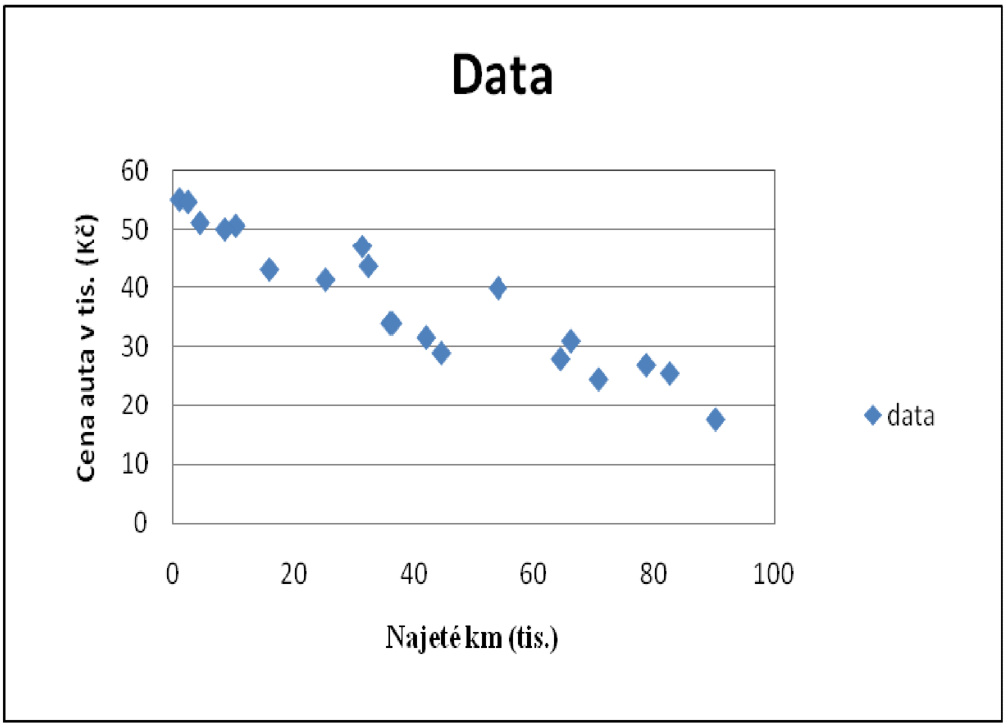
Po teoretické části, jejímž cílem bylo objasnit podstatu regresní analýzy se zaměřím na praktické příklady. Oba příklady jsou zpracovány v tabulkovém programu MS Excel využitím nainstalovaného balíku Analýza dat (Příloha 1, 2, 3), který obsahuje metody regrese. Zadání tohoto příkladu je ze sbírek úloh ze statistiky od Mgr. Slívka z interních zdrojů VOŠ Česká Třebová [10]. Nejprve je uvedeno zadání příkladu spolu s tabulkou a grafem dat. V dalších částech jsou zobrazeny jednotlivé regresní funkce a podstatné údaje z regresní analýzy zpracované nástrojem MS Excel. Následuje výběr optimální regresní funkce.

7.1 Zadání

U dvaceti prodaných ojetých automobilů určité značky byla zjištěna cena *y* [tis. Kč] a počet najetých kilometrů [tis. km] *x*. Nalezněte nejvhodnější regresní křivku vystihující závislost ceny automobilů na počtu najetých kilometrů.

Tabulka 1: Data

x (km)	*y* (cena)
1,1	55
2,5	54,6
10,4	50,6
4,5	51,1
31,4	47
8,6	50
32,4	43,6
25,3	41,3
16	43
54	39,9
36	34
66,2	31
44,5	29
42	31,6
36,4	34
82,6	25,6
64,5	28
70,8	24,6
78,7	27
90,2	17,6



Obrázek 2: Vstupní data

Výše jsou zobrazena data v tabulce (Tabulka 1), kde *x* (najeté kilometry) vyjadřují nezávislé proměnné a *y* zobrazují závislé proměnné (cena vozů). Cena automobilu je závislá na najetých km. Bodový graf (Obrázek 2) vedle tabulky s daty zobrazuje vstupní hodnoty.

7.2 Analýza regresní přímky

Níže můžeme vidět (Obrázek 3, str. 30), že regresní přímka sestrojena z nezávislých proměnných *x* a očekávaných hodnot *Y* (Tabulka 2, str. 29) vytvořených pomocí regresní analýzy dobře prokládá data a tedy vystihuje závislost. K posouzení závislosti jak v tomto případě, tak i v případech ostatních budeme používat jednak vizuálního hodnocení proložení dat regresní funkcí a hlavně ukazatele regresní analýzy. Tyto ukazatele jsou pro každou regresní funkci zobrazeny ve vytvořené *regresní statistice (*pod obrázkem 3 v případě regresní přímky*)*, kde použijeme hodnotu spolehlivosti R (index korelace), násobné R (index determinace) a chybu střední hodnoty (residuální směrodatná odchylka s). O indexu korelace a indexu determinace víme z teoretické části, že čím větší hodnoty mají, tím je model lepší. Chyba střední hodnoty by naopak měla být co nejnižší. V ukazatelích ANOVA, které jsou také vytvořeny pro každou regresní funkci v rámci analýzy, se zaměříme na tučně zvýrazněná rezidua a významnost F. I tyto parametry by měly být co nejnižší.

Jak již bylo napsáno výše, regresní funkce přímky dobře prokládá data. Potom můžeme říci, že hodnoty skutečně naměřených dat se příliš neodchylují od jejich odhadů vyjádřených regresní funkcí. Tomu by odpovídaly i velmi dobré parametry regresní statistiky pod grafem regresní přímky jako je index korelace a index determinace, kdy se oba parametry blíží jedné. Jak víme z teoretické části (Kapitola 5), tyto parametry nám vypovídají o těsnosti závislosti regresní funkce. V případě plné funkční závislosti jedné proměnné na druhé, by oba koeficienty měly hodnotu 1. Chybu střední hodnoty, nebo-li reziduální rozptyl a rezidua můžeme porovnávat až s výsledky ostatních regresních funkcí. Pro reziduální rozptyly a stejně tak pro hodnoty reziduí platí, že čím jsou hodnoty těchto parametrů nižší, tím je regresní model lepší. U obou parametrů je to proto, že regresní křivka je tím lepší, čím blíže je všem naměřeným datům. Právě tuto vzdálenost křivky od skutečně naměřených dat oba parametry popisují. Potom platí, že čím menší jsou hodnoty těchto parametrů, tím menší je vzdálenost odhadu křivky od dat, a tím lepší je proložení dat křivkou.

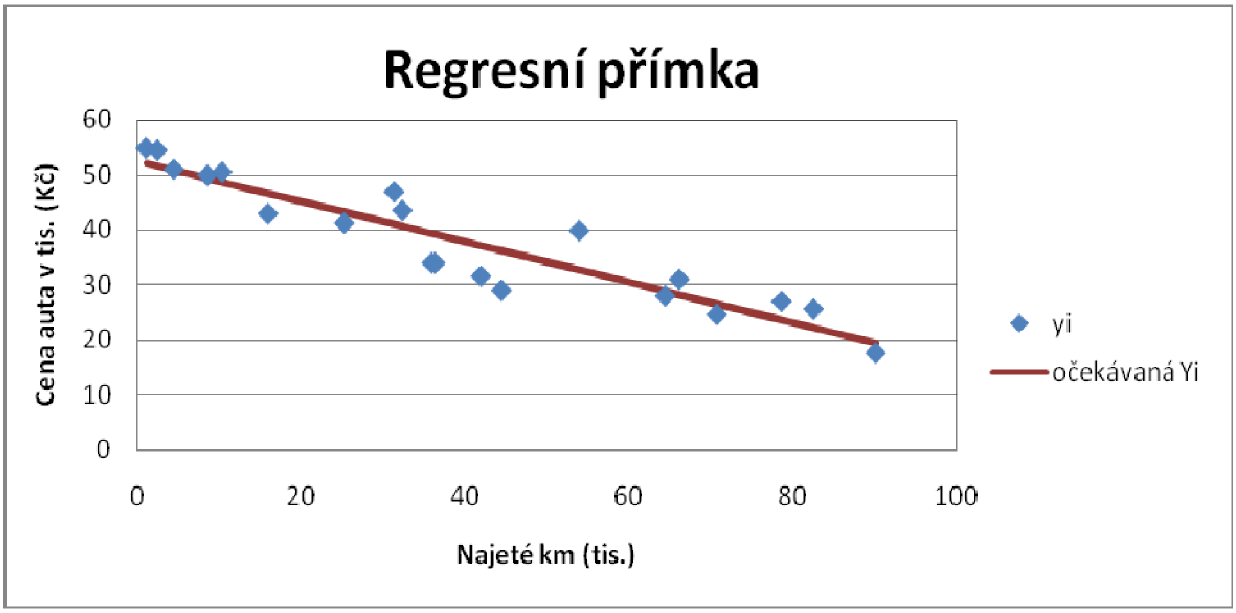
Parametr významnost F, sice není v teoretické části popisován, ale to z toho důvodu, že nás „pouze“ informuje o celkové významnosti statistického modelu. Platí, že čím je regresní model statisticky významnější, tím je hodnota tohoto parametru nižší. **Za statisticky nevýznamný model se pokládá takový model, který by měl parametr významnost F větší jak 0,05**. Pokud toto platí, tak s jistotou 95% zamítáme hypotézu o významnosti modelu.

Potom můžeme říci, že tato regresní přímka je modelem statisticky významným. Jinak o regresní přímce samotné můžeme říci, že parametr b_0 neboli její konstanta protíná osu y zhruba v hodnotě 52,5 a její směrnice b_1 je záporná, protože přímka má tvar klesající.

V tabulce dole (Tabulka 2) jsou obsaženy jednotlivé nezávislé hodnoty *x* (počet ujetých km) a závislé hodnoty *y* (cena). Tato data, na rozdíl od pole *očekávaná Y* jsem do tabulky zadal a jsou vstupními daty. Pole *očekávaná Y* bylo vytvořeno regresním modelem. Totéž platí pro tyto tabulky u ostatních regresních modelů, které jsou dále uváděny v příloze (Příloha 4 až 16). Hodnoty *očekávaná Y* si můžeme představit jako body, kterými naše regresní křivka bude procházet. Je to v podstatě již mnohokrát zmiňovaný odhad regresního modelu v tomto případě pro regresní přímku. Čím jsou tato *očekávaná Y* podobnější skutečným hodnotám *y*, tím lépe model prokládá data. Již z této tabulky bychom alespoň částečně mohli odhadnout, že regresní přímka bude data dobře prokládat, protože ve většině případů odchylka hodnot pole *očekávaná Y* a *y* není velká.

Tabulka 2: Vstupní data a očekávané hodnoty regresní přímky

x	***y***	***Očekávaná Y***
1,1	55	52,15117161
2,5	54,6	51,63792228
10,4	50,6	48,74172963
4,5	51,1	50,90470895
31,4	47	41,04298968
8,6	50	49,40162163
32,4	43,6	40,67638302
25,3	41,3	43,27929033
16	43	46,68873231
54	39,9	32,75767906
36	34	39,35659902
66,2	31	28,28507776
44,5	29	36,24044238
42	31,6	37,15695904
36,4	34	39,20995636
82,6	25,6	22,27272846
64,5	28	28,90830909
70,8	24,6	26,5986871
78,7	27	23,70249445
90,2	17,6	19,48651781



Obrázek 3: Regresní přímka

VÝSLEDEK

Regresní statistika	
Násobné R	**0,932100798**
Hodnota spolehlivosti R	**0,868811898**
Nastavená hodnota spolehlivosti R	0,86152367
Chyba střední hodnoty	**4,150930879**
Pozorování	20

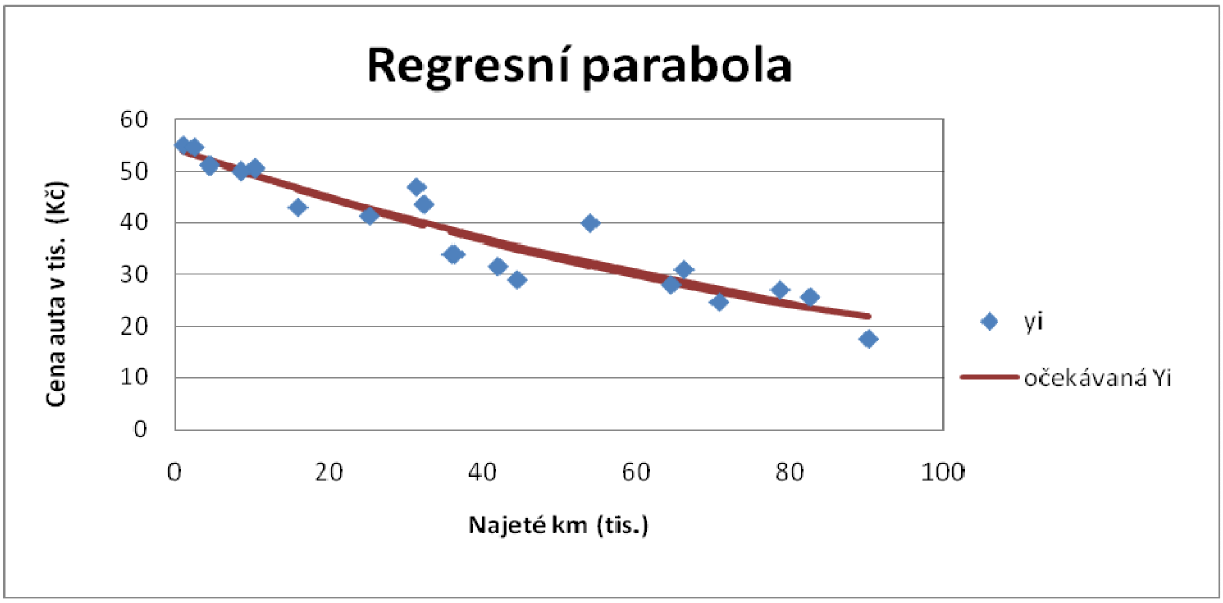
ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	2053,973	2053,97	119,21	**2,2732E-09**
Rezidua	18	**310,1441**	17,2302		
Celkem	19	2364,118			

7.3 Analýza regresní paraboly

Opět je vytvořena tabulka (Příloha 4) s potřebnými daty. Tabulka obsahuje nezávislá i závislá vstupní data *x* a *y*. Mimo očekávaných hodnot *Y* vytvořených regresní analýzou je zde navíc druhá mocnina *x*, kterou zadáváme v MS Excel do nástroje analýzy dat.

Z níže uvedeného grafu (Obrázek 4) můžeme vidět, že i regresní parabola velmi dobře prokládá data. Na tomto příkladu lze poznat, že vizualizace k posouzení optimálního modelu vždy nestačí. K přesnému porovnání (např. mezi přímkou a parabolou) je nutné použít číselné údaje z ANOVY a *regresní statistiky*. Je nutné použít údajů z analýzy dat, protože kvalita regresní funkce není vždy pouhým porovnáním proložení regresních funkcí daty ihned patrná. Kdybychom se spoléhali pouze na náš vlastní úsudek, lehce bychom mohli způsobit chybu. Z regresní analýzy můžeme prozatím zhodnotit, že regresní parabola má o něco lepší parametry než regresní přímka. Index korelace i index determinace, které určují míru těsnosti přiléhání křivky k datům jsou v případě přímky i paraboly blížící se jedné. Významnost F určující statistickou významnost modelu je v obou případech v pořádku. Chybu střední hodnoty neboli reziduální rozptyl čtverců a rezidua má ale parabola lepší.



Obrázek 4: Regresní parabola

VÝSLEDEK

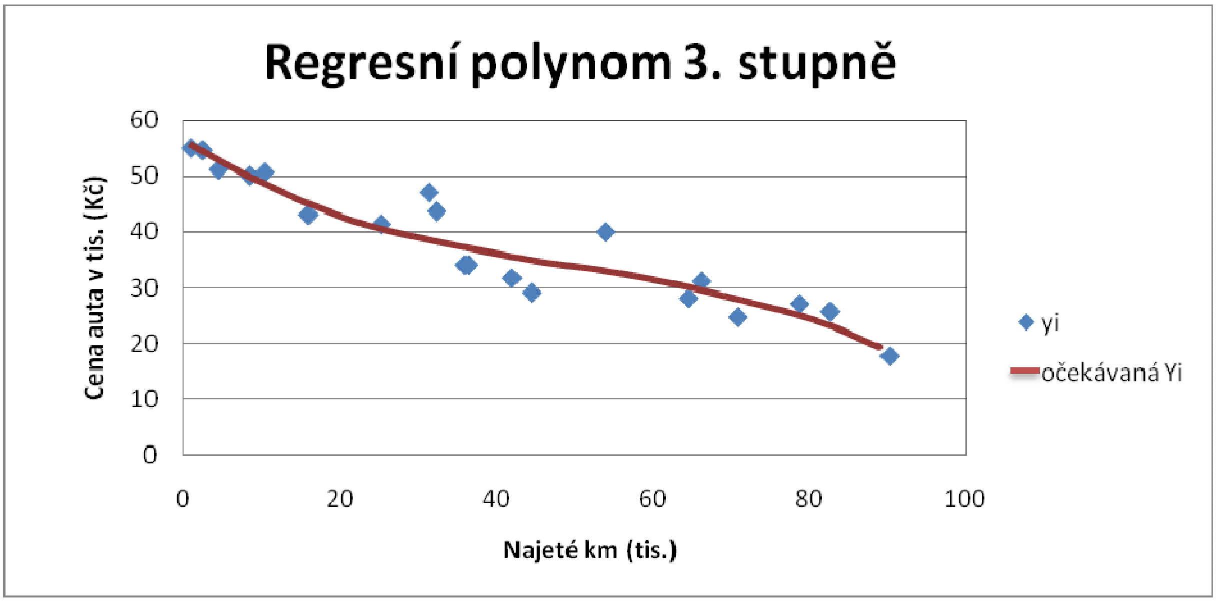
Regresní statistika	
Násobné R	**0,938077576**
Hodnota spolehlivosti R	**0,879989539**
Nastavená hodnota spolehlivosti R	0,865870661
Chyba střední hodnoty	**4,08525924**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	2	2080,398668	1040,199	62,32716	**1,4906E-08**
Rezidua	17	**283,718832**	16,68934		
Celkem	19	2364,1175			

7.4 Analýza regresního polynomu 3. stupně

Vytvořená tabulka (Příloha 5) obsahuje závislé proměnné *y* a nezávisle proměnné *x*. V tabulce je zařazena nově vytvořená proměnná x^3, protože se jedná o polynom 3. stupně. Hodnoty proměnné x^3 spolu s hodnotami *y* byly dále zpracovány pro účely analýzy. Opět jsme dostali hodnoty *očekávaná Y*, vytvořené analýzou regresního polynomu 3. stupně. Polynom na základě porovnání očekávaných *Y* se skutečnými hodnotami *y* (Příloha 5) prokládá vstupní data velmi dobře, protože chyby dané rozdílem obou proměnných nejsou velké. Více než z přílohy se ale jistě dozvíme z grafického výstupu a především z parametrů regresní analýzy.



Obrázek 5: Regresní polynom

Z grafického výstupu vidíme, že polynom dle očekávání optimálně prokládá data. Stejně jako v předchozích případech nevidíme žádné vybočující hodnoty skutečných dat *y*, které by nebyly křivkou proloženy. Proto bude nutné použít údaje z *regresní statistiky* a ANOVY pro konečné porovnání modelů. Výsledky regresní analýzy jsou podobné jako v případě regresního modelu paraboly, ale ve všech případech je tento model ještě lepší. Porovnáme-li index korelace a index determinace, je těsnost přiléhání skutečných dat ke křivce lepší. Pokud porovnáme rezidua a chybu střední hodnoty, zjistíme, že i tyto vlastnosti má polynom prozatím nejlepší.

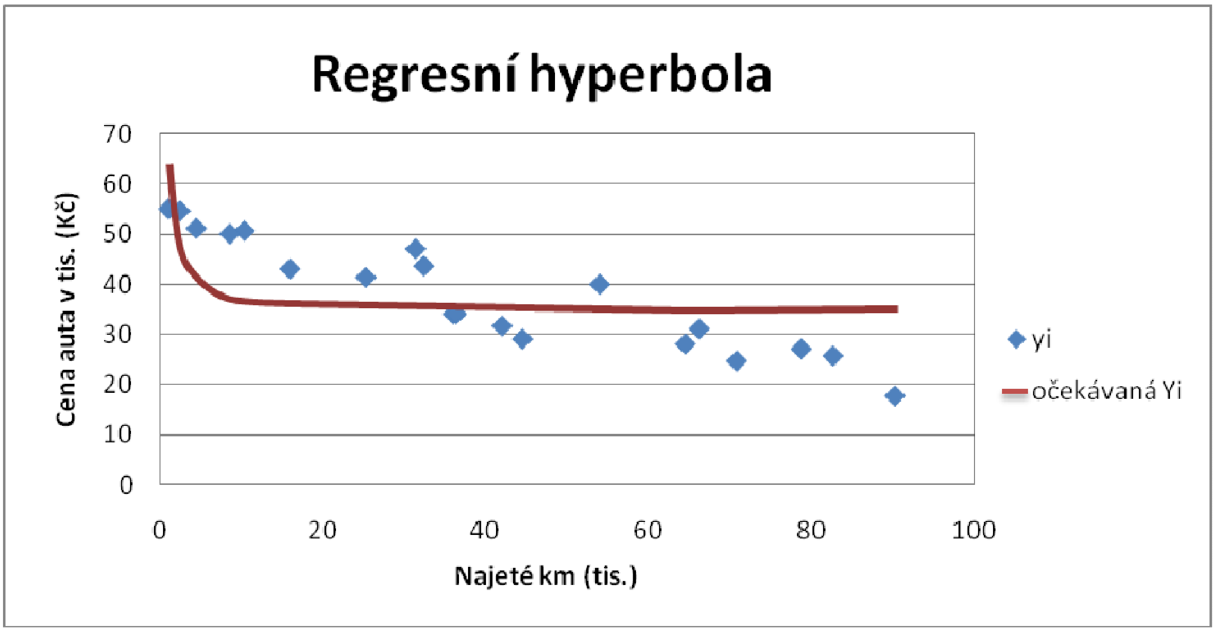
Regresní statistika	
Násobné R	**0,943032237**
Hodnota spolehlivosti R	**0,8893098**
Nastavená hodnota spolehlivosti R	0,868555388
Chyba střední hodnoty	**4,044167395**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	3	2102,433	700,811	42,8492	**7,14368E-08**
Rezidua	16	**261,6846**	16,35529		
Celkem	19	2364,118			

7.5 Analýza regresní hyperboly

Do tabulky vytvořené pro analýzu regresní hyperboly (Příloha 6) jsou opět zanesené hodnoty vstupních proměnných *y* a *x*. Pro účely regresní analýzy byla vytvořená proměnná *1/x*, protože se jedná o hyperbolu a z teoretické části víme, že regresní hyperbola má tvar $y = \beta_0 + \frac{\beta_1}{x}$. Koeficient β_0 bude vypočítán analytickým nástrojem, ale koeficient β_1 rovný jedné jsme zvolili my, protože základní tvar funkce hyperboly je $y = 1/x$. Porovnáním *očekávaných* hodnot *Y* se skutečnými hodnotami *y* v příloze můžeme vidět, že hyperbolická závislost mezi daty patrně nebude nejvhodnější. Více se ale dozvíme z regresní analýzy.



Obrázek 6: Regresní hyperbola

Z obrázku vidíme, že regresní hyperbola dle předpokladu zdaleka neprokládá vstupní data stejně dobře jako předchozí modely. Z grafu je jasně patrný rozdíl mezi skutečnými daty a regresní křivkou. Nevhodně je proložena zejména část křivky u počátku osy x a poté se regresní křivka chová skoro jako přímka bez větších změn. Z tohoto důvodu si myslím, že analýzu dat pomocí MS Excel bychom ani uvádět nemuseli. Pro pořádek si ji ale v tomto i dalších případech uvádět budeme. Alespoň tak poznáme, zda je model statisticky významný. Uvedu jen pro porovnání, že index determinace a index korelace je mnohem horší (nižší) než v předchozích případech. Chyba střední hodnoty a rezidua jsou také významně horší (vyšší). Tedy rozdíl mezi skutečnými daty a odhadovanými hodnotami vyjádřený regresní hyperbolickou závislostí mezi daty je velký. I přesto je model statisticky významný, protože platí 0,0045 < 0,005.

VÝSLEDEK

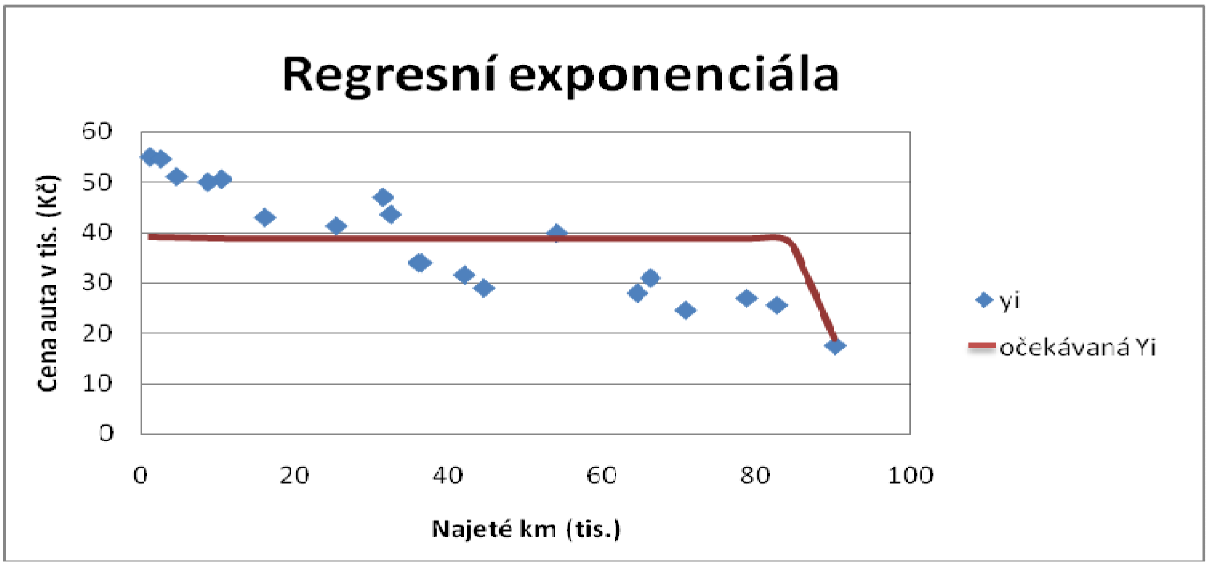
Regresní statistika	
Násobné R	**0,606868177**
Hodnota spolehlivosti R	**0,368288985**
Nastavená hodnota spolehlivosti R	0,333193928
Chyba střední hodnoty	**9,108723128**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	870,6784	870,6784	10,49404	**0,00454991**
Rezidua	18	**1493,439**	82,96884		
Celkem	19	2364,118			

7.6 Analýza regresní exponenciály

Již z tabulky uvedené v příloze (Příloha 7) vidíme, že regresní model exponenciální závislosti mezi daty bude asi zcela nepoužitelný, protože rozdíl mezi *očekávanými Y* a skutečnými *y* je velký. Navíc se tento model chová u většiny hodnot téměř konstantně a nemá ani „snahu“ data optimálně prokládat.



Obrázek 7: Regresní exponenciála

Z obrázku vidíme, že tento model opravdu není vhodný. Parametry regresní analýzy jsou přesto pro srovnání uvedeny, ale dále se jimi zabývat nebudeme. Nebudeme se jimi zabývat, protože se stačí podívat na parametr **významnost F**. Jeho hodnota v případě regresní exponenciály je vyšší než 0,05 a tedy **model zamítáme jako statisticky nevýznamný.**

VÝSLEDEK

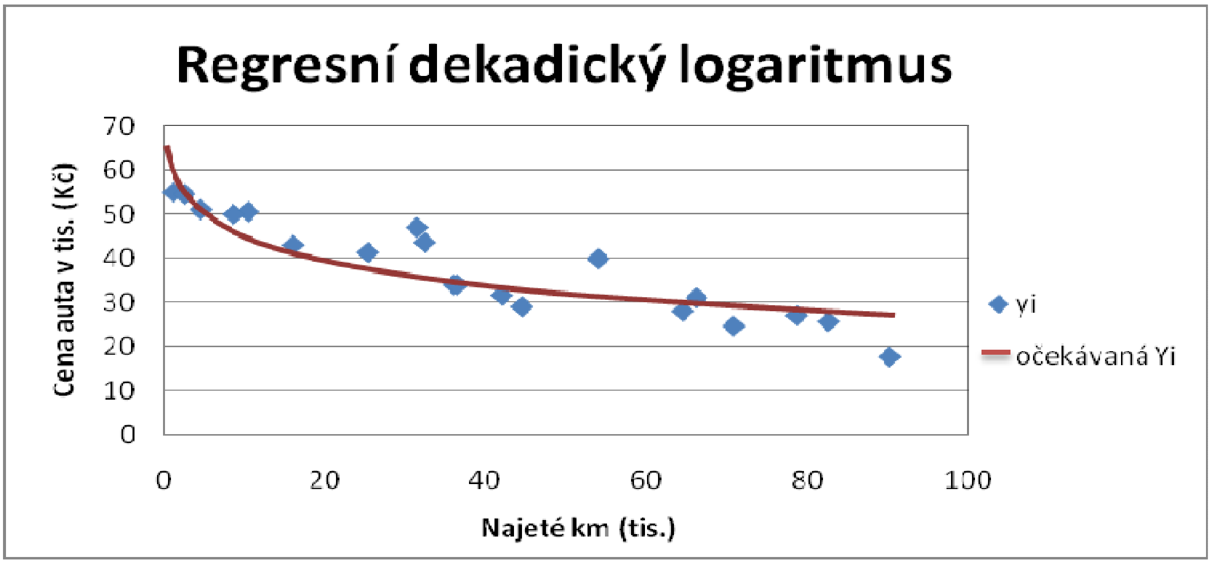
Regresní statistika	
Násobné R	**0,429022**
Hodnota spolehlivosti R	**0,18406**
Nastavená hodnota spolehlivosti R	0,13873
Chyba střední hodnoty	**10,35207**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	435,14	435,14	4,060452	**0,059083429**
Rezidua	18	**1928,977**	107,1654		
Celkem	19	2364,118			

7.7 Analýza regresního dekadického logaritmu

Z rozdílů v tabulce (Příloha 8) mezi *očekávanými Y* a vstupními *y* vidíme, že model regresního dekadického logaritmu prokládá data o poznání lépe než regresní hyperbola nebo exponenciála. Zdali jsou ale data proložena lépe než v případě regresního polynomu nebo regresního modelu přímky přímo z tabulky zjistit nemůžeme. Proto se zaměříme na grafický výstup a parametry regresní analýzy.



Obrázek 8: Regresní log

V tomto případě nám opět pro konečné posouzení optimálního regresního modelu pomůže analýza dat, protože samotný graf nám data prokládá velmi dobře, stejně jako v případech regresního modelu paraboly, přímky a polynomu 3. stupně. I zde se nám potvrzuje, že spoléhat se na pouhé vizuální hodnocení z obrázku grafu je značně rizikové. Dle mého názoru zde totiž není velký rozdíl mezi proložením dat logaritmickou křivkou a zatím nejlepšího modelu regresního polynomu 3. stupně. Porovnáním obou modelů ale rychle zjistíme, že ve všech parametrech více vyhovuje polynom.

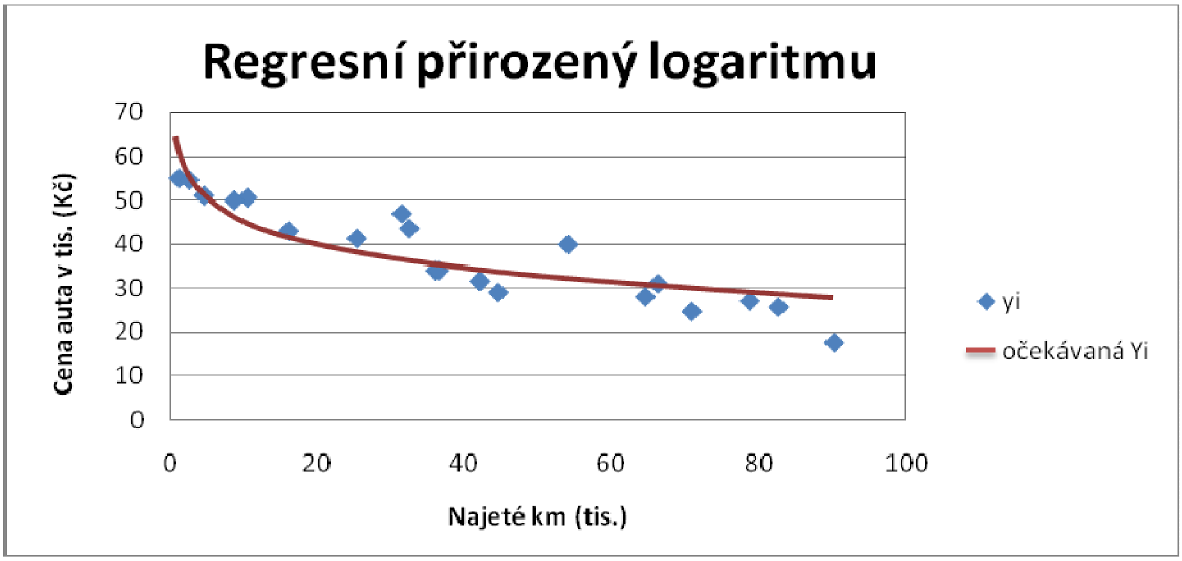
VÝSLEDEK

Regresní statistika	
Násobné R	**0,87890062**
Hodnota spolehlivosti R	**0,77246631**
Nastavená hodnota spolehlivosti R	0,75982555
Chyba střední hodnoty	**5,46664828**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1826,201	1826,201	61,10916	**3,40055E-07**
Rezidua	18	**537,9164**	29,88424		
Celkem	19	2364,118			

7.8 Analýza regresního přirozeného logaritmu



Obrázek 9: Regresní ln

Jak vidíme z obrázku, grafy dekadického a přirozeného logaritmus jsou téměř totožné, regresní statistiky jsou si také velmi podobné. Proto je zbytečné analýzu blíže popisovat.

VÝSLEDEK

Regresní statistika	
Násobné R	**0,878900625**
Hodnota spolehlivosti R	**0,772466309**
Nastavená hodnota spolehlivosti R	0,759825548
Chyba stř. hodnoty	**5,466648281**
Pozorování	20

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	1826,201	1826,201	61,11	**3,40055E-07**
Rezidua	18	**537,9164**	29,88424		
Celkem	19	2364,118			

7.9 Hodnotící tabulka

Při ukázkách dříve znázorněných regresních modelů bylo mnohokrát upozorněno, že některé regresní modely můžeme již pouhým vizuálním zhlédnutím předem odmítnout jako funkce nevhodně prokládající data. Ovšem u jiných modelů, jako například u regresního polynomu nebo regresní paraboly (v tomto příkladě) situace již zdaleka není tak jasná, jak by se nám zamlouvalo. Nejlepším řešením je proto sestrojit hodnotící tabulku, která bude porovnávat parametry kvality všech použitých regresních funkcí, ačkoliv jsme tyto výsledky již částečné porovnávali. Tabulka je vytvořená pro maximální přehlednost srovnávací analýzy parametrů jednotlivých regresních modelů. Bude obsahovat index determinace I^2, index korelace I a reziduální směrodatnou odchylku s z regresní statistiky. Dále významnost F a rezidua RSC z ANOVY. Jak vidíme, tabulka dohromady obsahuje pět kritérií, což je dle názoru odborníků dostatečný počet pro správný výběr optimálního modelu.

Tabulka 3: Hodnotící tabulka

Regresní model / Vlastnosti	**I**	**I^2**	**RSC**	**F**	**s**
PŘÍMKA	0,868812	0,932101	310,1441	2,27E-09	4,15093
PARABOLA	0,87999	0,938078	283,7188	1,49E-08	4,08526
POLYNOM 3. STUPNĚ	**0,88931**	**0,943032**	**261,6846**	**7,14E-08**	**4,04417**
HYPERBOLA	0,368289	0,606868	1493,439	0,00455	9,10872
EXPONENCIÁLA	0,18406	0,42902	1928,977	0,059083	10,3521
LOG	0,772466	0,878901	537,9164	3,40E-07	5,46665
LN	0,772466	0,878901	537,9164	3,40E-07	5,46665

7.10 Výběr optimálního regresního modelu

Zopakujme, že optimální regresní model by měl splňovat maximální index determinace, což spolu s minimálním RSC a vyhovující hladinou významnosti F jsou asi nejpodstatnější parametry. Dále je vhodná minimální směrodatná odchylku s a vysoký index korelace. Z hodnotící tabulky můžeme vyčíst, že v téměř všech parametrech je nejlepším modelem polynom 3. stupně. Má ze všech modelů nejvyšší index determinace, nejnižší rezidua RSC a současně je i v dalších parametrech nejlepším modelem. Pro tento typ regresní funkce rovněž svědčí i fakt optimálního proložení vstupními daty touto regresní funkcí. Z celkového výsledku analýzy jsme se tedy dozvěděli, že závislost ceny automobilu na počtu najetých kilometrů je závislostí polynomickou. Druhým modelem, který dobře vystihuje závislost dat je parabola. Naopak i bez vizuálního hodnocení jednotlivých grafů můžeme pomocí orientace v tabulce ihned vyloučit exponenciální regresní funkci, která v téměř všech parametrech nevyhovuje a navíc je jediným modelem statisticky nevýznamným v tomto příkladě.

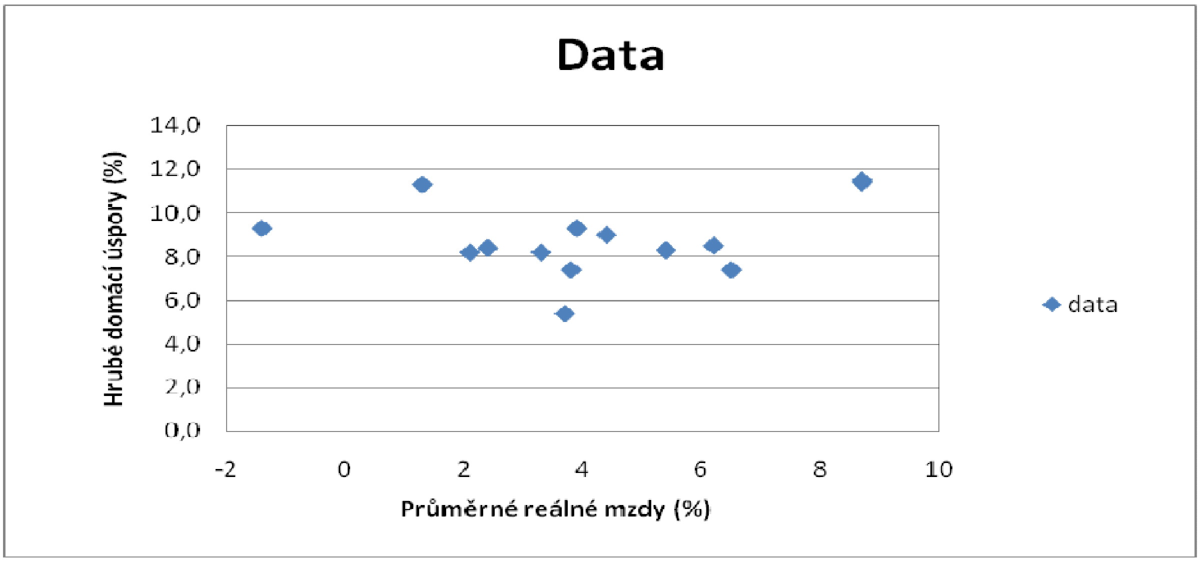
8 Příklad 2

8.1 Zadání

Tentokrát byl vybrán příklad z ekonomické praxe, kdy jsem zjišťoval závislost míry hrubých domácích úspor na měsíční hrubé reálné mzdě. Opět je mým cílem zjistit pomocí proložení regresních funkcí daty ideální závislost mezi zmíněnou hrubou mzdou a úsporami domácností. Data jsem opatřil z ekonomického serveru *Měšec.cz* dostupná online z www http://www.mesec.cz/dane/ekonomika/pruvodce/mzdy/ [11],[12]. Růsty či poklesy obou skupin dat jsou vyjádřeny v procentech a vztahují se k určitému období. Nezávislými daty *x* jsou hrubé mzdy a na nich závislými daty *y* jsou úspory domácností.

Tabulka 4: Vstupní data

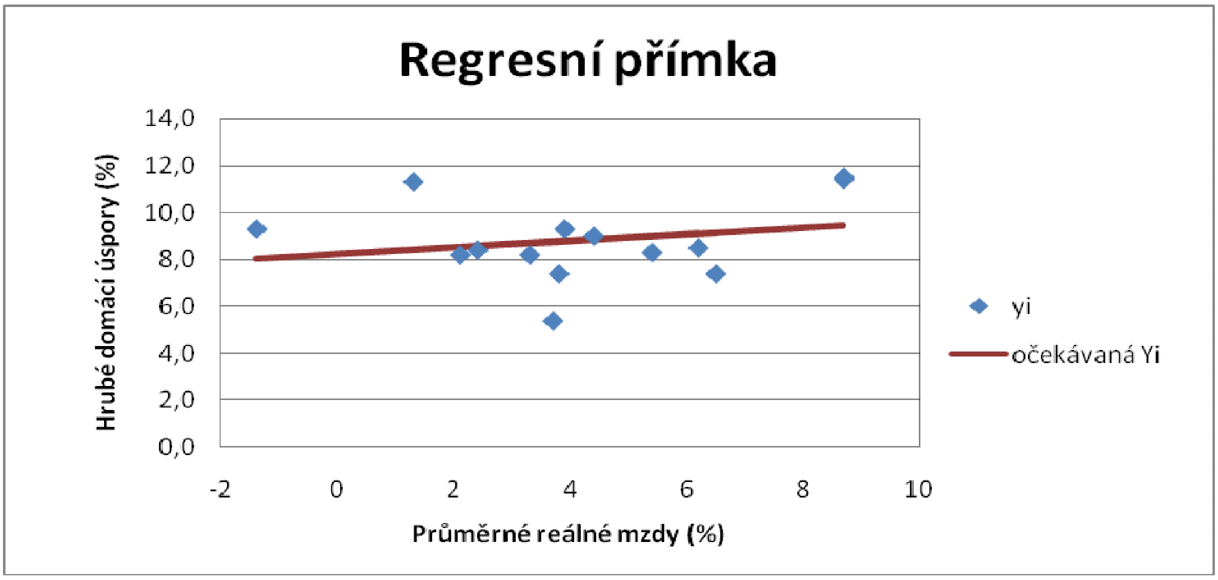
rok	průměrné mzdy *x*	úspory *y*
1995	8,7	11,4
1996	8,7	11,5
1997	1,3	11,3
1998	-1,4	9,3
1999	6,2	8,5
2000	2,4	8,4
2001	3,8	7,4
2002	5,4	8,3
2003	6,5	7,4
2004	3,7	5,4
2005	3,3	8,2
2006	3,9	9,3
2007	4,4	9,0
2008	2,1	8,2



Obrázek 10: Vstupní data

8.2 Analýza regresní přímky

V analýze se bude postupovat obdobným způsobem jako v předchozím příkladě. K posouzení závislosti budeme používat předběžné vizuální ukázky grafů a opět se hlavně zaměřím na ukazatele regresní analýzy. V *regresní statistice* opět použijeme hodnotu spolehlivosti R (index korelace), násobné R (index determinace) a chybu střední hodnoty (residuální směrodatná odchylka s). V ukazateli ANOVA se zaměříme na tučně zvýrazněná rezidua a významnost F. Tabulky potřebné k vytvoření regresní analýzy jsou opět uvedeny v příloze stejně jako v příkladě kapitoly sedm.



Obrázek 11: Regresní přímka

Z grafu můžeme vidět, že regresní funkce neprokládá data příliš dobře a neprokládá vůbec odlehlé hodnoty. Protože se jedná o první regresní model, nemůžeme porovnávat hodnoty parametrů regresní analýzy. Model je ale stejně statisticky nevýznamný, protože parametr významnost F je větší než 0,05.

VÝSLEDEK

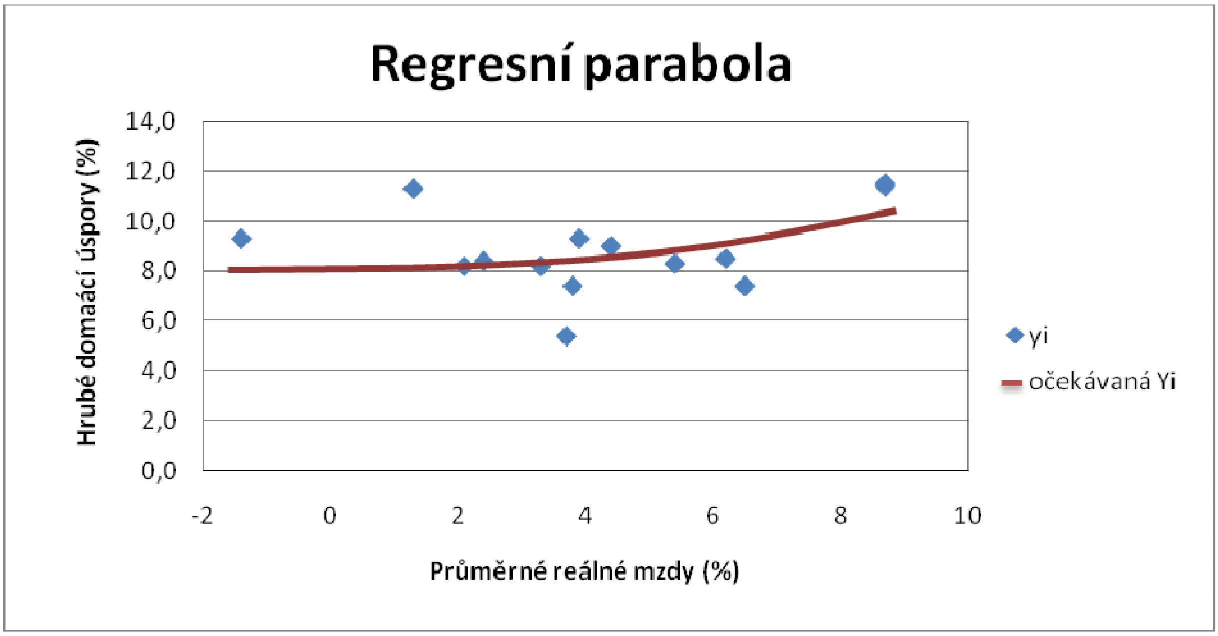
Regresní statistika	
Násobné R	**0,22925818**
Hodnota spolehlivosti R	**0,05255931**
Nastavená hodnota spolehlivosti R	-0,02639408
Chyba střední hodnoty	**1,721339**
Pozorování	14

ANOVA

	Rozdíl	SS	MS	F	Významnost F
Regrese	1	1,972475969	1,972476	0,665701	**0,430453**
Rezidua	12	**35,55609546**	2,963008		
Celkem	13	37,52857143			

8.3 Analýza regresní paraboly

V příloze je uvedena tabulka se vstupními daty (Příloha 11). Opakují se závislá a nezávislá vstupní data navíc doplněná o druhou mocninu *x*. Tato data zpracovávám v programu MS Excel. Jednotlivá *x* jsou nezávislá vstupní data, zatímco jednotlivá *y* jsou závislé veličiny. *Očekávaná Y* jsou odhadované veličiny, které odhadla regresní analýza použitá v MS Excel.



Obrázek 12: Regresní parabola

Z grafu můžeme posoudit, že regresní parabola prokládá data lépe než regresní přímka. Počáteční fáze regresní paraboly je téměř totožná s regresní přímkou, avšak v dalším průběhu vidíme, že model má alespoň „snahu“ proložit okolí odlehlé hodnoty vpravo. I v tomto případě nám více řekne hodnotící tabulka (Tabulka 5), ve které porovnáme jednotlivé parametry paraboly ve srovnání s ostatními regresními funkcemi. O vyloučení některých odlehlých hodnot nemůže být ani uvažováno, protože zjištěných údajů je příliš málo.

Níže je zobrazen výsledek regresní analýzy. Z analýzy můžeme vyčíst, že parabola má lepší parametry než přímka. Například parametry index determinace nebo index korelace mají vyšší hodnoty než přímka a naopak třeba parametr RSC má hodnotu nižší. Přesto je také tento model **statisticky nevýznamný,** ačkoliv je významnost F nižší než v případě regresní přímky.

VÝSLEDEK

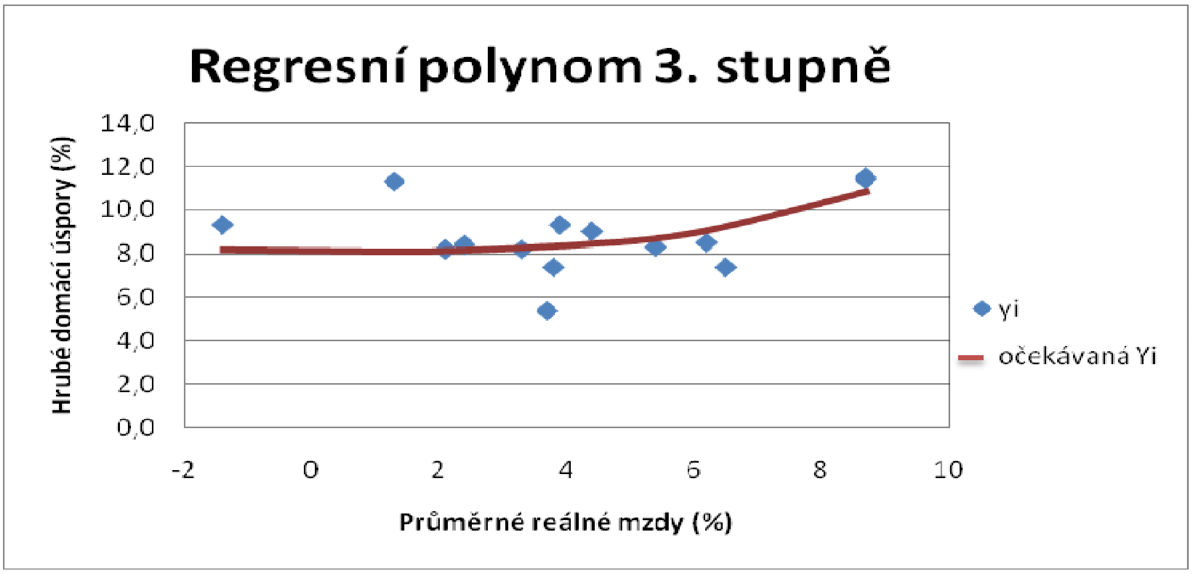
Regresní statistika	
Násobné R	**0,4440804**
Hodnota spolehlivosti R	**0,1972074**
Nastavená hodnota spolehlivosti R	0,130308
Chyba střední hodnoty	**1,5844994**
Pozorování	14

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	7,401	7,4009	2,948	**0,111667982**
Rezidua	12	**30,13**	2,5106		
Celkem	13	37,53			

8.4 Analýza regresního polynomu 3. stupně

Tabulka v příloze zobrazuje vstupní závislé a nezávislé veličiny (Příloha 12). V případě polynomu 3. stupně navíc zahrnuje třetí mocninu nezávislých vstupních hodnot *x*.



Obrázek 13: Regresní polynom

Z grafu je patrné, že regresní polynom 3. stupně je dobrou funkcí pro vystižení závislosti mezi daty, protože dobře prokládá zjištěné hodnoty. Můžeme jej tedy již v této fázi označit jako vhodný z vizuálního hlediska. I z výsledků parametrů je vidět, že proložení daty je celkem dobré.

Tento model je prvním modelem, který je statisticky významný. Z toho vyplývá, že index determinace a index korelace je větší, než v případě paraboly a naopak chyba střední hodnoty a rezidua mají hodnoty nižší. Toto vše ukazuje na vyšší kvalitu regresního modelu než v předchozích případech.

VÝSLEDEK

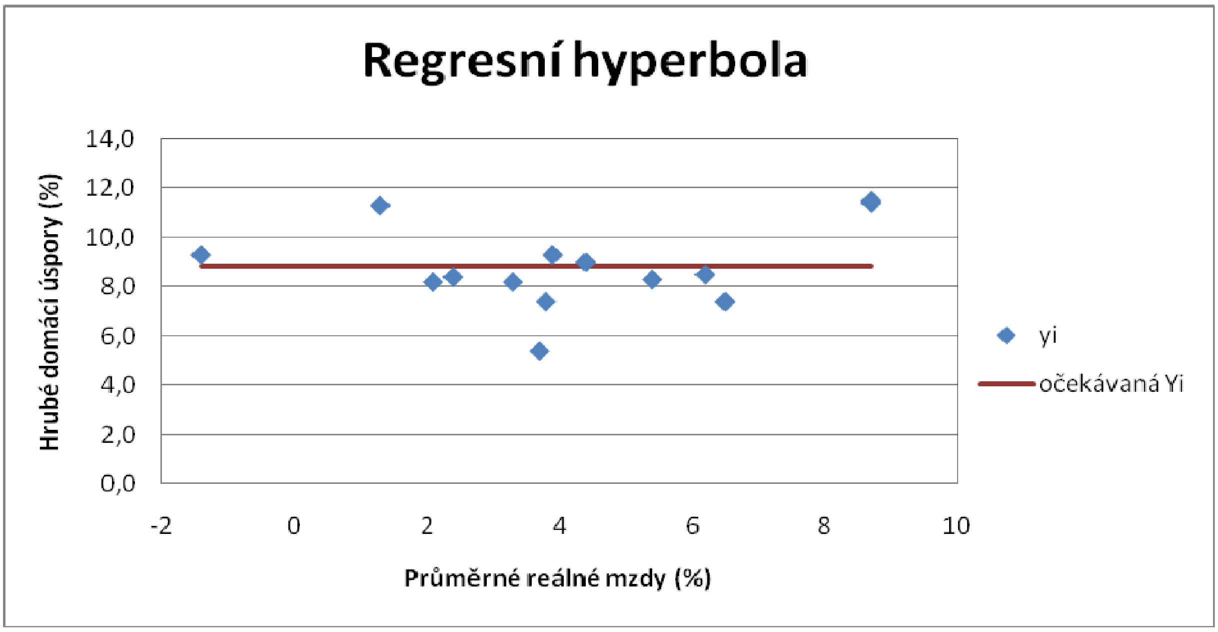
Regresní statistika	
Násobné R	**0,53196969**
Hodnota spolehlivosti R	**0,28299175**
Nastavená hodnota spolehlivosti R	0,22324106
Chyba střední hodnoty	**1,49745048**
Pozorování	14

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	10,62028	10,62028	4,736209107	**0,050226943**
Rezidua	12	**26,9083**	2,242358		
Celkem	13	37,52857			

8.5 Analýza regresní hyperboly

Obrázek 14 regresní hyperboly nám také jednoznačně neurčí, zda lépe prokládá data regresní hyperbola nebo třeba přímka. Z analýzy ale vidíme, že parametry regresní hyperboly nejsou příliš dobré. Hyperbola má dost nízký index determinace a velmi nízký index korelace v porovnání s předchozími modely. Naopak rezidua a třeba významnost F mají velké hodnoty. Model je statisticky nevýznamný a zatím je svými parametry nejhorším modelem vystihující závislost mezi daty. Tabulka s daty k tomuto modelu je označena jako Příloha 13.



Obrázek 14: Regresní hyperbola

VÝSLEDEK

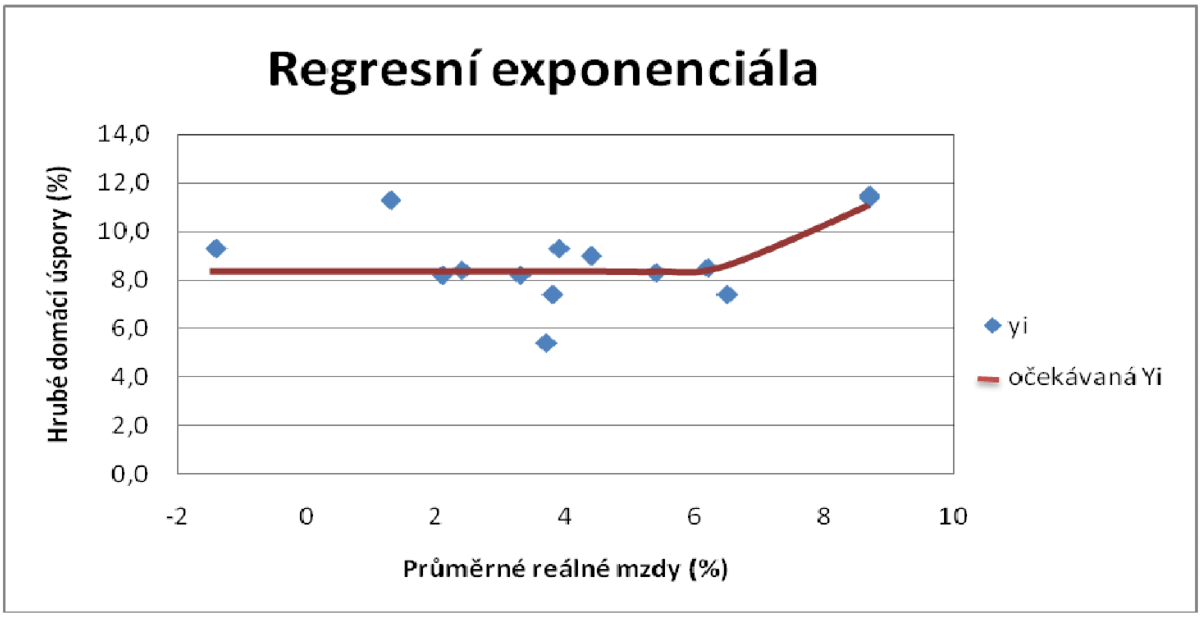
Regresní statistika	
Násobné R	**0,002097621**
Hodnota spolehlivosti R	**4,40002E-06**
Nastavená hodnota spolehlivosti R	-0,083328567
Chyba střední hodnoty	**1,768436369**
Pozorování	14

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	0,000165	0,0001651	5,3E-05	**0,994321715**
Rezidua	12	**37,52841**	3,1273672		
Celkem	13	37,52857			

8.6 Analýza regresní exponenciály

Z grafu můžeme vidět, že regresní exponenciála dobře prokládá vstupní data a možná bude nejvhodnější funkcí pro vystižení závislosti. Nejen že dobře prokládá oblast, kde je nejvíce naměřených hodnot vstupních dat, ale navíc prokládá odlehlou hodnotu v pravé části grafu, což u předchozích regresních modelů neplatilo.



Obrázek 15: Regresní exponenciála

Vše se objasní závěrečným porovnáním parametrů všech modelů. Model regresní exponenciály má ale prozatím nejlepší parametry i ve srovnání s polynomem nebo parabolou, které prokládaly data celkem dobře, ačkoliv model paraboly nebyl statisticky významný. Vidíme, že indexy korelace a determinace mají prozatím největší hodnoty. Naopak rezidua nebo významnost F jsou výrazně nižší než v předchozích případech. Významnost F je menší než 0,05 a proto je model **statisticky významný**. Zatím se jedná o nejlepší model vystihující závislost mezi daty nejen svými parametry, ale i z vizuálního hlediska.

VÝSLEDEK

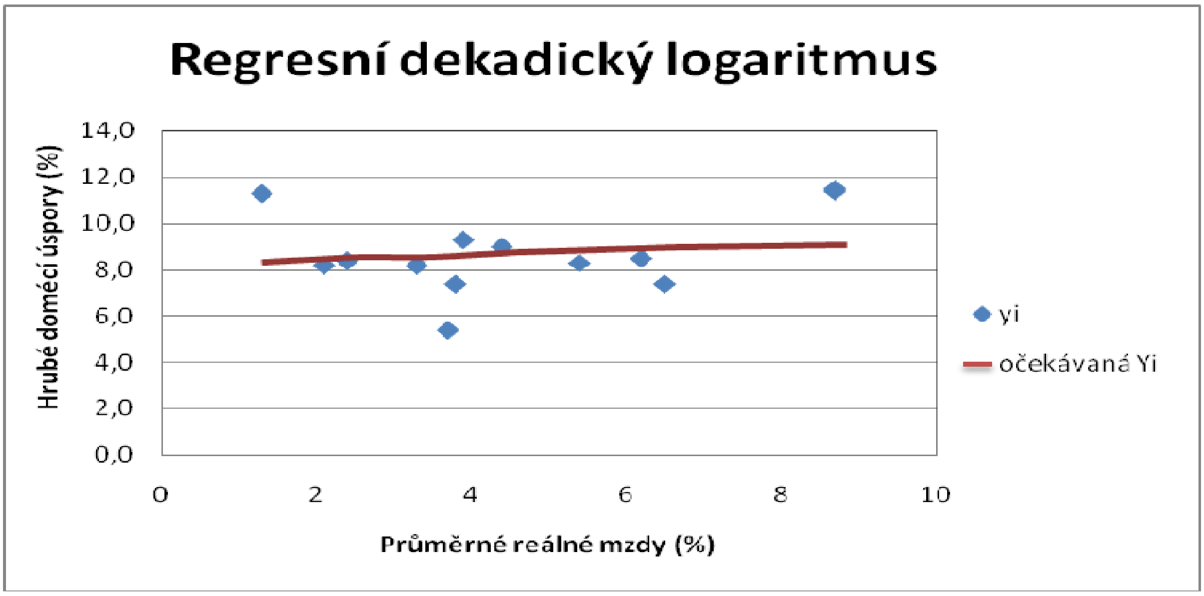
Regresní statistika	
Násobné R	**0,636106828**
Hodnota spolehlivosti R	**0,404631897**
Nastavená hodnota spolehlivosti R	0,355017889
Chyba střední hodnoty	**1,364530272**
Pozorování	14

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	15,18526	15,185257	8,1556	**0,014466167**
Rezidua	12	**22,34331**	1,8619429		
Celkem	13	37,52857			

8.7 Analýza regresního dekadického logaritmu

Proložení regresní funkce daty (Obrázek 16) je podobné jako v případě přímky. Rovněž parametry jsou podobné, což vysvětluje podobný tvar regresní funkce. Níže je uveden opět výstup z regresní analýzy. Jak vidíme, parametry jsou horší ve srovnání s exponenciálou nebo polynomem. Model je statisticky nevýznamný. I tak ale zahrneme tento model do konečného hodnocení v tabulce. Ze závěrečné hodnotící tabulky se můžeme vzájemným porovnáním dozvědět důležité věci.



Obrázek 16: Regresní log

Z hodnotící tabulky poznáme nejen který model je nejoptimálnější pro odhad vývoje dat, ale i které modely je třeba zavrhnout jako vyloženě nehodící se pro budoucí predikci dat. Nebo můžeme odhadnout, které modely by se naopak daly použít, kdybychom dostali větší počet dat z analyzované oblasti (hrubé domací úspory závislé na reálných mzdách) a kdybychom v návaznosti na tuto skutečnost mohli vybočující hodnoty odstranit. Potom by třeba exponenciála nebyla příliš vhodná, ale spíše by se více hodila přímková regrese nebo model logaritmu.

VÝSLEDEK

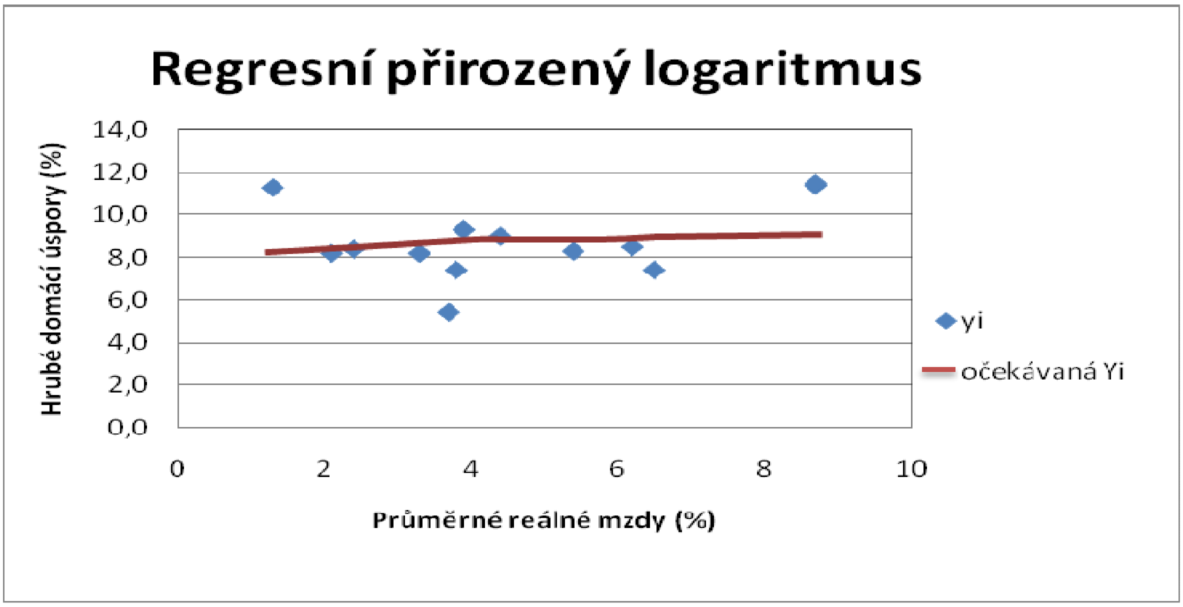
Regresní statistika	
Násobné R	**0,118832891**
Hodnota spolehlivosti R	**0,014121256**
Nastavená hodnota spolehlivosti R	-0,07550408
Chyba střední hodnoty	**1,828130192**
Pozorování	13

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	0,526571	0,52657	0,15756	**0,699007022**
Rezidua	11	**36,76266**	3,34206		
Celkem	12	37,28923			

8.8 Analýza regresního přirozeného logaritmu

Průběh regresní funkce přirozeného logaritmu se od logaritmické funkce příliš neliší, proto jej nebudu blíže popisovat. Pouze zbývá dodat, že u přirozeného i dekadického logaritmu jsem vyřadil hodnotu – 1,4 v nezávislých datech *x*, protože funkce obou logaritmů je definována pouze pro kladné hodnoty.



Obrázek 17: Regresní ln

VÝSLEDEK

Regresní statistika	
Násobné R	**0,117833**
Hodnota spolehlivosti R	**0,014121**
Nastavená hodnota spolehlivosti R	-0,075504
Chyba střední hodnoty	**1,82813**
Pozorování	13

ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	0,526570777	0,5266	0,1576	**0,699007022**
Rezidua	11	**36,76565999**	3,3421		
Celkem	12	37,28923077			

8.9 Hodnotící tabulka

Jak vidíme, tabulka opět obsahuje index determinace I^2, index korelace I a reziduální směrodatnou odchylku s z regresní statistiky. Dále významnost F a rezidua RSC z ANOVY. Dohromady obsahuje pět kritérií, které jsou postačující pro posouzení modelu.

Tabulka 5: Hodnotící tabulka

Regresní model / Vlastnosti	**I**	**I^2**	**RSC**	**F**	**s**
PŘÍMKA	0,052559	0,229258	35,5561	0,430453	1,721339
PARABOLA	0,197207	0,44408	30,12766	0,111668	1,584499
POLYNOM 3. STUPNĚ	0,282992	0,53197	26,9083	0,050227	1,49745
HYPERBOLA	4,40E-06	0,002098	37,52841	0,994322	1,768436
EXPONENCIÁLA	**0,404632**	**0,636107**	**22,34331**	**0,014466**	**1,36453**
LOG	0,014121	0,118833	36,76266	0,699007	1,82813
LN	0,014121	0,117833	36,76566	0,699007	1,82813

8.10 Výběr optimálního regresního modelu

Nejprve bych chtěl říci, že kdyby nám šlo o rychlou analýzu, zaměřil bych se v popisech modelů nejprve na parametr významnost F, který je velmi důležitý z hlediska použitelnosti modelů. Zjistil bych, že statisticky významné jsou pouze regresní modely polynomu 3. stupně a exponenciály. Poté bych porovnával pouze parametry těchto modelů a ostatní regresní modely bych vůbec neuvažoval. Protože však statisticky významné modely vyšly v tomto příkladě pouze dva, dovolil jsem si analyzovat i ostatní regresní modely. Máme k dispozici alespoň názornou ukázku, že modely statisticky nevýznamné mají parametry skutečně výrazně horší.

Vzájemným porovnáním parametrů (Tabulka 5) nám vychází, že nejlepším regresním modelem pro vystižení závislosti mezi daty, je závislost exponenciální. Je nejlepším modelem, protože hodnoty index korelace i index determinace jsou ze všech uvedených modelů nejvyšší a naopak chybu střední hodnoty má exponenciální model nejnižší. V ukazatelích ANOVA jsme se zaměřili na tučně zvýrazněná rezidua RSC a významnost F. Tyto hodnoty jsou také nejnižší, což vhodnost exponenciálního modelu jen potvrzuje. Mimoto, jak již bylo napsáno, je exponenciální model ještě s polynomem 3. stupně jediným modelem, který je statistický významný.

Výrazně nejhorším modelem pro vystižení závislosti je hyperbolický model, který nevyhovuje ve srovnání s ostatními modely ve všech parametrech. Celkem slušné parametry pro vystižení závislosti mezi daty má také polynom 3. stupně a regresní parabola, ačkoliv je statisticky nevýznamná. Z výsledku tabulky potom můžeme usoudit, že závislost procentního růstu úspor domácností *y* na velikosti růstu procent hrubých mezd *x* je závislostí exponenciální. Je třeba říci, že to není ani příliš překvapivé zjištění, protože si jistě dokážeme představit exponenciální růst (či pokles) rodinných úspor (např. na bydlení, spoření, dovolenou a ostatní výdaje) v závislosti na výši mezd.

Závěr

Na závěr se dá říci, že záměr bakalářské práce byl splněn. Cílem zadání práce bylo teoreticky popsat metody a pojmy regresní analýzy a následně vytvořit a vyhodnotit jednotlivé regresní modely u zpracovaných příkladů. V teoretické části jsem mimo jiné charakterizoval jednotlivé regresní modely použité v praktické části práce a jednotlivé ukazatelé kvality regresní funkce, kterých se pro posouzení kvality modelu následně využívalo. V praktické reprezentaci metod regresní analýzy pro všechny tyto modely vytvořila analýza v programu MS Excel určité parametry. Z těchto parametrů jsem se zaměřil na index determinace I^2, index korelace I, chybu střední hodnoty s, dále významnost F a rezidua RSC. Následně jsem pomocí očekávaných hodnot *Y*, které jsou výstupem pro každý model vždy vytvořil příslušnou regresní funkci prokládající vstupní data. Nejprve se u většiny modelů provedlo vizuální zhodnocení proložení regresní funkce daty, avšak hlavní důraz pro určení kvality regresního modelu se kladl na porovnání jednotlivých parametrů. Porovnáním se určila vhodnost regresních modelů a dle toho byl vybrán ten model, který nejlépe prokládá zvolená data. Výběr nejlepšího modelu byl vždy zdůvodněn a zároveň se poukázalo na regresní modely, které jsou nevhodné pro popis závislosti mezi daty. Příklady byly zpracovány dva proto, aby bylo možno porovnat výstupy dvou odlišných případů. Z výsledků vidíme, že výstupy mohou být pro odlišná data zcela jiná. Zatímco v prvním příkladě vyšel pouze jeden statisticky nevýznamný model, v druhém případě je nevýznamných modelů pět.

U příkladu v sedmé kapitole byla analyzována data vztahující se k závislosti mezi stářím a cenou automobilů. Naměřených hodnot je dvacet a bylo zjištěno, že tato data jsou proložena nejlépe polynomem 3. stupně. U tohoto příkladu se příliš nedalo spoléhat na subjektivní vizuální hodnocení, protože i další modely data dobře prokládaly, ačkoliv měli horší parametry. Výsledky analýzy jsem tedy odvodil hlavně pomocí hodnotící tabulky (Tabulka 3), kde jsou vypsány hodnoty nejdůležitějších parametrů pro určení kvality modelů. Naopak se dá říci, že model, který absolutně nevyhovuje jak z hlediska vizuálního, tak z hlediska parametrického je model regresní exponenciály. Regresní analýza nám dala predikci, že i pro budoucnost můžeme očekávat polynomickou závislost mezi cenou automobilu a jeho najetými kilometry.

V druhém příkladě v kapitole osmé byla analyzována závislost mezi výší hrubých reálných mezd a úsporami domácnosti, přičemž bylo k dispozici 14 naměřených hodnot. Bylo zjištěno, že data jsou nejlépe proložena exponenciální funkcí. Exponenciální funkce byla již z pouhého vizuálního hodnocení nejlepším regresním modelem, což bylo potvrzeno porovnáním parametrů s ostatními modely. Dalším regresním modelem dobře popisujícím závislost mezi daty byl model polynomu 3. stupně. Nejhorším modelem z hlediska parametrů, je regresní hyperbola, ačkoliv v tomto příkladě bylo více statisticky nevýznamných modelů. Na základě této analýzy můžeme predikovat, že i v budoucnosti se závislost mezi mzdou a úsporami bude řídit exponenciální funkční závislostí.

Je třeba zmínit, že při modelování dat v obou příkladech byl použit pouze polynom 3. stupně. Obvykle se sice jeví regresní polynomy vyšších stupňů jako vhodnější, protože lépe aproximují data, ale na druhou stranu čím vyšší je stupeň regresního polynomu, tím obtížnější se s ním pracuje. Vzrůstá i riziko, že se regresní funkce bude zbytečně snažit popsat náhodné výchylky od celkového trendu a důsledkem bude, že regresní funkce bude pro předpověď naprosto nevhodná.

Použitá literatura

[1] CYHELSKÝ, L. *Úvod do teorie statistiky*. vydání 2. Praha : SNTL, 1981. 352 s. ISBN L31-C3-3-41/38253.

[2] CYHELSKÝ, L., NOVÁK, I. *Statisitka 1. díl*. vydání 1. Praha: SNTL, 1967. 287 s. ISBN L31-C3-4-41/3740/1.

[3] CYHELSKÝ, L. a kolektiv. *Základy teorie statistiky pro ekonomy*. vydání 1. Praha: SNTL, 1979. 365 s. ISBN L31-C3-4-41/38141.

[4] HENDL, J. *Přehled statistických metod – zpracování dat.* vydání 2. Praha: PORTÁL, 2006. 583 s. ISBN 80-7367-123-9.

[5] HINDLS, R., HRONOVÁ, S., SEGER, S. *Statistika pro ekonomy*. vydání 3. Praha: PROFESSIONAL PUBLISHING, 2003. 415 s. ISBN 80-86419-34-7 .

[6] HINDLS, R., KAŇKOVÁ, R., NOVÁK, I. *Metody statistické analýzy pro ekonomy*. vydání 1. Praha: MANAGEMENT PRESS, 1997. 250 s. ISBN 80-85943-44-1 .

[7] KUBANOVÁ, J. *Statistické metody pro ekonomickou a technickou praxi*. vydání 3. Bratislava: STATIS, 2008. 247 s. ISBN 978-80-85659-474.

[8] MAREK, L. *Statistika pro ekonomy – aplikace.* vydání 1. Praha: PROFESSIONAL PUBLISHING, 2005. 423 s. ISBN 80-86419-68-1.

[9] MELOUN, M. a MILITKÝ, J. *Statistické zpracování experimentálních dat*. vydání 1. Praha: EAST PUBLISHING, 1998. 824 s. ISBN 80-7219-003-2.

[10] SLÍVKO, P. *Sbírka úloh ze statistiky*. vydání 2. Česká Třebová : Interní zdroj VOŠ, 2003. 64 s.

[11] *Měšec.cz : Vývoj reálných mezd v ČR, 1995-2008* [online]. 2008 [cit. 2010-02-10]. Dostupné z WWW: <http://www.mesec.cz/dane/ekonomika/pruvodce/mzdy/>.

[12] *Měšec.cz : Vývoj spotřeby v ČR, 1995-2008* [online]. 2008 [cit. 2009-02-11]. Dostupné z WWW: <http://www.mesec.cz/dane/ekonomika/pruvodce/spotreba/>

[13] Regresní analýza : Metoda nejmenších čtverců. In *Less8reg2*. Praha :fzp.ujep.cz, 14.12.2008 [cit.2010-02-07].Dostupné z WWW: <http://fzp.ujep.cz/~synek/statistika/prednasky/?C=D;O=AD;O=A>.

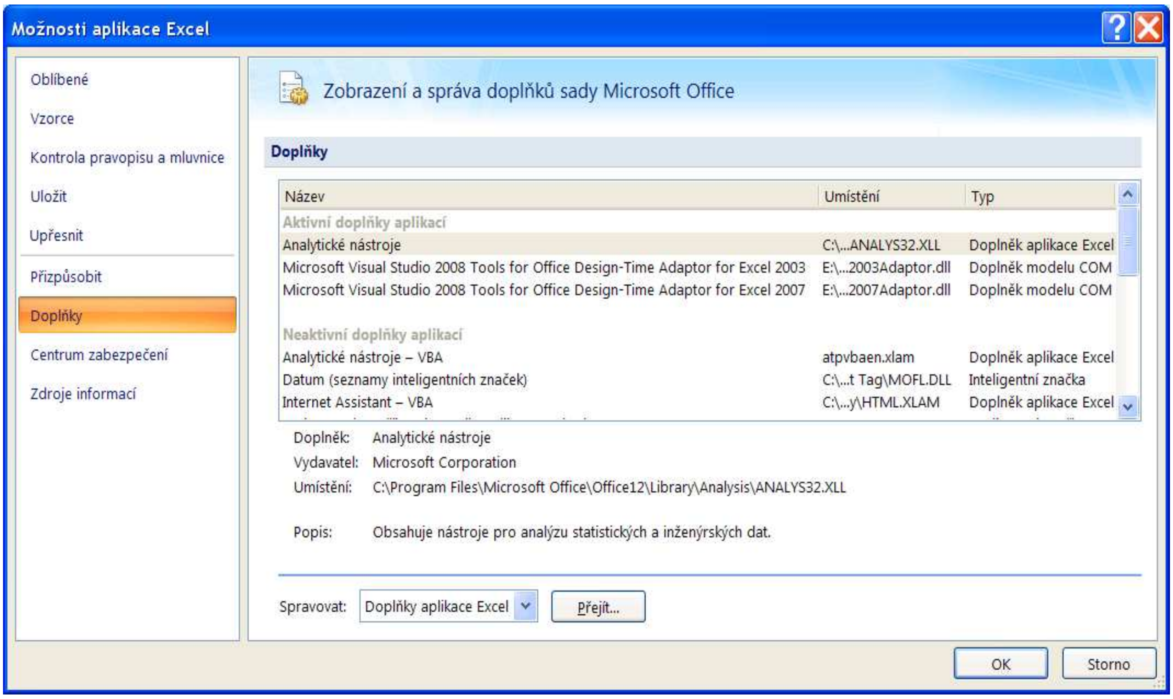
Seznam zkratek

X	Náhodná veličina
x	Hodnota statistického znaku
$\bar{x}$	Aritmetický průměr
$\hat{y}$	Odhad hodnoty statistického znaku
I^2	Index determinace
I	Index korelace
s	Chyba střední hodnoty
RSC	Residuální součet čtverců
ε	Náhodná chyba

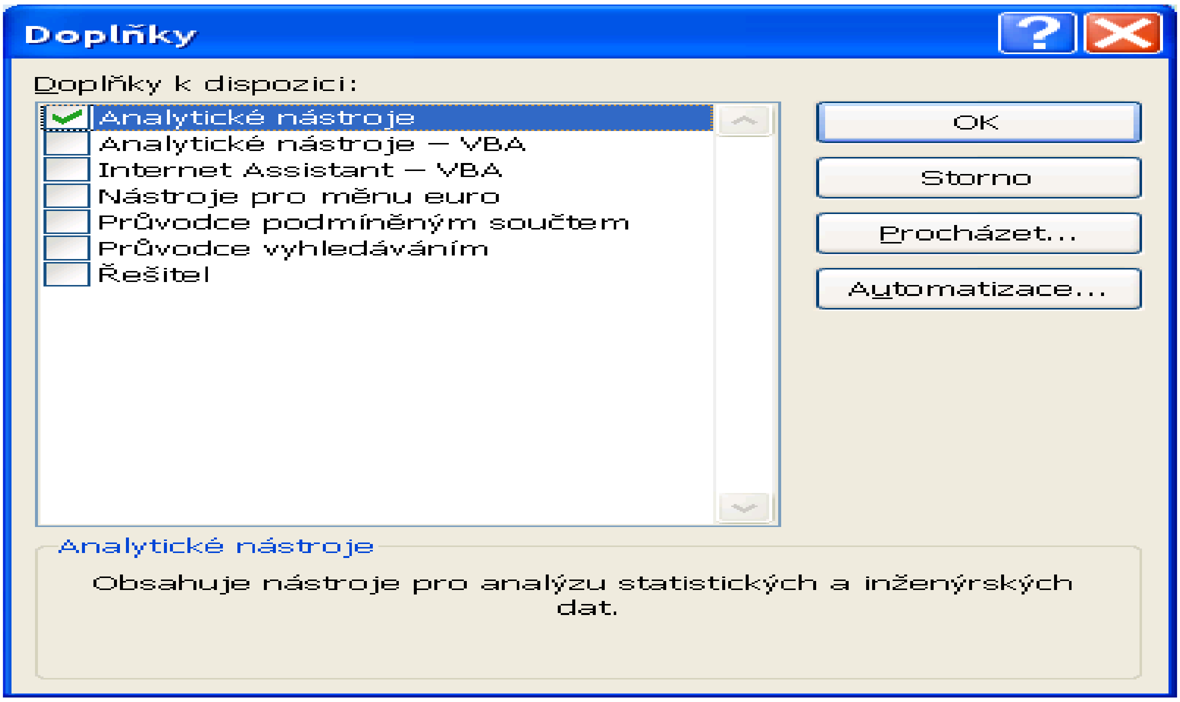
Seznam obrázků

Seznam tabulek

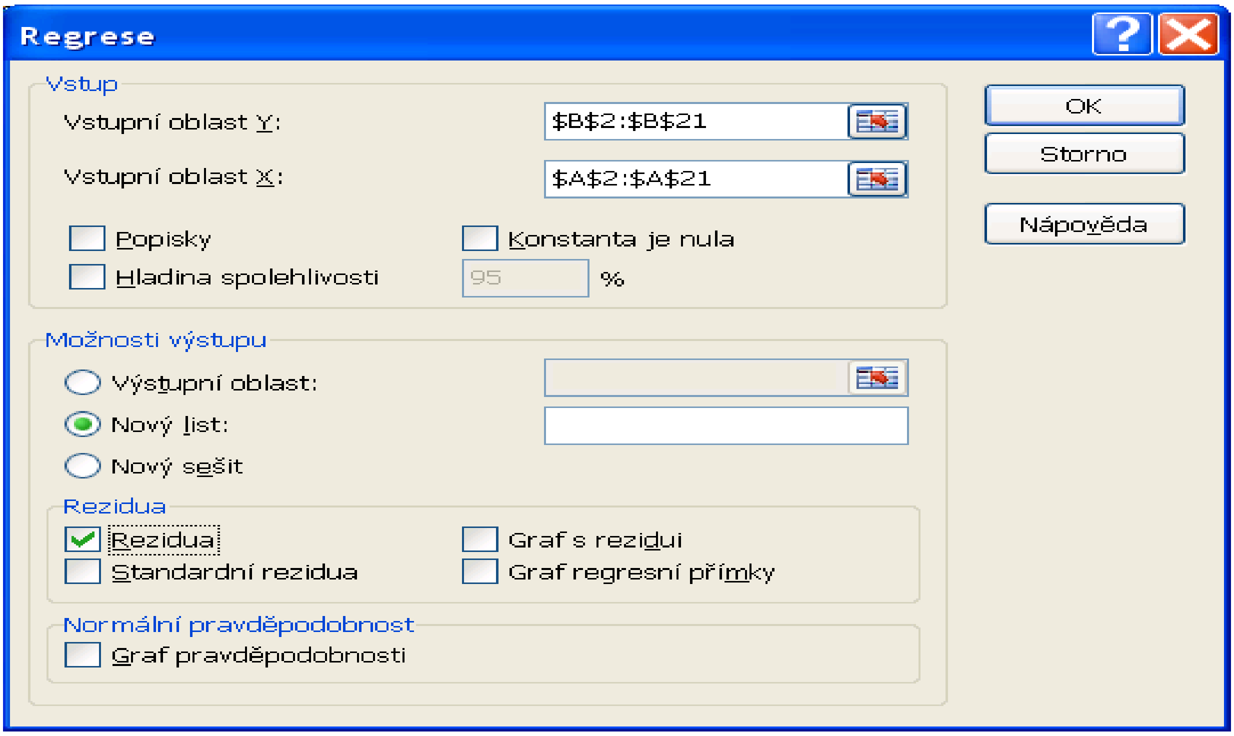
Seznam příloh



Příloha 1:Doplňky MS EXCEL



Příloha 2:Doplňky Analytické nástroje



Příloha 3:Tvorba regrese

Příloha 4: Vstupní data a očekávané hodnoty regresní paraboly

y	*x*	x^2	*Očekávaná Y*
55	1,1	1,21	53,82247464
54,6	2,5	6,25	53,11456238
50,6	10,4	108,16	49,24341337
51,1	4,5	20,25	52,11468955
47	31,4	985,96	39,97318052
50	8,6	73,96	50,10699059
43,6	32,4	1049,76	39,56872156
41,3	25,3	640,09	42,51318169
43	16	256	46,6263869
39,9	54	2916	31,65297648
34	36	1296	38,14050572
31	66,2	4382,44	27,87519932
29	44,5	1980,25	34,94121424
31,6	42	1764	35,8569682
34	36,4	1324,96	37,98450458
25,6	82,6	6822,76	23,5853046
28	64,5	4160,25	28,37160609
24,6	70,8	5012,64	26,58069467
27	78,7	6193,69	24,52351661
17,6	90,2	8136,04	21,9039083

Příloha 5: Vstupní data a očekávané hodnoty regresního polynomu

y	*x*	x^2	x^3	***Očekávaná Y***
55	1,1	1,21	1,331	55,19770375
54,6	2,5	6,25	15,625	54,11086475
50,6	10,4	108,16	1124,864	48,70128831
51,1	4,5	20,25	91,125	52,62797842
47	31,4	985,96	30959,14	39,01475281
50	8,6	73,96	636,056	49,83149238
43,6	32,4	1049,76	34012,22	38,67778191
41,3	25,3	640,09	16194,28	41,26999187
43	16	256	4096	45,53242352
39,9	54	2916	157464	32,67394298
34	36	1296	46656	37,52870221
31	66,2	4382,44	290117,5	29,35965971
29	44,5	1980,25	88121,13	35,11928959
31,6	42	1764	74088	35,79368381
34	36,4	1324,96	48228,54	37,40665426
25,6	82,6	6822,76	563560	23,35087601
28	64,5	4160,25	268336,1	29,85809251
24,6	70,8	5012,64	354894,9	27,91759929

Příloha 6: Vstupní data a očekávané hodnoty regresní hyperboly

y	*x*	*1/x*	***Očekávaná Y***
55	1,1	0,909091	63,68111014
54,6	2,5	0,4	47,3649453
50,6	10,4	0,096154	37,62679472
51,1	4,5	0,222222	41,66723694
47	31,4	0,031847	35,5657897
50	8,6	0,116279	38,27180028
43,6	32,4	0,030864	35,53428697
41,3	25,3	0,039526	35,81188448
43	16	0,0625	36,54820209
39,9	54	0,018519	35,13861278
34	36	0,027778	35,43536843
31	66,2	0,015106	35,02923457
29	44,5	0,022472	35,26531744
31,6	42	0,02381	35,30818744
34	36,4	0,027473	35,42558527
25,6	82,6	0,012107	34,9331113
28	64,5	0,015504	35,04199467
24,6	70,8	0,014124	34,9977796
27	78,7	0,012706	34,95233923
17,6	90,2	0,011086	34,90041867

Příloha 7: Vstupní data a očekávané hodnoty regresní exponenciály

y	***x***	***exp(x)***	**Očekávaná Y**
55	1,1	3,004166	38,99567094
54,6	2,5	12,182494	38,99567094
50,6	10,4	32859,626	38,99567094
51,1	4,5	90,017131	38,99567094
47	31,4	4,334E+13	38,99567094
50	8,6	5431,6596	38,99567094
43,6	32,4	1,178E+14	38,99567094
41,3	25,3	9,72E+10	38,99567094
43	16	8886110,5	38,99567094
39,9	54	2,831E+23	38,99567094
34	36	4,311E+15	38,99567094
31	66,2	5,627E+28	38,99567094
29	44,5	2,119E+19	38,99567094
31,6	42	1,739E+18	38,99567094
34	36,4	6,432E+15	38,99567094
25,6	82,6	7,46E+35	38,98496003
28	64,5	1,028E+28	38,99567094
24,6	70,8	5,598E+30	38,99567086
27	78,7	1,51E+34	38,99545413
17,6	90,2	1,491E+39	17,59317991

Příloha 8: Vstupní data a očekávané hodnoty regresního logaritmu

y	***x***	***log(x)***	**Očekávaná Y**
55	1,1	0,041393	62,77040937
54,6	2,5	0,39794	56,25698855
50,6	10,4	1,017033	44,94736672
51,1	4,5	0,653213	51,59366017
47	31,4	1,49693	36,18060167
50	8,6	0,934498	46,45511737
43,6	32,4	1,510545	35,93187569
41,3	25,3	1,403121	37,89431057
43	16	1,20412	41,5296604
39,9	54	1,732394	31,87913382
34	36	1,556303	35,09597601
31	66,2	1,820858	30,26306572
29	44,5	1,64836	33,41426572
31,6	42	1,623249	33,87298937
34	36,4	1,561101	35,00830983
25,6	82,6	1,91698	28,50710411
28	64,5	1,80956	30,46946306
24,6	70,8	1,850033	29,73009075
27	78,7	1,895975	28,89083004
17,6	90,2	1,955207	27,80878105

Příloha 9: Vstupní data a očekávané hodnoty regresního ln

y	***x***	***ln(x)***	***Očekávaná Y***
55	1,1	0,09531	62,77040937
54,6	2,5	0,916291	56,25698855
50,6	10,4	2,341806	44,94736672
51,1	4,5	1,504077	51,59366017
47	31,4	3,446808	36,18060167
50	8,6	2,151762	46,45511737
43,6	32,4	3,478158	35,93187569
41,3	25,3	3,230804	37,89431057
43	16	2,772589	41,5296604
39,9	54	3,988984	31,87913382
34	36	3,583519	35,09597601
31	66,2	4,19268	30,26306572
29	44,5	3,795489	33,41426572
31,6	42	3,73767	33,87298937
34	36,4	3,594569	35,00830983
25,6	82,6	4,41401	28,50710411
28	64,5	4,166665	30,46946306
24,6	70,8	4,259859	29,73009075
27	78,7	4,365643	28,89083004
17,6	90,2	4,502029	27,80878105

Příloha 10: Vstupní data a očekávané hodnoty regresní přímky

x	***y***	***Očekávaná Y***
8,7	11,4	9,458575959
8,7	11,5	9,458575959
1,3	11,3	8,419269122
-1,4	9,3	8,040062573
6,2	8,5	9,107458785
2,4	8,4	8,573760679
3,8	7,4	8,770386297
5,4	8,3	8,995101289
6,5	7,4	9,149592846
3,7	5,4	8,75634161
3,3	8,2	8,700162862
3,9	9,3	8,784430984
4,4	9,0	8,854654419
2,1	8,2	8,531626618

Příloha 11: Vstupní data a očekávané hodnoty regresní paraboly

x	***y***	***x^2***	***Očekávaná Y***
8,7	11,4	75,69	10,36669544
8,7	11,5	75,69	10,36669544
1,3	11,3	1,69	8,125553874
-1,4	9,3	1,96	8,133731013
6,2	8,5	38,44	9,238553233
2,4	8,4	5,76	8,24881666
3,8	7,4	14,44	8,511696509
5,4	8,3	29,16	8,957501967
6,5	7,4	42,25	9,353941738
3,7	5,4	13,69	8,488982236
3,3	8,2	10,89	8,404182285
3,9	9,3	15,21	8,535016496
4,4	9,0	19,36	8,660702138
2,1	8,2	4,41	8,20793097

Příloha 12: Vstupní data a očekávané hodnoty regresního polynomu

x	***y***	***x^2***	***x^3***	***Očekávaná Y***
8,7	11,4	75,69	658,503	10,80376202
8,7	11,5	75,69	658,503	10,80376202
1,3	11,3	1,69	2,197	8,182310624
-1,4	9,3	1,96	-2,744	8,162575026
6,2	8,5	38,44	238,328	9,125477286
2,4	8,4	5,76	13,824	8,228751789
3,8	7,4	14,44	54,872	8,392707835
5,4	8,3	29,16	157,464	8,802486112
6,5	7,4	42,25	274,625	9,270456642
3,7	5,4	13,69	50,653	8,375856087
3,3	8,2	10,89	35,937	8,317076678
3,9	9,3	15,21	59,319	8,410470273
4,4	9,0	19,36	85,184	8,513781595
2,1	8,2	4,41	9,261	8,210526018

Příloha 13: Vstupní data a očekávané hodnoty regresní hyperboly

x	***y***	***1/x***	***Očekávaná Y***
8,7	11,4	0,114942529	8,829680164
8,7	11,5	0,114942529	8,829680164
1,3	11,3	0,769230769	8,822368149
-1,4	9,3	-0,714285714	8,838947226
6,2	8,5	0,161290323	8,829162203
2,4	8,4	0,416666667	8,826308239
3,8	7,4	0,263157895	8,82802378
5,4	8,3	0,185185185	8,828895166
6,5	7,4	0,153846154	8,829245396
3,7	5,4	0,27027027	8,827944295
3,3	8,2	0,303030303	8,827578185
3,9	9,3	0,256410256	8,828099188
4,4	9,0	0,227272727	8,828424815
2,1	8,2	0,476190476	8,825643029

Příloha 14: Vstupní data a očekávané hodnoty regresní exponenciály

x	***y***	***exp(x)***	***Očekávaná Y***
8,7	11,4	6002,912	11,368424250
8,7	11,5	6002,912	11,368424250
1,3	11,3	3,669297	8,337845407
-1,4	9,3	0,246597	8,336116395
6,2	8,5	492,749	8,584909036
2,4	8,4	11,02318	8,341560295
3,8	7,4	44,70118	8,358573084
5,4	8,3	221,4064	8,447837537
6,5	7,4	665,1416	8,671994915
3,7	5,4	40,4473	8,356424193
3,3	8,2	27,11264	8,349688051
3,9	9,3	49,40245	8,360947976
4,4	9,0	81,45087	8,377137563
2,1	8,2	8,16617	8,340117049

Příloha 15: Vstupní data a očekávané hodnoty regresního log

x	***y***	***log(x)***	***Očekávaná Y***
8,7	11,4	0,939519	9,077670849
8,7	11,5	0,939519	9,077670849
1,3	11,3	0,113943	8,362033838
6,2	8,5	0,792392	8,950135719
2,4	8,4	0,380211	8,592843817
3,8	7,4	0,579784	8,76583986
5,4	8,3	0,732394	8,898127492
6,5	7,4	0,812913	8,967924592
3,7	5,4	0,568202	8,755800302
3,3	8,2	0,518514	8,712729256
3,9	9,3	0,591065	8,77561862
4,4	9,0	0,643453	8,821030364
2,1	8,2	0,322219	8,542574443

Příloha 16: Vstupní data a očekávané hodnoty regresního ln

x	***y***	***ln(x)***	***Očekávaná Y***
8,7	11,4	2,163323	9,077670849
8,7	11,5	2,163323	9,077670849
1,3	11,3	0,262364	8,362033838
6,2	8,5	1,824549	8,950135719
2,4	8,4	0,875469	8,592843817
3,8	7,4	1,335001	8,76583986
5,4	8,3	1,686399	8,898127492
6,5	7,4	1,871802	8,967924592
3,7	5,4	1,308333	8,755800302
3,3	8,2	1,193922	8,712729256
3,9	9,3	1,360977	8,77561862
4,4	9,0	1,481605	8,821030364
2,1	8,2	0,741937	8,542574443