Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

# BAKALÁŘSKÁ PRÁCE

Tomáš Gubanec

# Regresní model ceny nemovitostí

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Martin Šmíd, PhD.
Studijní program: Matematika
Studijní obor: Obecná matematika

2006

Na tomto místě bych rád poděkoval vedoucímu práce RNDr. Martinovi Šmídovi, PhD. za pomoc při návrhu modelů, zapůjčení literatury a mnoho cenných připomínek, které přispěly k celkovému zkvalitnění práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 20.května 2006 Tomáš Gubanec

# Obsah

Název práce: Regresní model ceny nemovitostí
Autor: Tomáš Gubanec
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: RNDr. Martin Šmíd, PhD.
e-mail vedoucího: martin@klec.cz

Abstrakt: Předložená práce se týká trhu s nemovitostmi. Je zaměřena na byty v Praze, speciálně na městské části Praha 1,2 a 11. Studujeme, do jaké míry cena bytu závisí na některých jeho vlastnostech, například výměře, poloze ve městě apod. Používáme dvě varianty lineárního regresního modelu. První je aditivní a druhý multiplikativní. U obou vytváříme výsledný model a odhadujeme regresní parametry. V závěru tyto postupy porovnáváme. Jedna část se zabývá porovnáním reálných cen se zákonem o oceňování majetku. Testujeme zde hypotézu, jestli tento zákon odpovídá skutečné situaci na trhu.

Klíčová slova: byt, nemovitost, cena, regrese, zákon

Title: A regression model of real estate prices
Author: Tomáš Gubanec
Department: Department of probabilty and mathematical statistics
Supervisor: RNDr. Martin Šmíd, PhD.
Supervisor's e-mail address: martin@klec.cz

Abstract: This present work is concerning about the real estate market. It is focused on flats in Prague, especially the districts Prague 1, 2 and 11. We study how much the flat's price depends on some of its characteristics, for example acreage, situation etc. We use two different linear regression models, the first one additive, the second one multiplikative. We also estimate parametres in either of them. At the end of the work we compare these two variants. One part is about evaluation of property. We test hypothesis whether the flat's offer price is equal to lawful price.

Keywords: flat, real property, price, regression, law

# Kapitola 1

# Průběh práce

I naprostému laikovi je jasné, že cena nemovitostí závisí na mnoha různých faktorech. Jako příklad uveďme velikost či stáří nemovitosti, kvalitu okolního prostředí, vzdálenost od metropole apod. Dále je pochopitelné, že u různých typů nemovitostí budou hrát důležitou roli různé faktory. Konkrétněji můžeme u bytů sledovat závislost ceny na počtu poschodí domu, ve kterém se byt nachází. Tento faktor se už neprojeví u rodinných domů, natož u pozemků. Na druhé straně plocha pozemků bývá zadána ve čtvercových jednotkách, ale u domů se často kalkuluje s počtem metrů krychlových obestavěného prostoru.

Při výběru typu nemovitosti, který budu dále zpracovávat, jsem se rozhodl pro byty. Hlavním důvodem byl značně větší počet faktorů, [1] ovlivňujících cenu, ve srovnání s ostatními nemovitostmi. Pro další zjednodušení jsem se rozhodl omezit pouze na pražské byty.

K tomu jsem se rozhodl mj. také proto, že jsem nezjistil existenci žádné podobné práce na toto téma.

Po matematické stránce budu používat teorii lineární regrese, kterou lze nalézt na stranách 111-116 a 187-197 v [Anděl 2003]. Ostatní případné odkazy jsou explicitně uvedeny.

[1] Některé z nich uvedu níže

## 1.1 Aditivní model

Začal jsem tedy zkoumat závislost ceny bytů v Praze na různých faktorech. Data jsem sbíral na stránkách http://www.sreality.cz/ z nabídek různých realitních kanceláří, k datu 1.března 2006. Postupně jsem získal ceny cca 130 bytů v závislosti na různých faktorech. Mezi tyto faktory jsem se snažil uvést takové, které jsou lidmi často vyhledávané a tudíž by jejich vliv na celkovou výši ceny bytu neměl být zanedbatelný. Jako příklad takovýchto veličin nás většinou napadne počet metrů čtverečných (tzv.výměra) bytu, počet pokojů v bytě, kolik podlaží má dům, v němž se byt nachází nebo poloha domu ve městě. Některými z nich bylo nutno zabývat se podrobněji.

Co se týče umístění domu ve městě, jednou z možností je zaměřit se na konkrétní lokalitu a zkoumat pouze byty v ní. Druhým extrémem by bylo sledovat závislost ceny na poloze v každé lokalitě. Jelikož je Praha rozdělena na alespoň dvacet městských částí, byl by takový model výrazně nepřehledný. Připadalo mi tedy rozumné vybrat pro popis polohy bytu v Praze jen několik z těchto oblastí. Rozhodl jsem se pro Prahy 1,2 a 11 (dále P1,P2 a P11). Ty jsem vybral vzhledem k rozdílu atraktivnosti centra (P1 a P2) a ,,klasického sídliště" v P11 na Jižním městě a dále také kvůli rozdílům cen podle zákona v těchto lokalitách.[2]

Ani počet pokojů v bytě nebyl z matematického hlediska bez problémů. Např. dva různé byty 2+1 a 3+kk mají shodně tři obytné místnosti. Abych předešel tomuto nedostatku, uvažoval jsem pro každý byt zvlášť počet pokojů bez kuchyně či kuchyňského koutu (dále jen KK) a jako další faktor přítomnost kuchyně či KK.

Další ,,problémovou" položkou byl stav domu, v němž se byt nachází. Toto nelze jednoznačně vyčíslit. Z pozorovaných dat bylo možno vyčíst několik možností, které prodejci uvádějí. Např.: *novostavba, byt po rekonstrukci, byt ve velmi dobrém či dobrém stavu* apod. Ohodnocení domu těmito výrazy je však čistě subjektivní názor prodejce, který samozřejmě může jednat ve svém zájmu. Přesto jsem se rozhodl zahrnout do modelu i tuto položku, s tím, že jsem použil pouze ty záznamy, ve kterých byl uveden jeden ze stavů ,,dobrý" a ,,velmi dobrý". Tyto dva byly totiž v údajích zastoupeny zdaleka nejčastěji.

[2]Na tomto místě je vhodné poznamenat, že nejdražší základní cena $m^2$ podlahové plochy bytu v Praze podle umístění je právě v P1 a to 49083 Kč. Na druhém místě je P2 se 38658 Kč. Následuje mnoho dalších a teprve potom P11 s 30000 Kč. Podrobnosti lze nalézt v příloze č. 17 v [zákon].

Pro zkoumání závislosti ceny na níže uvedených faktorech jsem nakonec použil klasický lineární regresní model

$$Y = a_1 L_1 + a_2 L_2 + a_{11} L_{11} + cV + b_1 X_1 + \ldots + b_7 X_7 + \epsilon, \tag{1.1}$$

kde

- $Y$ je nabídková cena (v Kč),
- $L_i = \begin{cases} 1 & \text{pokud je byt v Praze } i, \\ 0 & \text{pokud není v Praze } i, \end{cases}$
- $V$ je výměra bytu (v $m^2$),
- $X_1$ určuje počet pokojů v bytě bez kuchyně či KK,
- $X_2 = \begin{cases} 1 & \text{pokud je v bytě kuchyň,} \\ 0 & \text{pokud je v bytě KK,} \end{cases}$
- $X_3$ udává počet podlaží v domě,
- $X_4$ značí, v kolikátém patře je byt,
- $X_5 = \begin{cases} 1 & \text{pokud byt není v přízemí,} \\ 0 & \text{pokud byt je v přízemí,} \end{cases}$
- $X_6 = \begin{cases} 1 & \text{pokud je byt v osobním vlastníctví,} \\ 0 & \text{pokud je byt v družstevním vlastnictví,} \end{cases}$
- $X_7 = \begin{cases} 1 & \text{pokud je stav budovy velmi dobrý,} \\ 0 & \text{pokud je stav budovy dobrý,} \end{cases}$
- $\epsilon$ je vektor chyb.

Původně jsem plánoval do tohoto modelu zahrnout ještě veličinu

$$X_8 = \begin{cases} 1 & \text{pokud je byt v cihlovém domě,} \\ 0 & \text{pokud je byt v panelovém domě.} \end{cases}$$

Mohlo by však dojít k problémům vyplývajících ze závislosti $X_8$ a $L_{11}$ (tento problém se nazývá *multikolinearita*), neboť se často vyskytují panelové domy v P11 a naopak cihlové v P1 a P2 . Mezi příslušnými soubory dat jsem vypočítal koeficient korelace. Ten vyšel roven přibližně $-0.95$, tedy v absolutní hodnotě blízko jedné, což vykazuje značnou závislost mezi těmito daty. Proto jsem se rozhodl nadále $X_8$ nepoužívat.

Nejprve jsem testoval hypotézu, zda vůbec model (1.1) je použitelný, tj. jestli vůbec existuje nějaká závislost $Y$ na regresorech. Toto se obvykle zjišťuje pomocí $F$-testu o regresním modelu, který má nulovou hypotézu, že jsou závislé proměnné rovny pouze konstantě plus šumu, tedy

$H_0$ : všechny koeficienty kromě absolutního členu jsou nulové, proti
$H_1$ : alespoň jeden z koeficientů kromě absolutního členu je nenulový.

I když tento test nelze použít přímo na model (1.1), který nemá absolutní člen, testoval jsem tímto způsobem ekvivalentní model, vzniklý z (1.1) vyškrtnutím jednoho z regresorů určujícího polohu a přidáním absolutního členu. Pomocí softwaru MS Excel jsem vypočítal hodnotu F-statistiky 71.74, která výrazně přesahuje kritickou hodnotu $1,87$,[3] což je příslušná mez pro zamítání hypotézy. Přikláníme se tedy k alternativě, že alespoň jeden z regresních koeficientů je nenulový. Navíc je vidět, že hodnota F-statistiky značně přesahuje kritickou hodnotu, což můžeme chápat tak, že model (1.1) není navržen úplně špatně.

Poté jsem se podrobněji zabýval jednotlivými koeficienty a pro každý z nich testoval hypotézu

$$H_0 : \text{koeficient je nulový proti } H_1 : \text{koeficient není nulový.}$$

Výsledky jsem obdržel opět za pomoci MS Excel. Hodnoty t-statistik jsou uvedeny v tabulce 1.1, významné koeficienty jsou zvýrazněny tučně.

K určení významnosti číselných výsledků pro jednotlivé testy jsem použil vypočtené p-value.[4]

Jak je tedy patrno z tabulky 1.1, koeficienty vycházející významně jsou $a_2$, $a_{11}$, $c$ a $b_6$. Jinými slovy cena bytů ve zvoleném segmentu trhu závisí

[3] Ve skutečnosti jsem použil hodnot $F_{12,120}(0,05) = 1,83$ a $F_{10,120}(0,05) = 1,91$, uvedených na str. 265 v [Anděl 2003] pro odhad potřebné $F_{11,120}(0,05) = 1,87$.

[4] Často se pro určení hladiny významnosti místo p-value používá srovnání s tabulkovými hodnotami t-rozdělení. Protože pro vyšší počty stupňů volnosti nejsou tabelovány, používá se jejich aproximace. Tím by zde mohlo dojít ke zbytečným nepřesnostem.

| koeficient | t-statistika | odhad koef. | p-value |
|---|---|---|---|
| $a_1$ | -0,185 | -140884 | 0,853364 |
| $\boldsymbol{a_2}$ | -2,518 | -1648296 | 0,013123 |
| $\boldsymbol{a_{11}}$ | -3,784 | -3392403 | 0,000242 |
| $\boldsymbol{c}$ | 10,102 | 64469 | $9,02.10^{-18}$ |
| $b_1$ | 0,198 | 38136 | 0,843289 |
| $b_2$ | -1,591 | -476564 | 0,114301 |
| $b_3$ | 0,685 | 76962 | 0,494571 |
| $b_4$ | 1,227 | 92583 | 0,222389 |
| $b_5$ | -0,565 | -230501 | 0,573169 |
| $\boldsymbol{b_6}$ | 2,899 | 922663 | 0,004447 |
| $b_7$ | 1,269 | 339045 | 0,206900 |

Tabulka 1.1: Výsledky testů o jednotlivých parametrech

| koeficient | odhad |
|---|---|
| $b_6$ | 1012947 |
| $c$ | 65417 |
| $a_2$ | -1236064 |
| $a_{11}$ | -2437656 |

Tabulka 1.2: Výsledné odhady významných parametrů

významně na výměře, faktu, zda je byt v osobním či družstevním vlastnictví a poloze v P11 či P2.

Použitím MS Excel při testování statistik o „nulovosti" jednotlivých koeficientů dostaneme i jejich bodové odhady. Zaokrouhlené výsledky jsou pro lepší přehlednost též uvedeny v tabulce 1.1.

Poté jsem z rovnice (1.1) odstranil ty regresory, jejichž koeficienty „vyšli" nevýznamně. Tím je „lépe" vyjádřena cena na zbylých faktorech. Následně jsem provedl nové odhady. Výsledky ukazuje tabulka 1.2.

Podívejme se nyní na interpretaci výsledků této části. Na základě dat, týkajících se náhodně vybraného bytu v P1, P2 nebo P11, můžeme odhadnout jeho tržní cenu pomocí následujícího vzorce, který dostaneme dosazením odhadnutých významných koeficientů do modelu (1.1)

$$Y = 65147V + 1012947X_6 - 1236064L_2 - 2437656L_{11}. \qquad (1.2)$$

Ze statistického hlediska by bylo správné otestovat nyní tento model na dalších bytech. Jelikož však data, použitá při jeho tvorbě, byla z aktuální nabídky v určitém časovém okamžiku, nemáme jich pro tuto práci dostatek. [5]

Překvapující výsledek je pro mě záporná hodnota odhadu koeficientu $b_5$ i když se regresor $X_5$ v daném modelu ukazuje jako nevýznamný. V praxi naopak bývá cena bytu nižší, pokud se nachází v přízemí. Lidé se totiž bojí například vykradení.

## 1.2 Srovnání se zákonem

V této části budeme sledovat, jak nabídková cena bytu v P1, P2 a P11 odpovídá ocenění bytu podle zákona. Nejprve se stručně podíváme na obecné schéma oceňování majetku. Přesné znění lze nalézt v paragrafu 26 v [zákon].

Zákon používá pro byty v Praze vzoreček

$$Y_z = gVE, \quad E = \frac{\sum_{k=1}^{22} v_k c_k}{\sum_{k=1}^{22} v_k}, \tag{1.3}$$

kde

- $Y_z$ je odhadní cena podle zákona (v Kč),
- $V$ je výměra bytu (v $m^2$),
- $k = 1, \ldots, 22$ jsou různé vlastnosti (znaky) bytu, uvedené v zákoně,
- $v_k$ je váha vlastnosti $k$,
- $c_k$ je ohodnocení bytu vlastností $k$,[6]
- $g = g_i$, pokud je byt v Praze $i$, přičemž $g_i$ jsou pevné konstanty určené zákonem pro $i = 1, \ldots, 28$.

Poslední položku ve výčtu můžeme ovšem vyjádřit jako

$$g = \prod_{i=1}^{28} (g_i)^{L_i},$$

[5] Také by bylo možno pracovat s cenami, aktuálními v jiném čase. V takovém případě by však došlo ke zkreslení pohybem cen na trhu.

[6] Pokud má byt znak $k$ průměrný, je $c_k = 1$, pro nadprůměrný je $c_k > 1$ a pro podprůměrný je $c_k < 1$.

kde

$$L_i = \begin{cases} 1 & \text{pokud je byt v Praze } i, \\ 0 & \text{pokud není v Praze } i. \end{cases}$$

Protože budeme vzorec (1.3) používat pro data, zpracovávaná v předchozí části, omezme se na $i \in \{1, 2, 11\}$. Potom

$$g = (g_1)^{L_1}(g_2)^{L_2}(g_{11})^{L_{11}}. \tag{1.4}$$

Hodnoty $c_k$ často nejsou v nabídce bytů k dispozici, neboť se většinou jedná o užší specifikace, proto budeme považovat $E$ za náhodné.

Z důvodu přehlednosti pišme nyní $\hat{x}$ místo $\ln x$, pro reálná kladná $x$.

Předpokládejme dále, že $Y_z = d \cdot Y \cdot f$ kde $d$ je konstanta (průměrný procentní podíl tržní a nabídkové ceny), $Y$ je nabídková cena a $f$ je multiplikativní šum, splňující $\hat{f} \sim \mathrm{N}(0, \sigma^2)$. Po zlogaritmování a použití (1.3) a (1.4) nám pak vychází

$$\hat{Y} + \hat{d} + \hat{f} = \hat{g}_1 L_1 + \hat{g}_2 L_2 + \hat{g}_{11} L_{11} + \hat{V} + \hat{E}.$$

Po úpravě dostáváme

$$\hat{Y} = \hat{g}_1 L_1 + \hat{g}_2 L_2 + \hat{g}_{11} L_{11} + \hat{V} - \hat{d} + [\hat{E} - \hat{f}],$$

což vypadá jako lineární regresní model, až na maličkost, že náhodná složka $[\hat{E} - \hat{f}]$ nemusí mít nulovou střední hodnotu. Ale pokud napíšeme

$$\hat{Y} = \hat{g}_1 L_1 + \hat{g}_2 L_2 + \hat{g}_{11} L_{11} + \hat{V} + \mu + \mathcal{E},$$

$$\mu = \mathbb{E}[\hat{E} - \hat{f}] - \hat{d} = \mathbb{E}\hat{E} - \hat{d}, \quad \mathcal{E} = [\hat{E} - \hat{f}] - \mathbb{E}\hat{E},$$

máme lineární regresní model

$$\hat{Y} = \mu + \alpha_1 L_1 + \alpha_2 L_2 + \alpha_{11} L_{11} + \gamma \hat{V} + \mathcal{E}, \tag{1.5}$$

ve kterém bychom rádi testovali hypotézu

$$H_0: \quad \alpha_1 = \hat{g}_1, \quad \alpha_2 = \hat{g}_2, \quad \alpha_3 = \hat{g}_3, \quad \gamma = 1. \tag{1.6}$$

Pokud by vyšel test významně, zákon nevystihuje trh, pokud nevýznamně, svědčilo by to pro zákon.

Malý problém ovšem nastává v tom, že model (1.5) nemá plnou hodnost. K tomu stačí uvážit, že obsahuje absolutní člen $\mu$, tudíž má matice regresorů v prvním sloupci samé jedničky. Na druhé straně vždy platí

$$L_1 + L_2 + L_{11} = 1, \tag{1.7}$$

protože každý byt leží v právě jedné části Prahy, tedy součet tří sloupců v matici dává další sloupec.

Závažnější problém testování hypotézy (1.6) spočívá v tom, že samostatné parametry $\alpha_1$, $\alpha_2$ a $\alpha_{11}$ nejsou odhadnutelné.[7] Tím pádem je nutno model (1.5) a hypotézu (1.6) modifikovat. Zavedeme-li

$$\theta = \mu + \alpha_{11}, \quad \zeta_1 = \alpha_1 - \alpha_{11}, \quad \zeta_2 = \alpha_2 - \alpha_{11},$$

dostaneme z (1.5) spolu s použitím (1.7) model

$$\hat{Y} = \theta + \zeta_1 L_1 + \zeta_2 L_2 + \gamma \hat{V} + \mathcal{E},$$

který má nyní plnou hodnost. V něm můžeme testovat hypotézu

$$H_0 : \quad \zeta_1 = \hat{g}_1 - \hat{g}_{11}, \quad \zeta_2 = \hat{g}_2 - \hat{g}_{11}, \quad \gamma = 1. \tag{1.8}$$

Parametry $\zeta_1$ (resp.$\zeta_2$) se dají interpretovat jako nějaký bonus za polohu v P1 (resp. P2) oproti P11. Víme (viz poznámka 2 na str. 6), jaké by tyto bonusy měly být podle zákona. Pokud vyjde test (1.8) nevýznamně, přikloníme se k variantě, že *rozdíly* cen odpovídají zákonu. Pokud významně, přikloníme se k tomu, že nabídkové ceny neodpovídají cenám podle zákona.

Pro test (1.8) jsem použil statistiku z Věty 9 na stranách 101, 102 v [Anděl 1985]. Výsledek 47,71 výrazně přesahuje hodnotu $F_{3,128}(0,05) \doteq 2,68$ potřebnou pro zamítnutí hypotézy.

Výsledky této části přesvědčivě svědčí proti tomu, že by 1.března 2006 zákon odpovídal realitě. To se ovšem dalo očekávat, na základě zkušeností s cenami na trhu.

[7] K důkazu neodhadnutelnosti použijeme první část Věty 8 na straně 137 v [Anděl 1985] a následující úvahu (s použítím stejného značení jako v odkázané literatuře). Řádky matice $X$ mají v našem případě na prvním místě jedničku. Na druhém, třetím a čtvrtém místě nulu nebo jedničku (podle části Prahy), ale podle (1.7) je součet těchto tří čísel vždy roven jedné. Na pátém místě je nějaké číslo odpovídající logaritmu výměry. Je-li tedy vektor $\boldsymbol{c}$ nějakou lineární kombinací řádků matice $X$, má na prvním místě nějaké číslo, stejné jako součet čísel na druhém, třetím a čtvrtém místě. Parametr $\alpha_1$ je podle odkazované Věty 8 odhadnutelný právě když vektor $\boldsymbol{c} = (0,1,0,0,0)^T$ je lineární kombinací řádků matice $X$. To ale není podle výše zmíněné úvahy nikdy. Analogicky bychom dostali neodhadnutelnost $\alpha_2$ a $\alpha_{11}$.

## 1.3 Multiplikativní model

V části 1.1 jsme se zabývali klasickým lineárním regresním modelem (1.1). Ze vztahů (1.3) a (1.4) v části 1.2 je však patrné, že cena by mohla na některých faktorech záviset multiplikativně a ne aditivně. Tuto myšlenku použijeme navzdory tomu, že jsme zamítli hypotézu (1.8), čímž jsme popřeli shodu zákona s realitou a tím jsme zpochybnili platnost vzorce (1.3), v němž nahradíme skutečnou cenu $Y_z$ nabídkovou cenou $Y$.

Tato motivace nás přivádí k následujícímu modelu, který budeme nazývat *multiplikativní*.

$$Y = {a_1}^{L_1} \cdot {a_2}^{L_2} \cdot {a_{11}}^{L_{11}} \cdot V^c \cdot {X_1}^{b_1} \cdot {b_2}^{X_2} \cdot {X_3}^{b_3} \cdot {X_4}^{b_4} \cdot {b_5}^{X_5} \cdot {b_6}^{X_6} \cdot {b_7}^{X_7} \cdot \epsilon,$$

kde popis jednotlivých faktorů je stejný jako v modelu (1.1), jenom $\epsilon$ je tentokrát multiplikativní šum, splňující $\ln e \sim \mathrm{N}(0, \sigma^2)$. Po zlogaritmování a použití stejného označení jako v předchozí kapitole ($\hat{x} := \ln x$) dostaneme klasický lineární regresní model

$$\hat{Y} = \hat{a}_1 L_1 + \hat{a}_2 L_2 + \hat{a}_{11} L_{11} + c\hat{V} + b_1 \hat{X}_1 + \hat{b}_2 X_2 + b_3 \hat{X}_3 + b_4 \hat{X}_4 + \hat{b}_5 X_5 + \hat{b}_6 X_6 + \hat{b}_7 X_7 + \hat{\epsilon}.$$

Dále analogickým postupem jako v části 1.1 jsem provedl $F$-test o regresním modelu, tj. testoval hypotézu

$H_0$ : všechny koeficienty kromě absolutního členu jsou nulové, proti
$H_1$ : alespoň jeden z koeficientů kromě absolutního členu je nenulový.

Hodnota F-statistiky vyšla $168,01$, tudíž se přikláníme k alternativě $H_1$.

Podívejme se ještě do tabulky 1.3 na výsledky testů hypotéz

$H_0$ : koeficient je nulový proti $H_1$ : koeficient není nulový.

Významně tedy vyšel test pro $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_{11}$, $c$, $\hat{b}_6$ a $\hat{b}_7$, takže cena závisí na poloze v P1, P2 i P11, výměře, osobním či družstevním vlastnictví a stavu budovy.

Podobně jako v části 1.1 jsem hledal odhady významných koeficientů v modelu, který už nedisponuje nevýznamnými regresory. Výsledky, spolu s přepočítanými hodnotami u původně logaritmovaných koeficientů jsou znázorněny v tabulce 1.4.

Na základě těchto hodnot můžeme sestavit výsledný model, popisující závislost nabídkové ceny na významných veličinách následovně

$$Y = 125046^{L_1} \cdot 96628^{L_2} \cdot 53659^{L_{11}} \cdot V^{0,83} \cdot 1,22^{X_6} \cdot 1,01^{X_7}. \tag{1.9}$$

| koeficient | t-stat. | p-hodnota |
|---|---|---|
| $\boldsymbol{\hat{a}_1}$ | 39,37 | $3,39.10^{-71}$ |
| $\boldsymbol{\hat{a}_2}$ | 39,10 | $7,3.10^{-71}$ |
| $\boldsymbol{\hat{a}_{11}}$ | 36,54 | $1,47 \cdot 10^{-67}$ |
| $\boldsymbol{c}$ | 10,75 | $2,24.10^{-19}$ |
| $b_1$ | 0,77 | 0,44092 |
| $\hat{b}_2$ | -0,35 | 0,72331 |
| $b_3$ | 0,33 | 0,74199 |
| $b_4$ | 1,84 | 0,06799 |
| $\hat{b}_5$ | -0,13 | 0,89799 |
| $\boldsymbol{\hat{b}_6}$ | 4,44 | $1,97.10^{-5}$ |
| $\boldsymbol{\hat{b}_7}$ | 2,48 | 0,01466 |

Tabulka 1.3: Výsledky testů v multiplikativním modelu

| koeficient | odhad | koeficient | odhad |
|---|---|---|---|
| $\hat{a}_1$ | 11,736 | $a_1$ | 125046 |
| $\hat{a}_2$ | 11,479 | $a_2$ | 96628 |
| $\hat{a}_{11}$ | 10,890 | $a_{11}$ | 53659 |
| $c$ | 0,830 | | |
| $\hat{b}_6$ | 0,199 | $b_6$ | 1,2207 |
| $\hat{b}_7$ | 0,093 | $b_7$ | 1.0978 |

Tabulka 1.4: Výsledné odhady významných parametrů

# Kapitola 2

# Výsledky

V této kapitole shrneme výsledky kapitoly 1.

Porovnejme nejprve části 1.1 a 1.3. Jak jsem již uvedl v motivaci k multiplikativnímu modelu, vezmeme-li v úvahu ocenění podle zákona, cena bytu závisí na poloze a výměře podle vzorce (1.3). Očekáváme tedy, že multiplikativní model bude lépe popisovat skutečnost, než aditivní model (1.1).

Srovnáme-li dosažené hodnoty $F$-testů u těchto dvou modelů (71,74 a 168,01), můžeme usoudit, že naše očekávání nebude marné.

Z dalších výsledků vidíme, že multiplikativní model odhalil významně stejné koeficienty (a tudíž regresory) jako aditivní model. Navíc však jako důležité parametry přibývají další dva a to poloha v P1 a stav domu.

Z navržených výsledných modelů v těchto dvou částech bychom se tedy raději přiklonili k multiplikativní verzi (1.9).

Povšimněme si ještě odhadnutých hodnot některých koeficientů v (1.9). Pohledem do tabulky 1.4, zjistíme, že $b_6 > 1$ i $b_7 > 1$, což odpovídá našemu očekávání, že cena bytu se zvýší, je-li v osobním vlastnictví (oproti družstvu) nebo je-li stav budovy v lepším stavu.

V části 1.2 jsme zamítli hypotézu (1.8), čímž jsme popřeli, že odhadní cena bytu je shodná s nabídkovou. Na druhé straně víme, že multiplikativní model vystihuje realitu přesněji, než aditivní verze. Tomu vlastně odpovídá výsledek (1.9), který je v součinovém tvaru a hodnoty základů mocnin $L_i$ výrazně přesahují skutečné $g_i$, uvedené např. v poznámce 2 na straně 6.

# Příloha

| č. | $L_1$ | $L_2$ | $L_3$ | $V$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | cena |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 68 | 3 | 0 | 3 | 3 | 1 | 0 | 1 | 3600000 |
| 2 | 0 | 1 | 0 | 89 | 3 | 1 | 6 | 6 | 1 | 1 | 0 | 4613400 |
| 3 | 0 | 1 | 0 | 136 | 1 | 1 | 4 | 4 | 1 | 0 | 1 | 3570000 |
| 4 | 0 | 1 | 0 | 78 | 3 | 0 | 6 | 3 | 1 | 0 | 1 | 3390000 |
| 5 | 0 | 1 | 0 | 45 | 1 | 0 | 8 | 3 | 1 | 0 | 1 | 3000000 |
| 6 | 0 | 0 | 1 | 63 | 3 | 0 | 9 | 9 | 1 | 0 | 0 | 1550000 |
| 7 | 0 | 1 | 0 | 64 | 3 | 1 | 4 | 1 | 0 | 1 | 0 | 2800000 |
| 8 | 0 | 1 | 0 | 37 | 1 | 1 | 6 | 5 | 1 | 0 | 1 | 2600000 |
| 9 | 1 | 0 | 0 | 150 | 3 | 1 | 5 | 4 | 1 | 1 | 1 | 12600000 |
| 10 | 1 | 0 | 0 | 115 | 3 | 1 | 6 | 2 | 1 | 1 | 1 | 5500000 |
| 11 | 0 | 1 | 0 | 36 | 1 | 1 | 4 | 2 | 1 | 0 | 1 | 1600000 |
| 12 | 0 | 1 | 0 | 68 | 3 | 0 | 4 | 4 | 1 | 0 | 1 | 3200000 |
| 13 | 0 | 0 | 1 | 84 | 3 | 1 | 7 | 4 | 1 | 1 | 1 | 2490000 |
| 14 | 1 | 0 | 0 | 90 | 3 | 0 | 5 | 3 | 1 | 1 | 1 | 11137500 |
| 15 | 0 | 0 | 1 | 66 | 3 | 1 | 9 | 8 | 1 | 0 | 0 | 1546000 |
| 16 | 0 | 1 | 0 | 70 | 2 | 1 | 5 | 2 | 1 | 0 | 1 | 3790000 |
| 17 | 0 | 0 | 1 | 65 | 3 | 0 | 8 | 7 | 1 | 0 | 1 | 1900000 |
| 18 | 0 | 0 | 1 | 43 | 2 | 0 | 8 | 1 | 0 | 0 | 1 | 1450000 |
| 19 | 0 | 0 | 1 | 65 | 3 | 1 | 9 | 7 | 1 | 0 | 0 | 1632000 |
| 20 | 0 | 0 | 1 | 82 | 4 | 1 | 12 | 1 | 0 | 0 | 1 | 2250000 |
| 21 | 0 | 1 | 0 | 72 | 2 | 1 | 5 | 1 | 0 | 1 | 0 | 3415500 |
| 22 | 1 | 0 | 0 | 120 | 4 | 1 | 6 | 5 | 1 | 1 | 1 | 8100000 |
| 23 | 1 | 0 | 0 | 90 | 2 | 1 | 5 | 2 | 1 | 1 | 1 | 5980000 |
| 24 | 0 | 0 | 1 | 64 | 3 | 0 | 8 | 6 | 1 | 0 | 1 | 1788000 |
| 25 | 1 | 0 | 0 | 97 | 3 | 1 | 6 | 3 | 1 | 1 | 0 | 6500000 |
| 26 | 0 | 0 | 1 | 72 | 3 | 0 | 12 | 5 | 1 | 0 | 1 | 1745000 |

| č. | $L_1$ | $L_2$ | $L_3$ | $V$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | cena |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | 0 | 0 | 1 | 72 | 3 | 0 | 12 | 4 | 1 | 0 | 1 | 1825000 |
| 28 | 1 | 0 | 0 | 142 | 5 | 1 | 4 | 4 | 1 | 1 | 0 | 11500000 |
| 29 | 1 | 0 | 0 | 80 | 3 | 1 | 6 | 3 | 1 | 1 | 1 | 7500000 |
| 30 | 0 | 0 | 1 | 61 | 2 | 0 | 11 | 4 | 1 | 0 | 1 | 1710000 |
| 31 | 0 | 0 | 1 | 68 | 3 | 0 | 9 | 7 | 1 | 0 | 1 | 1782000 |
| 32 | 0 | 1 | 0 | 190 | 4 | 0 | 7 | 7 | 1 | 1 | 1 | 10850000 |
| 33 | 0 | 0 | 1 | 92 | 4 | 1 | 4 | 3 | 1 | 0 | 1 | 2370000 |
| 34 | 0 | 0 | 1 | 68 | 3 | 0 | 10 | 8 | 1 | 0 | 1 | 1683000 |
| 35 | 1 | 0 | 0 | 82 | 2 | 1 | 3 | 2 | 1 | 1 | 1 | 4000000 |
| 36 | 0 | 0 | 1 | 65 | 3 | 0 | 8 | 8 | 1 | 0 | 0 | 1673100 |
| 37 | 0 | 0 | 1 | 61 | 2 | 0 | 11 | 5 | 1 | 0 | 0 | 1650000 |
| 38 | 0 | 1 | 0 | 60 | 2 | 0 | 6 | 6 | 1 | 1 | 1 | 3990000 |
| 39 | 0 | 1 | 0 | 138 | 4 | 1 | 5 | 4 | 1 | 0 | 1 | 7249900 |
| 40 | 0 | 1 | 0 | 98 | 4 | 1 | 5 | 3 | 1 | 0 | 1 | 4850000 |
| 41 | 1 | 0 | 0 | 16 | 1 | 0 | 5 | 5 | 1 | 1 | 0 | 1855000 |
| 42 | 0 | 0 | 1 | 47 | 1 | 1 | 10 | 8 | 1 | 1 | 0 | 1470000 |
| 43 | 1 | 0 | 0 | 105 | 3 | 1 | 4 | 2 | 1 | 1 | 1 | 11335500 |
| 44 | 1 | 0 | 0 | 23 | 1 | 0 | 8 | 6 | 1 | 1 | 1 | 2020000 |
| 45 | 1 | 0 | 0 | 40 | 2 | 0 | 7 | 6 | 1 | 1 | 0 | 3250000 |
| 46 | 0 | 1 | 0 | 152 | 4 | 1 | 6 | 5 | 1 | 1 | 1 | 12848456 |
| 47 | 0 | 1 | 0 | 103 | 2 | 1 | 6 | 4 | 1 | 1 | 1 | 7951340 |
| 48 | 0 | 1 | 0 | 117 | 2 | 1 | 6 | 3 | 1 | 1 | 1 | 8443995 |
| 49 | 0 | 1 | 0 | 45 | 1 | 0 | 8 | 3 | 1 | 0 | 1 | 3000000 |
| 50 | 0 | 0 | 1 | 68 | 3 | 0 | 9 | 1 | 0 | 0 | 1 | 1707750 |
| 51 | 0 | 0 | 1 | 44 | 2 | 0 | 7 | 4 | 1 | 1 | 0 | 1560600 |
| 52 | 0 | 0 | 1 | 68 | 3 | 1 | 12 | 2 | 1 | 0 | 1 | 2060000 |
| 53 | 1 | 0 | 0 | 108 | 3 | 1 | 5 | 5 | 1 | 1 | 1 | 8250000 |
| 54 | 0 | 1 | 0 | 97 | 3 | 1 | 5 | 1 | 0 | 1 | 1 | 6790000 |
| 55 | 0 | 0 | 1 | 64 | 3 | 0 | 10 | 4 | 1 | 0 | 0 | 1455300 |
| 56 | 1 | 0 | 0 | 61 | 1 | 1 | 5 | 5 | 1 | 0 | 1 | 3800000 |
| 57 | 0 | 0 | 1 | 42 | 2 | 0 | 11 | 2 | 1 | 0 | 1 | 1673000 |
| 58 | 0 | 1 | 0 | 180 | 5 | 1 | 6 | 6 | 1 | 1 | 1 | 15981000 |
| 59 | 0 | 0 | 1 | 92 | 4 | 1 | 4 | 3 | 1 | 0 | 1 | 2370000 |
| 60 | 0 | 0 | 1 | 84 | 3 | 1 | 7 | 4 | 1 | 1 | 1 | 2490000 |
| 61 | 0 | 0 | 1 | 84 | 3 | 1 | 7 | 5 | 1 | 0 | 0 | 2147310 |
| 62 | 0 | 1 | 0 | 64 | 3 | 1 | 4 | 1 | 0 | 1 | 0 | 2800000 |

| č. | $L_1$ | $L_2$ | $L_3$ | $V$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | cena |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 63 | 0 | 1 | 0 | 69 | 3 | 0 | 6 | 4 | 1 | 0 | 0 | 3250000 |
| 64 | 1 | 0 | 0 | 178 | 4 | 1 | 6 | 3 | 1 | 1 | 1 | 12900016 |
| 65 | 1 | 0 | 0 | 220 | 6 | 1 | 5 | 5 | 1 | 1 | 1 | 17000000 |
| 66 | 1 | 0 | 0 | 109 | 3 | 1 | 7 | 3 | 1 | 1 | 0 | 6590000 |
| 67 | 1 | 0 | 0 | 92 | 3 | 1 | 5 | 5 | 1 | 1 | 1 | 7290000 |
| 68 | 0 | 1 | 0 | 45 | 2 | 0 | 6 | 2 | 1 | 1 | 0 | 2970000 |
| 69 | 0 | 1 | 0 | 52 | 2 | 1 | 5 | 1 | 0 | 1 | 1 | 2785860 |
| 70 | 0 | 0 | 1 | 43 | 2 | 0 | 12 | 5 | 1 | 0 | 0 | 1336500 |
| 71 | 0 | 1 | 0 | 57 | 2 | 0 | 5 | 1 | 0 | 0 | 1 | 2767050 |
| 72 | 0 | 1 | 0 | 72 | 3 | 0 | 6 | 3 | 1 | 1 | 1 | 3500000 |
| 73 | 0 | 0 | 1 | 88 | 3 | 1 | 5 | 3 | 1 | 0 | 0 | 2380950 |
| 74 | 0 | 1 | 0 | 42 | 1 | 0 | 4 | 2 | 1 | 1 | 1 | 2705000 |
| 75 | 0 | 0 | 1 | 70 | 3 | 1 | 12 | 5 | 1 | 0 | 1 | 1700000 |
| 76 | 0 | 0 | 1 | 63 | 3 | 0 | 8 | 8 | 1 | 0 | 0 | 1600000 |
| 77 | 1 | 0 | 0 | 110 | 3 | 1 | 6 | 3 | 1 | 1 | 0 | 6990000 |
| 78 | 0 | 0 | 1 | 64 | 3 | 0 | 9 | 6 | 1 | 0 | 0 | 2050000 |
| 79 | 0 | 1 | 0 | 90 | 2 | 1 | 4 | 1 | 0 | 0 | 1 | 3799000 |
| 80 | 0 | 1 | 0 | 93 | 2 | 1 | 5 | 1 | 0 | 0 | 0 | 4320000 |
| 81 | 1 | 0 | 0 | 80 | 2 | 1 | 5 | 2 | 1 | 1 | 1 | 11000000 |
| 82 | 1 | 0 | 0 | 60 | 2 | 0 | 5 | 2 | 1 | 1 | 1 | 3950100 |
| 83 | 0 | 0 | 1 | 84 | 3 | 1 | 11 | 5 | 1 | 0 | 1 | 2120000 |
| 84 | 0 | 1 | 0 | 40 | 2 | 0 | 5 | 1 | 0 | 1 | 1 | 2950000 |
| 85 | 0 | 1 | 0 | 53 | 3 | 0 | 5 | 3 | 1 | 1 | 1 | 3950000 |
| 86 | 0 | 0 | 1 | 65 | 3 | 0 | 8 | 6 | 1 | 0 | 0 | 1600000 |
| 87 | 0 | 0 | 1 | 69 | 3 | 1 | 12 | 1 | 0 | 1 | 1 | 2702700 |
| 88 | 0 | 1 | 0 | 73 | 1 | 1 | 5 | 5 | 1 | 1 | 1 | 4120000 |
| 89 | 0 | 1 | 0 | 155 | 4 | 1 | 3 | 2 | 1 | 1 | 1 | 7500000 |
| 90 | 0 | 1 | 0 | 42 | 1 | 0 | 4 | 2 | 1 | 1 | 1 | 2870000 |
| 91 | 0 | 0 | 1 | 45 | 2 | 0 | 8 | 5 | 1 | 1 | 0 | 1534500 |
| 92 | 0 | 1 | 0 | 98 | 4 | 1 | 5 | 3 | 1 | 0 | 1 | 4850000 |
| 93 | 0 | 0 | 1 | 30 | 1 | 0 | 7 | 2 | 1 | 0 | 0 | 899000 |
| 94 | 1 | 0 | 0 | 100 | 4 | 0 | 6 | 3 | 1 | 1 | 1 | 7128000 |
| 95 | 0 | 1 | 0 | 117 | 4 | 1 | 4 | 1 | 0 | 1 | 1 | 6600000 |
| 96 | 0 | 1 | 0 | 68 | 3 | 0 | 5 | 3 | 1 | 0 | 1 | 3267000 |
| 97 | 0 | 1 | 0 | 68 | 3 | 0 | 3 | 3 | 1 | 0 | 0 | 3264000 |
| 98 | 1 | 0 | 0 | 115 | 4 | 1 | 4 | 2 | 1 | 1 | 1 | 9800000 |

| č. | $L_1$ | $L_2$ | $L_3$ | $V$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | cena |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 99 | 0 | 0 | 1 | 42 | 2 | 0 | 12 | 2 | 1 | 0 | 1 | 1522000 |
| 100 | 1 | 0 | 0 | 115 | 3 | 1 | 5 | 2 | 1 | 1 | 1 | 5580000 |
| 101 | 0 | 0 | 1 | 42 | 2 | 0 | 12 | 10 | 1 | 0 | 1 | 1419000 |
| 102 | 0 | 0 | 1 | 41 | 1 | 1 | 12 | 9 | 1 | 0 | 1 | 1412000 |
| 103 | 1 | 0 | 0 | 65 | 2 | 1 | 3 | 2 | 1 | 1 | 0 | 4000000 |
| 104 | 1 | 0 | 0 | 120 | 4 | 0 | 5 | 2 | 1 | 1 | 1 | 7100000 |
| 105 | 1 | 0 | 0 | 85 | 2 | 1 | 5 | 1 | 0 | 1 | 1 | 5490000 |
| 106 | 1 | 0 | 0 | 114 | 3 | 1 | 5 | 2 | 1 | 1 | 1 | 9500020 |
| 107 | 1 | 0 | 0 | 72 | 2 | 1 | 3 | 2 | 1 | 1 | 1 | 4160000 |
| 108 | 0 | 0 | 1 | 42 | 2 | 0 | 12 | 2 | 1 | 0 | 1 | 1595000 |
| 109 | 0 | 0 | 1 | 61 | 2 | 0 | 11 | 4 | 1 | 0 | 1 | 1710000 |
| 110 | 1 | 0 | 0 | 100 | 3 | 1 | 4 | 1 | 0 | 1 | 0 | 8770000 |
| 111 | 0 | 0 | 1 | 43 | 2 | 0 | 8 | 5 | 1 | 0 | 0 | 1485010 |
| 112 | 1 | 0 | 0 | 100 | 4 | 1 | 4 | 2 | 1 | 1 | 1 | 6900000 |
| 113 | 0 | 0 | 1 | 45 | 2 | 0 | 8 | 4 | 1 | 0 | 0 | 1603810 |
| 114 | 1 | 0 | 0 | 82 | 3 | 1 | 4 | 4 | 1 | 1 | 1 | 6990000 |
| 115 | 0 | 1 | 0 | 77 | 2 | 1 | 4 | 1 | 0 | 0 | 1 | 3761010 |
| 116 | 1 | 0 | 0 | 110 | 3 | 1 | 5 | 2 | 1 | 1 | 1 | 6599950 |
| 117 | 1 | 0 | 0 | 108 | 3 | 1 | 4 | 1 | 0 | 1 | 0 | 9550000 |
| 118 | 0 | 0 | 1 | 88 | 3 | 1 | 5 | 3 | 1 | 0 | 1 | 2445000 |
| 119 | 1 | 0 | 0 | 112 | 4 | 0 | 4 | 3 | 1 | 1 | 1 | 10781100 |
| 120 | 1 | 0 | 0 | 54 | 1 | 1 | 5 | 2 | 1 | 1 | 1 | 4940100 |
| 121 | 1 | 0 | 0 | 109 | 3 | 1 | 7 | 3 | 1 | 1 | 0 | 6590000 |
| 122 | 1 | 0 | 0 | 92 | 3 | 1 | 5 | 5 | 1 | 1 | 1 | 7290000 |
| 123 | 1 | 0 | 0 | 111 | 3 | 1 | 6 | 3 | 1 | 1 | 1 | 8662500 |
| 124 | 0 | 0 | 1 | 42 | 2 | 0 | 12 | 2 | 1 | 0 | 1 | 1593000 |
| 125 | 1 | 0 | 0 | 90 | 2 | 1 | 5 | 1 | 0 | 0 | 1 | 8250000 |
| 126 | 0 | 0 | 1 | 61 | 2 | 0 | 11 | 5 | 1 | 0 | 0 | 1650000 |
| 127 | 1 | 0 | 0 | 40 | 1 | 1 | 7 | 4 | 1 | 1 | 0 | 3696660 |
| 128 | 0 | 1 | 0 | 125 | 2 | 1 | 7 | 6 | 1 | 1 | 1 | 8240000 |
| 129 | 0 | 1 | 0 | 91 | 4 | 0 | 4 | 2 | 1 | 0 | 0 | 4350000 |
| 130 | 0 | 1 | 0 | 93 | 2 | 1 | 5 | 1 | 0 | 0 | 0 | 4250000 |
| 131 | 1 | 0 | 0 | 97 | 3 | 1 | 6 | 3 | 1 | 1 | 0 | 6500000 |
| 132 | 1 | 0 | 0 | 102 | 3 | 1 | 6 | 4 | 1 | 1 | 1 | 7250000 |

Tabulka 2.1: Pozorovaná data

# Literatura

[Anděl 1985] Anděl, J.: *Matematická statistika*, SNTL/ALFA, Praha, 1985.

[Anděl 2003] Anděl, J.: *Statistické metody*, Matfyzpress, Praha, 2003.

[zákon] *Úplné znění vyhlášky č. 540/2002 Sb., kterou se provádějí některá ustanovení zákona č. 151/1997 Sb., o oceňování majetku a o změně některých zákonů (zákon o oceňování majetku).*

PŘIJATO K OBHAJOBĚ

31 -05- 2006

O. Jol

PŘEDSEDA KOMISE PRO BSZZ
STUDIJNÍ PROGRAM MATEMATIKA