

**Vysoká škola ekonomická v Praze**

**Fakulta informatiky a statistiky**

**Studijní program: Kvantitativní metody v ekonomice**

**Studijní obor: Statistické a pojistné inženýrství**

Diplomant: Ivana Novosádová

Vedúci diplomovej práce: Ing. Mgr. Martin Dlouhý, Dr., MSc.

**MODELY HODNOTENIA EFEKTÍVNOSTI A ICH APLIKÁCIA NA NEMOCNICE  
V ČR**

školský rok 2005/2006

## **Prehlásenie**

Prehlasujem, že som diplomovú prácu spracovala samostatne a že som uviedla všetky použité pramene a literatúru, z ktorých som čerpala.

V Prahe dňa

## **Pod'akovanie**

Chcela by som pod'akovať Ing. Mgr. Martinovi Dlouhému, Dr., MSc za cenné pripomienky, a trpezlivosť pri mojom spracovávaní diplomovej práce.

## Modely hodnotenia efektívnosti nemocníc v ČR

Efektívnosť zdravotníckych zariadení patrí k jednej z často diskutovaných tém. V oblasti zdravotníctva je táto otázka diskutovanejšia o to viac z dôvodu, že zdravotná starostlivosť je financovaná z verejných zdrojov. Základom pre hodnotenie efektívnosti či je správny výber vstupov a výstupov.

Cieľom tejto práce bolo ohodnotiť efektívnosti nemocníc a liečební dlhodobochorých pomocou dvoch analytických metód a výsledky vzájomne porovnať.

Analyzované dáta pochádzajú z krajských ročeniek „Zdravotníctví“, ktoré vydávajú krajské pobočky Ústavu zdravotníckych informácií a štatistiky ČR. V ročenkách boli dáta vhodné pre naše analýzy publikované pre celkovo 22 nemocníc a 8 liečební dlhodobochorých. Základný súbor bol rozdelený na dva podsúbory – pre nemocnice a LDN a to z dôvodu zabezpečenia vyššieho stupňa homogenity. Premennou reprezentujúcou výstup je počet hospitalizovaných, vstup predstavuje počet lekárov pripadajúcich na zariadenie a počet lôžok v zariadení.

Dáta boli analyzované pomocou analýzy obalu dát (DEA) a pomocou ekonometrickej analýzy. DEA predstavuje metódu, ktorá je založená na teórii lineárneho programovania a ktorá označí podľa zvoleného typu modelu, jednotku za efektívnu alebo neefektívnu. V prípade neefektívnej jednotky, identifikuje zdroj neefektívnosti a príslušné cieľové hodnoty pre dosiahnutie efektívnej hranice. V rámci DEA modelov boli použité model s konštantnými výnosmi z rozsahu (CCR model) a model s variabilnými výnosmi z rozsahu (BCC model). V časti venovanej ekonometrickej analýze som sa pokúsila o odhad Cobb-Douglasovej produkčnej funkcie.

Ako efektívna, bola označená v modeli CCR jedna nemocnica, v prípade zmeny charakteru výnosov (model BCC) bolo celkovo efektívnych šesť jednotiek. Pre súbor LDN model CCR identifikoval 2 efektívne liečebne, model BCC jednotiek päť. Pri odhade Cobb-Douglasovej produkčnej funkcie sa na základe t-testov sa ukázalo, že premenná predstavujúca počet lekárov je štatisticky nevýznamná. Z dôvodu silnej lineárnej závislosti s premennou počet lôžok, bola z ďalších analýz vylúčená. Zamerala som sa na analýzu prípadných odľahlých pozorovaní (ako odľahlé sa ukázali dve jednotky a boli z ďalších analýz vylúčené) a testovanie vlastností reziduí. Konečný tvar modelu bol definovaný pre 20 nemocníc. Vysvetľovanou premennou je premenná predstavujúca počet hospitalizovaných, vysvetľujúcou je premenná predstavujúca počet lôžok. Pre súbor LDN sa nepodarilo definovať Cobb-Douglasovu funkciu, predovšetkým z dôvodu malého počtu pozorovaní.

## **Models of Efficiency Evaluation of Hospitals in the Czech Republic**

The efficiency of hospitals belongs among the most discussed topics of public policy. The fundamental problem of the evaluation of both the efficiency and inefficiency in health care is the definition of inputs and outputs.

To measure efficiency of Czech hospitals by two analytical methods, and to provide comparison of these two methods.

The data comes from the regional yearbooks of the Institute of Health Information and Statistics of the Czech Republic (ÚZIS CR). The first sample consists of 22 acute hospitals, the second sample consists of 8 long-term hospitals. The reason for dividing the data was to guarantee, at least to some extent, homogeneity of product. The single output is the number of admissions, the outputs are the number of physicians and the number of the beds.

The data were analysed by two methods of quantitative economic analysis of production function: data envelopment analysis (DEA) and econometric analysis. DEA is based on the theory of linear programming. In the case that a unit is inefficient, DEA defines the sources of inefficiency and the corresponding target values. There are two DEA models used for the evaluation of the efficiency: the model with the constant–returns–to–scale (CCR model), and the model with variable–returns–to–scale (BCC model). In econometric analysis, the authors used the Cobb-Douglas production function.

One acute hospital was identified as efficient by the CCR model and six hospitals by the BCC model. From the sample of long-term hospitals, the CCR model identified two efficient hospitals and the BCC model identified five efficient hospitals. Further, we formulate the Cobb–Douglas production function. The results of the statistical analysis pointed out that the number of the beds and the number of the physicians are strongly linearly correlated. Statistical tests showed that the number of physicians is not of statistically significant variable. On the basis of the statistical tests we decided to reduce the number of independent variables by excluding the number of physicians from the model. The analysis of influential units marked two acute hospitals, which were excluded from following analysis. The final model was formulated for 20 units and included one dependent variable (the number of admissions) and one independent variable (the number of the beds). For the sample of the long-term hospitals the Cobb-Douglas production function was not estimated due to the small number of the units.

# Obsah

<b>1</b>	<b>ÚVOD, CIELE A ŠTRUKTÚRA PRÁCE.....</b>	<b>2</b>
<b>2</b>	<b>DÁTA A METÓDY.....</b>	<b>4</b>
<b>3</b>	<b>DEA MODELY.....</b>	<b>8</b>
3.1	Klasifikácia DEA modelov.....	10
3.2	CCR model.....	11
3.3	BCC model.....	14
3.4	DEA modely v zdravotníctve.....	16
3.5	Tvorba modelu.....	18
<b>4</b>	<b>EKONOMETRICKÁ ANALÝZA.....</b>	<b>21</b>
<b>5</b>	<b>ANALÝZY A VÝSLEDKY.....</b>	<b>24</b>
5.1	DEA modely.....	24
5.1.1	DEA modely pre nemocnice.....	24
5.1.2	DEA modely pre LDN.....	28
5.2	Odhad Cobb-Douglasovej produkčnej funkcie.....	31
5.2.1	Odhad Cobb-Douglasovej produkčnej funkcie pre nemocnice.....	33
5.2.2	Odhad Cobb-Douglasovej produkčnej funkcie pre LDN.....	54
<b>6</b>	<b>ZÁVERY.....</b>	<b>57</b>
<b>7</b>	<b>POUŽITÁ LITERATÚRA.....</b>	<b>60</b>

# 1 Úvod, ciele a štruktúra práce

Efektívnosť, redukcia neefektívnosti a zdroje neefektívnosti, nielen v oblasti zdravotníctva, sú v súčasnej dobe často diskutované témy. Podľa ekonomickej teórie racionality by mal každý jedinec svojim „racionálnym“ správaním maximalizovať svoj úžitok. Maximálny úžitok je dosiahnutý pri optimálnom využití zdrojov. Ale čo je to „optimálne využitie“? Ako zistiť, že práve ten ktorý spôsob využívania je ten „najefektívnejší“? Tieto otázky si kladieme sami (alebo by sme si ich mali klásť) ako jedinci, ale aj spoločnosť ako celok. Využíva spoločnosť pri prerozdelení verejných prostriedkov tieto prostriedky efektívne alebo nie?

Zdravotníctvo je jednou z oblastí ktoré je financované z verejných zdrojov (z daní a zdravotného poistenia). Zdravotníctvo v ČR prechádza transformáciou už niekoľko rokov. Je táto postupná transformácia úspešná, zlepšuje sa využívanie prostriedkov, alebo je aj naďalej zdravotníctvo „deravým rybníkom“, do ktorého sa stále len prilievajú peniaze? Pri skúmaní využitia prostriedkov, ako aj pri hodnotení efektívnosti jednotlivých zariadení, je potrebná určitá dávka opatrnosti a nadhľadu.

Pre moje analýzy používam súbor 22 nemocníc a 8 liečební dlhodobó chorých (v ďalšom texte používam českú skratku LDN), ktoré som získala z publikácií „Zdravotníctví“ vydávané krajskými pobočkami Ústavu zdravotníckej informatiky a štatistiky Českej republiky. Dáta pochádzajú z roku 2003. Vstupmi v analýzach sú údaje o počte lôžok jednotlivých zariadení a počte lekárov v lôžkovej časti pripadajúcich na zariadenie. Výstupy predstavujú údaje o počte hospitalizovaných.

Tieto údaje sú analyzované najprv pomocou analýzy obalu dát (Data Envelopment Analysis, v ďalšom texte už len DEA) - prvá časť kapitoly 5. Pomocou DEA odhadnem efektívnosť jednotlivých zariadení za predpokladu konštantných výnosov z rozsahu (modely CCR) a za predpokladu, že výnosy z rozsahu majú variabilný charakter (modely BCC). V druhej polovici kapitoly 5 sa pokúsím o odhad Cobb-Douglasovej funkcie pre nemocnice a LDN, testovanie modelu ako celku, ako aj štatistickej významnosti jednotlivých premenných.

Túto tému som si vybrala preto, lebo diskusia o otázkach efektívnosti a spôsobe využívania finančných prostriedkov v zdravotníctve v súčasnosti nadobúda na intenzívnosti a aktuálnosti. Ale aj z dôvodu, že DEA analýza, ako aj odhad produkčných funkcií v tejto oblasti, je stále niečo „nové“ a „neprebádané“. Mojou snahou a ani cieľom tejto práce nie je neefektívnosť v zdravotníctve riešiť (to nie je v mojich silách), ale naznačiť ako je možné neefektívnosť kvantifikovať.

Medzi obmedzenia, ktoré limitujú analýzy, patria skutočnosti ako: malý počet jednotiek v analyzovanom súbore, fakt, že nie je zaručená homogenita a porovnateľnosť nemocníc medzi sebou (chýbajú údaje o prípadovom mixe). Problémom je aj to, že chýba nejaký „univerzálny“ ukazovateľ kvality poskytovanej starostlivosti. Tieto problémy a nedostatky sprevádzali a aj sprevádzajú takmer všetky analýzy efektívnosti v oblasti zdravotníctva. To všetko môže skresliť výsledky a závery (napr. porovnanie fakultných a okresných nemocníc). Ale aj napriek tomu si myslím, že je lepšia obmedzená analýza ako analýza žiadna.

Práca je formálne rozdelená do šiestich kapitol. V druhej kapitole popisujem jednotlivé metódy, ktoré použijem pre analýzu efektívnosti nemocníc. Použité metódy teoreticky rozoberám postupne v tretej (DEA modely) a štvrtej kapitole (Ekonomická analýza). Piata kapitola je venovaná analýzam, výsledkom a interpretáciám. Diskutujem v nej o možnostiach ďalších analýz. V závere (šiesta kapitola) sumarizujem výsledky.



## 2 Dáta a metódy

Vo svojej práci sa budem snažiť zúročiť teoretické poznatky získané počas môjho štúdia na Vysoké škole ekonomickej a následne aplikovať tieto poznatky na reálne dáta. Dáta pochádzajú z publikácii „Zdravotníctví“ vydávané Ústavom zdravotníckej informatiky a štatistiky ČR ([www.uzis.cz](http://www.uzis.cz)). Použité údaje boli publikované za rok 2003. Žiaľ ukazovatele použité v mojich analýzach, neboli za roky predchádzajúce ani za rok nasledujúci publikované. Preto je súbor dát, ktorý je použitý pre analýzy, početne obmedzený.

Pre analýzy boli použité údaje:

- počet lekárov pripadajúcich na lôžkovú časť,
- počet lôžok,
- počet hospitalizovaných pacientov.

Z celkového počtu viac ako 300 zdravotníckych zariadení boli kompletne údaje, ktoré sú použité v práci, publikované len pre 22 nemocníc a 8 LDN. U zvyšných zariadení boli údaje publikované len čiastočne. Dôvodom bola neochota údaje publikovať alebo údaj nebol v čase publikácie známy. Použité ukazovatele sú vo fyzických jednotkách a preto je možné aplikovať na tieto údaje analýzu obalu dát (prvá časť piatej kapitoly) a ekonometrickú analýzu (druhá časť piatej kapitoly).

Na začiatok uvádzam kompletnú tabuľku dát (tabuľka č. 1), ktoré boli použité v ďalších analýzach, a to sumárne, pre nemocnice i LDN.

Vstupy predstavujú:

- údaje o počte lekárov v lôžkovej časti,
- počte lôžok.

Výstupom je

- počet hospitalizovaných.

**Tabuľka č. 1:** Vstupné údaje

Zariadenie	počet hospitalizovaných	počet lôžok v lôž.časti	Počet lôžok
Fakultní nemocnice Na Bulovce	38697	160,39	1199
Nemocnice Jihlava	24090	80,87	740
Nemocnice Znojmo	21498	82,08	701
Nemocnice Třebíč	20638	75,05	647
Nemocnice Havlíčkův Brod	20637	74,99	575
Nemocnice Kroměříž	17329	51,39	459
Nemocnice Tábor	16923	67,70	564
Nemocnice Prostějov	16566	60,33	477
Nemocnice Vyškov	15826	53,25	498
Nemocnice Písek	15664	40,04	382
Nemocnice Rudolfa a Stefanie, Benešov	14823	50,10	436
Okres.nemocnice J. Hradec	14704	76,52	437
Nemocnice Chrudim	13270	52,64	440
Nemocnice Pelhřimov	12234	47,29	351
Oblastní nemocnice Jičín	12224	51,78	354
Nemocnice Český Krumlov	9482	34,89	282
Nemocnice Domažlice	6950	22,20	206
Nemocnice Stodůlky	6874	19,40	212
Nemocnice Prachatice	6740	27,50	191
Nemocnice Rokycany	6477	22,77	178
Nemocnice Blansko	4925	74,28	213
Nemocnice na Žižkově	1919	8,90	106
LDN na Vojkově, Tehovec	471	6,50	100
LDN Jablonec nad Nisou	454	2,79	67
LDN Lomnice nad Popelkou	347	2,50	59
LDN Hradec Králové	332	2,00	99
LDN Horažďovice	293	4,45	140
LDN Nejdek	173	3,26	90
LDN Chitussiho 1	169	4,18	100
LDN Opočno	122	1,20	53

**Zdroj:** [www.uzis.cz](http://www.uzis.cz)

V ďalších analýzach rozlišujem základný súbor na dve samostatné skupiny: na samostatný súbor pre nemocnice a samostatný súbor pre LDN. Dôvodom je snaha o zabezpečenie homogenity jednotlivých súborov (napr. z hľadiska prípadového mixu). LDN majú oproti nemocniciam výrazne nižší počet lekárov v lôžkovej časti, ako aj nižší počet lôžok a nižší počet hospitalizovaných. Táto disproporcia by mohla spôsobiť nesprávnosť výsledkov analýz a ich následnú mylnú interpretáciu.

Vzhľadom k charakteru a povahe dát je možné k ich analýze použiť dva druhy metód. Najprv teoreticky vysvetlené (tretia a štvrtá kapitola) a následne (v piatej kapitole) aplikované budú metódy:

- Analýza obalu dát
- Ekonometrická analýza produkčných funkcií.

### **Analýza obalu dát (Data Envelopment Analysis - DEA)**

Je relatívne nová (v porovnaní so štatistickými a inými metódami), neparametrická metóda, ktorá je jedným z možných prístupov hodnotenia efektívnosti produkčných jednotiek. Na rozdiel od tradičných prístupov DEA umožňuje individuálne hodnotenie efektívnosti jednotlivých produkčných jednotiek vzhľadom k celému súboru jednotiek, čo patrí medzi jej najväčšie výhody. Okrem rozdelenia jednotiek na efektívne a neefektívne je možné pre neefektívne jednotky identifikovať zdroj neefektívnosti a určiť tak spôsob, akým môže jednotka dosiahnuť hranicu efektívnosti prostredníctvom redukcie, resp. navýšenia vstupov či výstupov (v závislosti od orientácie modelu). Medzi ďalšie prednosti tejto metódy patrí možnosť identifikácie „vzoru“ pre neefektívnu jednotku. Vzorom pre neefektívnu jednotku je jednotka efektívna s podobnou kombináciou vstupov a výstupov. V neposlednom rade je nespornou výhodou možnosť využitia programového riešenia úloh, ktoré je užívateľsky veľmi prijateľné. Medzi softwarové systémy, ktoré sú zamerané na riešenie analýzy obalu dát, je napr. Frontier Analyst, DEA Solver Pro, OnFront a iné. Nevýhodou je deterministický prístup tejto metódy – priamy predpoklad o type výnosov (konštantné – model CCR, variabilné – model BCC). Štatistické testovanie „významnosti“ jednotlivých vstupov a výstupov nie je ešte tak prepracované ako napr. v ekonometrickej analýze. Na druhej strane je to oblasť v ktorej sa dá očakávať ďalší rozvoj, čím určite dôjde k ďalšiemu vylepšeniu tohto prístupu.

### **Ekonometrická analýza produkčných funkcií**

Ekonometrická analýza patrí k tradičným prístupom hodnotenia efektívnosti jednotiek (podobne ako napr. pomerové ukazovatele). Tento prístup je starší, prepracovanejší. V tejto oblasti je rozvinutejšia metodológia, prepracovanejšie testy hypotéz (testy o významnosti jednotlivých premenných a atď.).

Na druhej strane nevýhodou je nutnosť definovania funkčného vzťahu premenných. Pre hodnotenie a tvorbu relevantných záverov musia byť splnené viaceré predpoklady (napr.

o normálnom rozdelení reziduí, o nezávislosti reziduí, atď.), čo môže výrazne komplikovať ďalšie analýzy. V praxi je častým problémom multikolinearita, autokorelácia, či heteroskedasticita a ich riešenie je komplikované.

### 3 DEA modely

Efektívnosť je možné definovať ako pomer výstupu, ktoré jednotka produkuje, k vstupu, ktoré jednotka pri činnosti spotrebováva. Matematické vyjadrenie efektívnosti zobrazuje vzorec (3.1):

$$\text{Efektívnosť} = \frac{\text{výstup}}{\text{vstup}}. \quad (3.1)$$

Pred podrobnejším výkladom teórií jednotlivých DEA modelov zadefinujeme základné pojmy, ktoré budem ďalej v texte používať bez ďalšieho vysvetľovania.

Základnou myšlienkou DEA modelov je, že ku každému problému je možné konštruovať tzv. **množinu prípustných riešení**, ktorá je tvorená všetkými možnými kombináciami vstupov a výstupov. Množinu prípustných riešení určuje tzv. **efektívna hranica**, ktorá je tvorená **efektívnymi jednotkami**. Za efektívnu jednotku je považovaná tá jednotka, pre ktorú reálne neexistuje iná, ktorá by dosiahla rovnakého výstupu pri použití menšieho množstva vstupov, resp. ktorá by dosiahla vyššiu produkciu výstupov pri použití rovnakého množstva vstupov. **Súborom homogénnych jednotiek** sa rozumie súbor jednotiek, ktoré produkujú rovnaké alebo ekvivalentné výstupy.

V reálnych situáciách je často potrebné uvažovať viacero vstupov, resp. viacero výstupov. Preto sa pre agregáciu používa systém váh. Uvažujme teda súbor homogénnych produkčných jednotiek  $U_1, U_2, \dots, U_n$ . Každá z jednotiek produkuje  $r$  výstupov a zároveň spotrebováva  $m$  vstupov. Označme  $\mathbf{X} = \{x_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n\}$  ako maticu vstupov a  $\mathbf{Y} = \{y_{ij}, i = 1, 2, \dots, r; j = 1, 2, \dots, n\}$  ako maticu výstupov. Vstupy a výstupy jednotky  $U_q$  obsahuje  $q$ -tý riadok matic  $X_q$  a  $Y_q$ . Mieru efektívnosti tejto jednotky môžeme všeobecne vyjadriť ako

$$\frac{\text{vážený súčet výstupov}}{\text{vážený súčet vstupov}} = \frac{\sum_i u_i y_{iq}}{\sum_j v_j x_{jq}}, \quad (3.2)$$

kde  $v_j = 1, 2, \dots, m$  sú váhy priradené j-temu vstupu a  $u_i, i = 1, 2, \dots, r$  sú váhy priradené i-temu výstupu.

DEA analýza spočíva v tom, že pri hodnotení efektívnosti produkčnej jednotky  $U_q$  sa maximalizuje jej miera efektívnosti (3.2) pri rešpektovaní podmienky, že miera efektívnosti všetkých ostatných jednotiek daného súboru nemôže byť väčšia než 1 (resp. 100 %). Váhy vstupov a výstupov musia byť zároveň väčšie ako nula, a to z toho dôvodu, aby boli všetky uvažované charakteristiky v modeli zahrnuté.

### 3.1 Klasifikácia DEA modelov

Základnou myšlienkou klasifikácie DEA modelov je ich orientácia na vstupy alebo na výstupy:

- **Modely orientované na výstupy** (*output-oriented models*)  
Tento typ modelov v prípade neefektívnej jednotky navrhuje zvýšenie množstva výstupu pri zachovaní množstva vstupu.
- **Modely orientované na vstupy** (*input-oriented models*)  
V prípade týchto modelov sa pre neefektívne jednotky redukuje množstvo vstupu pri zachovaní súčasného množstva výstupu.
- **Aditívne, resp. odchýlkové modely** (*additive, slack-based models*)  
Tento typ modelov kombinuje oba predchádzajúce prístupy. Pre dosiahnutie efektívnej hranice dochádza súčasnou redukciou, resp. zvýšením vstupov i výstupov. (Podrobnejší výklad tejto časti DEA modelov presahuje rámec tejto práce, pre podrobnejšie informácie vid' Jablonský, Dlouhý, 2004; Cooper, Seiford, Charnes, 1998.)

Ďalším kritériom možnej klasifikácie modelov je predpoklad charakteru výnosov z rozsahu:

- V prípade konštantných výnosov z rozsahu (tj. pri zvýšení vstupu o jednu jednotku sa zvýši výstup presne o jednu jednotku), konštruujeme kónický obal dát. V takomto prípade hovoríme o **CCR modeloch** (vid' ďalej podkapitolu 3.2).
- Ak uvažujeme variabilné výnosy z rozsahu (tj. pri zvýšení vstupu o jednu jednotku výstup rastie o viac ako jednu jednotku / má za následok rast výstupu aspoň o jednu jednotku), definujeme konvexný obal dát a ide o **BCC modely** (vid' podkapitola 3.3) a modely odchýlkové.

### 3.2 CCR model

Tento model bol navrhnutý ako prvý z DEA modelov v roku 1978 pánmi Charnesom, Cooperom, Rhodesom a z ich mien je odvodený aj názov tohto modelu.

Základným predpokladom modelu sú konštantné výnosy z rozsahu:

$$f(tX, tY) = tf(X, Y) = tQ,$$

kde X znamená množstvo spotrebovaných vstupov, Y množstvo produkovaných výstupov, t je ľubovoľná konštanta, pre ktorú platí  $t \neq 0$ . (viď Soukupová, Horejší, Macáková, Soukup, 2004)

CCR model maximalizuje mieru efektívnosti q-tej jednotky (definovanú v účelovej funkcii) za predpokladu dodržania podmienok, že miery efektívnosti ostatných jednotiek sú menšie alebo rovné 1 (obmedzujúce podmienky úlohy). Pomocou váh tak dostávame pre každú jednotku virtuálny vstup a virtuálny výstup. CCR model teda počíta váhy vstupov ( $v_j$ ) a váhy výstupov ( $u_i$ ) tak, aby to bolo pre hodnotenú jednotku čo najvýhodnejšie z hľadiska efektívnosti pri maximálnej jednotkovej efektívnosti ostatných jednotiek. Tento model je teda možné formulovať ako úlohu lineárneho lomeného programovania:

$$\begin{aligned} \text{maximalizovať} \quad & \frac{\sum_i u_i y_{iq}}{\sum_j v_j x_{jq}} \\ \text{za podmienok} \quad & \frac{\sum_i u_i y_{ik}}{\sum_j v_j x_{jk}} \leq 1, & k = 1, 2, \dots, n, \\ & u_i \geq \varepsilon, & i = 1, 2, \dots, r, \\ & v_j \geq \varepsilon, & j = 1, 2, \dots, m. \end{aligned} \quad (3.2.1)$$

Úlohy lomeného lineárneho programovania je možné pomocou jednoduchým matematických úprav previesť na klasickú úlohu lineárneho programovania, a to tak, že buď sa položí menovateľ účelovej funkcie rovný 1 a dostaneme tak **primárny CCR model orientovaný na vstupy (CCR-I)** (3.2.2), alebo čitateľ účelovej funkcie a dostaneme **primárny CCR model orientovaný na výstupy (CCR-O)** (3.2.3).



$$\begin{aligned}
&\text{Maximalizovať} && \sum_i u_i y_{iq}, \\
&\text{za podmienok} && \sum_j v_j x_{jq} = 1, \\
&&& \sum_i^r u_i y_{ik} \leq \sum_j^m v_j x_{jk}, \quad k = 1, 2, \dots, n, \\
&&& u_i \geq \varepsilon, \quad i = 1, 2, \dots, r, \\
&&& v_j \geq \varepsilon, \quad j = 1, 2, \dots, m.
\end{aligned} \tag{3.2.2}$$

$$\begin{aligned}
&\text{Minimalizovať} && \sum_j v_j x_{jq}, \\
&\text{za podmienok} && \sum_i u_i y_{iq} = 1, \\
&&& \sum_i^r u_i y_{ik} \leq \sum_j^m v_j x_{jk}, \quad k = 1, 2, \dots, n, \\
&&& u_i \geq \varepsilon, \quad i = 1, 2, \dots, r, \\
&&& v_j \geq \varepsilon, \quad j = 1, 2, \dots, m.
\end{aligned} \tag{3.2.3}$$

Z interpretačného a výpočtového hľadiska je výhodnejšie pracovať s modelmi duálne združenými k úlohám primárnym. Duálny model dostaneme tak, že každému obmedzeniu primárnej úlohy (3.2.2) priradíme jednu duálnu premennú a každej primárnej premennej priradíme jedno duálne obmedzenie. Riešením duálnej úlohy dostaneme rovnaké výsledky ako riešením úlohy primárnej. (Pre podrobnejší výklad teórie duality vid' Jablonský, Lagová, 2004.) Pre úplnosť uvádzam maticovú formuláciu duálnych modelov.

**Duálny CCR model orientovaný na vstupy** je možné v maticovom vyjadrení zapísať ako úlohu typu:

$$\begin{aligned}
&\text{minimalizovať} && f = \theta_q - \varepsilon(e^T s^+ + e^T s^-), \\
&\text{za podmienok} && Y\lambda - s^+ = y_q, \\
&&& X\lambda + s^- = \theta_q x_q, \\
&&& \lambda, s^+, s^- \geq 0,
\end{aligned} \tag{3.2.4}$$

kde  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ ,  $\lambda \geq 0$  je vektor váh priradený jednotlivým produkčným jednotkám,  $s^+$  a  $s^-$  sú vektory prídavných premenných v obmedzeniach pre vstupy a výstupy,

$e^T$  je jednotkový vektor a  $\varepsilon$  je infinitezimálna konštanta (ľubovoľne malá konštanta, vo všeobecnosti volená napr.  $10^{-6}$ , alebo  $10^{-8}$ ).

Pri hodnotení jednotky  $U_q$  sa model snaží nájsť virtuálnu jednotku charakterizovanú vstupmi  $X\lambda$  a výstupmi  $Y\lambda$ , ktoré sú lineárnou kombináciou vstupov a výstupov ostatných jednotiek súboru, a ktoré sú lepšie (resp. nie sú horšie) než vstupy a výstupy hodnotenej jednotky  $U_q$ . Platí teda, že  $X\lambda \leq X_q$  a  $Y\lambda \geq Y_q$ . Jednotka  $U_q$  je označená ako efektívna, ak virtuálna jednotka s uvedenými vlastnosťami neexistuje, resp. virtuálna jednotka je totožná s hodnotenou jednotkou, tzn. že platí  $X\lambda = X_q$  a  $Y\lambda = Y_q$ . Pri praktickom matematickom výpočte sa táto skutočnosť prejaví tým, že hodnota premennej  $\theta$  je rovná 1 a všetky prídavné premenné  $s^+$ ,  $s^-$  sú rovné 0.

Jednotka je považovaná za CCR efektívnu, ak hodnota účelovej funkcie v modeli (3.2.4) je rovná 1. Optimálna hodnota účelovej funkcie  $f^*$  sa označuje ako miera efektívnosti hodnotenej jednotky. V prípade neefektívnych jednotiek, je hodnota účelovej funkcie menšia ako 1 a aby sa jednotka dostala na efektívnu hranicu a stala sa jednotkou efektívnou, mala by znížiť (redukovať) množstvo vstupov pri zachovaní objemu výstupu.

Tak, ako je možné definovať primárny CCR model orientovaný na vstupy a k nemu združený duálny model, je možné k primárnemu CCR modelu orientovanému na výstupy (3.2.3) definovať duálny CCR model orientovaný na výstupy.

#### **Duálny CCR model orientovaný na výstupy:**

$$\begin{aligned} \text{maximalizovať} \quad & g = \phi_q + \varepsilon(e^T s^+ + e^T s^-), \\ \text{za podmienok} \quad & Y\lambda - s^+ = \phi y_q, \\ & X\lambda + s^- = x_q, \\ & \lambda, s^+, s^- \geq 0. \end{aligned} \tag{3.2.5}$$

Pre tento model platí obdobná interpretácia ako pre model (3.2.4). Hodnotená jednotka je označená za efektívnu v prípade, že účelová funkcia  $g^* = 1$ . Ak je hodnota  $g^*$  väčšia ako 1, je jednotka neefektívna a pre dosiahnutie efektívnej hranice by mala zvýšiť množstvo produkovaných výstupov pri zachovaní spotrebovávaného množstva vstupov. Pre hodnoty účelových funkcií CCR modelov (3.2.4) a (3.2.5) platí vzťah  $f^* = 1/g^*$ .

### 3.3 BCC model

Banker, Charnes a Cooper (1984) modifikovali základný predpoklad CCR modelu o charaktere výnosov z rozsahu. Meno tohto modelu, rovnako ako modelu CCR, je odvodený zo začiatočných písmen pánov, ktorý ho matematicky ako prví formulovali. V modeli BCC sú uvažované variabilné výnosy z rozsahu (klesajúce, rastúce, resp. konštantné):

$$f(tX, tY) < \text{resp.} = \text{resp.} > tf(X, Y) = tQ,$$

kde  $X$  je množstvo spotrebovaných vstupov,  $Y$  množstvo produkovaných výstupov,  $t$  je ľubovoľná konštanta, pre ktorú platí  $t \neq 0$ . (viď Soukupová, Horejší, Macáková, Soukup, 2004). Z tejto definície vyplýva, že konštantné výnosy z rozsahu sú špeciálnym prípadom variabilných výnosov, takže model CCR je možné označiť za špeciálny prípad modelu BCC. Na základe predpokladu o variabilných výnosoch z rozsahu je obal dát definovaný ako konvexný. Z toho vyplýva, že v modeloch BCC je viacej efektívnych jednotiek ako v modeloch CCR (v CCR modeloch je efektívna (v typickom prípade) len jedna jednotka) a platí, že efektívnosť v BCC modeloch by nemala byť menšia (presnejšie povedané, nemala by byť horšia) ako v modeloch CCR. Pre úplnosť a názornú ukážku zdefinujeme primárne BCC modely orientované na vstupy (3.3.1) a na výstupy (3.3.2).

#### Primárny BCC model orientovaný na vstupy (BCC-I):

$$\begin{aligned} &\text{Maximalizovať} && \sum_i^r u_i y_{iq} + \mu, \\ &\text{za podmienok} && \sum_i^r u_i y_{ik} + \mu \leq \sum_j^m v_j x_{jk}, \quad k = 1, 2, \dots, n, \\ &&& \sum_j^m v_j x_{jq} = 1, \\ &&& u_i \geq \varepsilon, \quad i = 1, 2, \dots, r, \\ &&& v_j \geq \varepsilon, \quad j = 1, 2, \dots, m, \\ &&& \mu - \text{ľubovoľné.} \end{aligned} \tag{3.3.1}$$

Je vidieť, že model (3.3.1) a model (3.2.2) sa líšia len v tom, akých hodnôt môže nadobúdať premenná  $\mu$ . V prípade BCC modelu je táto premenná, čo sa hodnôt týka, neobmedzená, v modeli CCR je táto premenná nulová. Analógia platí aj pre modely orientované na výstupy (3.2.3 a 3.3.2). V primárnom BCC modeli premenná  $\nu$  môže nadobúdať ako kladných, tak záporných hodnôt. Naopak, v modeli CCR je táto premenná výhradne nulová.

**Primárny BCC model orientovaný na výstupy (BCC-O):**

$$\begin{aligned}
&\text{minimalizovať} && \sum_j^m v_j x_{iq} + v, \\
&\text{za podmienok} && \sum_i^r u_i y_{ik} \leq \sum_j^m v_j x_{jk} + v, \quad k = 1, 2, \dots, n, \\
&&& \sum_j^m v_j y_{jq} = 1, \\
&&& u_i \geq \varepsilon, \quad i = 1, 2, \dots, r, \\
&&& v_j \geq \varepsilon, \quad j = 1, 2, \dots, m, \\
&&& v - \text{ľubovoľné.}
\end{aligned} \tag{3.3.2}$$

Pre definovanie duálnych BCC modelov stačí rozšíriť model (3.2.4) a model (3.2.5) o podmienku konvexnosti obalu ( $e^T \lambda = 1$ ). Dostávame tak model:

$$\begin{aligned}
&\text{minimalizovať} && f = \theta_q - \varepsilon(e^T s^+ + e^T s^-), \\
&\text{za podmienok} && Y\lambda - s^+ = y_q, \\
&&& X\lambda + s^- = \theta_q x_q, \\
&&& e^T \lambda = 1, \\
&&& \lambda, s^+, s^- \geq 0,
\end{aligned} \tag{3.3.3}$$

ktorý označujeme ako **duálny BCC model orientovaný na vstupy**.

Analogicky sa dá definovať **duálny BCC model orientovaný na výstupy** (3.3.4):

$$\begin{aligned}
&\text{maximalizovať} && g = \phi_q + \varepsilon(e^T s^+ + e^T s^-), \\
&\text{za podmienok} && Y\lambda - s^+ = \phi_q Y_q, \\
&&& X\lambda + s^- = X_q, \\
&&& e^T \lambda = 1, \\
&&& \lambda, s^+, s^- \geq 0.
\end{aligned} \tag{3.3.4}$$

Aby bola jednotka v modeloch BCC označená za efektívnu, musia byť hodnoty pomocných premenných  $s^+$ ,  $s^-$  nulové a hodnota radiálnej premennej  $\phi_q$ , resp.  $\theta_q$  je rovná 1, tzn. hodnota účelovej funkcie v efektívnom prípade je rovná 1.

### 3.4 DEA modely v zdravotníctve

Všeobecným problémom analýz zdravotníctva je nepoužiteľnosť všeobecnej definície produkcie. Použiť definíciu produkcie (výroby) ako transformáciu zdrojov na finálne produkty, nie je možné. V klasickom poňatí produkcie je zákazník na konci produkčného procesu, nie je jeho priamou súčasťou. Zákazník proces výroby ovplyvňuje nepriamo – prostredníctvom dopytu, ktorý reprezentuje jeho potreby a prania. V oblasti zdravotníctva je zákazníkom pacient a ako taký je priamym účastníkom „produkčného“ procesu, ktorým je proces liečenia. V oblasti zdravotníctva je „produkcia“ ihneď spotrebovávaná, nedá sa uložiť na sklad alebo inak uskladniť. To čo je produkované, je ihneď aj spotrebované.

Počiatky DEA modelov sa objavujú v roku 1957 kedy Farell ako prvý navrhol modely pre ohodnotenie efektívnosti. Prvé DEA modely boli prezentované v roku 1978 pánmi Charnesom, Cooperom a Rhodesom. DEA modely sa ďalej vyvíjali a stali sa populárne vďaka svojim výhodám, ktoré už boli spomenuté. Výhodou analýzy obalu dát v zdravotníctve je fakt, že dokáže pracovať s viacnásobnými vstupmi a výstupmi. Pre zdravotníctvo je typické, že často nepoznáme cenové ohodnotenie, a to z dôvodu, že neexistuje vôbec alebo je ohodnotenie stanovené umelo, administratívne.

Prvé aplikácie DEA analýzy v oblasti zdravotníctva sa objavili na začiatku 80. rokov. Jedným z prvých, ktorý využil DEA v tejto oblasti, bol aj pán H. D. Sherman (1981). DEA modely boli použité pri analýze efektívnosti zdravotníckeho personálu. Informácie o konkrétnych aplikáciách DEA modelov v oblasti zdravotníctva je možné nájsť v prehľadovej publikácii Hollingsworth (2003). Na základe uskutočnených aplikácií DEA modelov v tejto oblasti vyplynulo, že DEA je vhodná pre porovnávanie jednotlivých zariadení medzi sebou (nemocnice, LDN, sanatória), konkrétnych oddelení (jednotky intenzívnej starostlivosti, dializačné centrá), ale aj pre vzájomné porovnanie lekárov rovnakých odborností (meranie výkonnosti praktických lekárov).

Žiadna analýza ale nie je taká jednoduchá, ako sa môže na začiatku javiť. Meranie výkonnosti, obzvlášť v zdravotníckych zariadeniach, je náročné. Ciele zdravotníckych zariadení sú komplexné, ťažko merateľné a protikladné. Meranie efektívnosti, či výkonnosti je komplikované z viacerých dôvodov.

- *Problém vymedzenia výkonnosti súvisiace s identifikáciou vhodných vstupov a výstupov.*

V jednotlivých štúdiách/analýzach/prácach s obdobným zameraním sú použité rôzne vstupy a výstupy. Samotná definícia výstupu jednotlivých zariadení je v zdravotníctve problematická.

- *Vzťahy medzi jednotlivými vstupmi a ich vzájomná substitúcia.*

Na základe možnej substitúcie vstupov a výstupov je možné identifikovať tzv. alternatívne spôsoby definície hranice produkčných možností. Príkladom by mohol byť vzťah preventívnej starostlivosti a dní hospitalizácie.

- *Homogenita jednotiek v rámci analýzy.*

Nemocnice, liečebne dlhodobo chorých, sanatóriá atď., každé toto zariadenie je niečím špecifické. Je logické, že pre zaistenie homogenity analyzovaného súboru nebude vhodné miešať takto rozdielne zariadenia do jedného súboru. Ale významný rozdiel môže byť aj v rámci skupiny nemocníc. Správna analýza a interpretácia výsledkov DEA analýz nemôže byť odtrhnutá od „pozadia“. Rozdielny prípadový mix v jednotlivých zariadeniach môže výrazne ovplyvniť výsledky a závery. Obdobne je tomu aj v prípade zariadení s rôznou organizačnou štruktúrou (nezisková organizácia, príspevková organizácia), resp. podľa typu majiteľa zariadenia (súkromné zariadenie, štátne zariadenie). V prípade medzinárodných porovnaní sa nesmie zabudnúť na systém financovania zdravotníckej starostlivosti v jednotlivých štátoch.

Z týchto dôvodov existuje v rámci DEA analýz, v oblasti zdravotníctva nekonzistentnosť výsledkov jednotlivých štúdií a tým ich vzájomná neporovnateľnosť, čím sa znižuje celkový prínos aplikácií DEA modelov v tejto oblasti.

### 3.5 Tvorba modelu

Na základe aplikácií modelov analýzy obalu dát bol zostavený „recept“, ako postupovať pri tvorbe modelov a pri zostavovaní DEA analýz. Tento postup samozrejme nie je špecifický len pre túto oblasť, ale s určitými modifikáciami a úpravami sa dá implementovať i do iných oblastí. Ako teda vyzerá „univerzálny recept“, výsledkom ktorého by mal byť správne zostavený model a na jeho základe správne zostavená analýza? Tento „návod na použitie“ by sa dal rozdeliť do niekoľkých krokov, ktoré na seba nadväzujú a vzájomne sa prelínajú.

#### 1. Identifikácia problémov a cieľov

Každá štúdia a analýza by mala začať vymedzením konkrétneho problému z reálneho života. Otázky, na ktoré sa budeme snažiť v rámci štúdie odpovedať, by mali byť základom. A na základe takto definovaných otázok potom k danej problematike hľadať dáta. Rozhodne by otázky nemali byť kladené na základe údajov, ktoré máme k dispozícii. Takisto je užitočné zistiť, či na danú tému neboli už obdobné analýzy uskutočnené.

#### 2. Zostavenie modelu

Neexistuje žiaden univerzálny postup ako zostaviť model. Každý model všeobecne, nielen modely v DEA, by mali podchycovať určitú reálnu časť ekonomickej reality. Otázkou je, či zostavený model dáva zmysel. Nikdy sa nám nepodarí zachytiť skutočnosť presne. Model má túto skutočnosť prezentovať zjednodušene, ale zároveň tak, aby boli zachované typické a výrazné črty. Preto zostaveniu samotného modelu by mala predchádzať štúdia oblasti, ku ktorej sa daný model vzťahuje. V tejto fázi sa zúročia poznatky načerpané v prvom kroku (skúsenosti z už uskutočnených analýz). Dôležité je vybrať vhodný typ modelu. V oblasti zdravotníctva sú najčastejšie používanými modelmi modely CCR a BCC, zriedkavejšie sa objavujú multiplikatívne a aditívne modely.

#### 3. Množina premenných ovplyvňujúcich daný problém

Nasleduje identifikácia množiny premenných. Dôležitými a často opomínanými faktormi sú faktory sociálne. Medzi významné faktory patria aj charakteristiky pacientov, faktory prostredia, štruktúra prípadového mixu atď. Cieľom tohto kroku je

vytvoriť skupinu premenných, ktoré súvisia s analyzovaným problémom. Nie vždy je však možné číselne podchytiť faktory, ktoré považujeme za významné.

#### 4. Výber faktorov

V ďalšom kroku sa vyberie z identifikovanej množiny premenných množina premenných, ktoré sa použijú ako vstupy a výstupy. Na základe takto vybraných premenných sa zozbierajú, resp. sa zostavia dáta. Častejší je prípad, že dáta sa nezískavajú experimentálne, ale na základe nejakých štatistických zisťovaní už boli zozbierané, takže v nejakej konkrétnej forme už existujú. Preto je časté, že pri zostavovaní základných vstupných dát je nutné pracovať s rôznymi zdrojmi (napr. nemocnice, databázy zdravotných poisťovní), aby sme dostali vstupné dáta vo vhodnej forme.

#### 5. Analýza faktorov pomocou štatistických metód

DEA predpokladá, že model počíta a ohodnocuje efektívnosť na porovnateľných jednotkách. Pred samotnou analýzou je dôležité položiť si otázku, či sú výstupy analyzovaných jednotiek homogénne alebo heterogénne. Z matematického hľadiska sa predpokladá, že existuje podmienený vzťah medzi vstupmi a výstupmi. Nepredpokladá sa, že zvyšovanie vstupov spôsobí znižovanie výstupu. V prípade korelácie a multikolinearity vstupov by malo dôjsť k ich redukcii. Bežným problémom bývajú aj chýbajúce údaje niektorých jednotiek. Jedným zo spôsobov, ako tento problém riešiť, je vynechanie týchto jednotiek, resp. substituovanie týchto údajov určitou vhodne zvolenou konštantou.

#### 6. Výpočet viacerých typov modelov

Sú výsledky logické? V prípade zostavenia viacerých typov modelov, zostávajú vypočítané výsledky stabilné?

#### 7. Analýza výsledkov pomocou štatistických metód

Výsledky DEA analýz by sa mali následne použiť pre testovanie hypotéz. Spájanie DEA so štatistikou je veľmi populárne a určite aj prínosné. Na základe ďalších štatistických postupov sa môžeme utvrdiť o tom, že výsledky sú relevantné a nielen výsledkom akejsi dobrej zhody náhod. V prípade, že sa dostaneme do situácie, že



vysvetľujúce premenné, resp. kontrolné premenné nemajú žiaden významný vzťah k výsledkom modelu, tak je niečo zle.

#### *8. Interpretácia a podelenie sa o výsledky*

Väčšina aplikácií DEA v oblasti zdravotníctva je zameraná skôr na metodológiu, čo môže zmenšiť prínos a význam takýchto štúdií. Pre následnú interpretáciu by sa nemalo zabúdať na spojenie s praxou a neponechávať výsledky len v teoretickej rovine.

## 4 Ekonometrická analýza

Všeobecná definícia **produkčnej funkcie** by sa dala slovne popísať ako funkčný vzťah medzi výstupom (produkciou) a vstupmi (faktormi). Matematické vyjadrenie je vyjadrené nasledovne:

$$Q = f(I_A, I_B, \dots, I_n), \quad (4.1)$$

kde  $Q$  je množstvo produkcie (výstup),  $I_i$  je množstvo spotrebovávaného  $i$ -teho vstupu.

V mojej práci sa zaoberám odhadom produkčných funkcií pre nemocnice a LDN. Konkrétne sa snažím o odhad Cobb-Douglasovej produkčnej funkcie, ktorá je najznámejším typom. (Pre podrobnejší výklad všeobecnej teórie produkčných funkcií vid' Soukupová, Horejší, Macáková, Soukup, 2004).

### Cobb-Douglasova produkčná funkcia

Tento typ produkčnej funkcie bol prvýkrát sformulovaný v roku 1928. Je pomenovaná podľa pánov, ktorí ju formulovali, tj. matematik Charles W. Cobb a ekonóm a neskorší senátor Paul H. Douglas. Douglas už ako študent pracoval pre Henryho L. Moorea. Jeho úlohou bolo odhadovanie a testovanie konceptu neoklasickej teórie. V rámci svojej ďalšej práce (pod vedením J.B. Clarka) sa zaoberal štatistickým testovaním teórie medznej produktivity. Výsledkom bolo následné formulovanie a matematické vyjadrenie najznámejšej produkčnej funkcie:

$$Q = f(L, K) = AL^a K^b, \quad a, b, A > 0, \quad (4.2)$$

kde  $Q$  je výstup (produkcia),  $K$  je množstvo použitého kapitálu,  $L$  množstvo spotrebovanej práce,  $a$ ,  $b$ ,  $A$  sú nenulové konštanty.

Znázornením dlhodobej produkčnej funkcie v trojdimenzionálnom priestore dostaneme tzv. produkčný kopec. Pre praktické znázornenie využívame premietnutie tohto 3D modelu do dvojdimenzionálneho priestoru, dostávame tak mapu izokvant. Izokvanta je krivka, ktorá je tvorená všetkými možnými kombináciami vstupu vedúcich k rovnakej produkcii výstupu.

### Výnosy z rozsahu

K produkčným funkciám v dlhom období (tj. v období, v ktorom sú všetky vstupy variabilné) sa vzťahuje pojem **výnosy z rozsahu** (výraz „výnosy z rozsahu“ nie je

ekvivalentom výrazu „výnosy z variabilného vstupu“, ktorý je typický pre produkčné funkcie v krátkom období). Všeobecne rozlišujeme tri druhy výnosov z rozsahu podľa toho, aký je vzťah medzi  $f(K,L)$  a po vynásobení jednotlivých vstupov rovnakou nenulovou konštantou „ $t$ “ ( $t < 0$ ),  $f(tK, tL)$ :

- konštantné:  $f(tK, tL) = tf(K,L)$
- rastúce:  $f(tK, tL) > tf(K,L)$
- klesajúce:  $f(tK, tL) < tf(K,L)$ .

Pre určenie charakteru výnosov z rozsahu pre Cobb-Douglasovu funkciu je potrebné poznať hodnotu jednotlivých konštánt. Po vynásobení vstupov nenulovou konštantou dostaneme:

$$(tK, tL) = A(tK)^a (tL)^b = At^{a+b} K^a L^b = t^{a+b} f(K,L). \quad (4.3)$$

Ak  $a + b = 1$  konštantné výnosy z rozsahu (zvýšenie vstupov o „ $t$ “ % vyvolá zvýšenie výstupu rovnako o „ $t$ “ %)

$a + b > 1$  rastúce výnosy z rozsahu (zvýšenie vstupov o „ $t$ “ % vyvolá zvýšenie výstupu o viac ako „ $t$ “ %)

$a + b < 1$  klesajúce výnosy z rozsahu (zvýšenie vstupov o „ $t$ “ % vyvolá zvýšenie výstupu o menej než „ $t$ “ %).

### Všeobecné vyjadrenie Cobb-Douglasovej produkčnej funkcie

Uvedené vyjadrenie, tak ako je vo forme (4.2), je všeobecné za predpokladu uvažovania 2 výrobných faktorov (práce a kapitálu). V prípade pripustenia existencie pôdy, resp. iného výrobného faktoru, sa len formálne zmení formulácia:

$$Q = f(L, K, T) = AL^a K^b T^c, \quad a, b, c, A > 0, \quad (4.4)$$

kde T je tretí výrobný faktor (pôda) a pre konštanty platí  $a + b + c = 1$ .

### Dôvody a dôsledky linearizácie Cobb-Douglasovej funkcie

Základná definícia Cobb-Douglasovej produkčnej funkcie, ako vidieť z (4.2, 4.4) je funkcia nelineárna v parametroch. Z dôvodu zložitého postupu pri odhadovaní parametrov funkcií v parametroch nelineárnych sa používa logaritmická transformácia funkcie:

$$\ln Q = \ln A + a \ln L + b \ln K. \quad (4.5)$$

Touto úpravou dostaneme funkciu lineárnu v parametroch. Neodhadujú sa teda parametre pôvodnej funkcie (4.2), ale funkcie, ktorá namiesto pôvodných dát obsahuje ich logaritmickú transformáciu. Linearita funkcie je pre odhady dôležitá, a to napr. z dôvodu následného použitia metódy najmenších štvorcov (MNS) pre odhad parametrov.

### Medzné produktivity

V súvislosti s produkčnými funkciami všeobecne sa definuje aj **medzný produkt** jednotlivých faktorov. Medzný produkt faktoru X ( $MP_X$ ) je prírastok produktu vyvolaný prírastkom faktoru o jednu jednotku za predpokladu nemennosti množstva iných faktorov.

Pre faktor kapitálu teda platí:

$$MP_K = \Delta Q / \Delta K, \quad (4.6)$$

kde  $MP_K$  je medzný produkt kapitálu,  $\Delta Q$  je prírastok produktu a  $\Delta K$  je prírastok kapitálu.

Úpravou dostaneme vzťah prírastku produktu a prírastok kapitálu pri nemennosti faktoru práce:

$$\Delta Q = MP_K \Delta K. \quad (4.7)$$

Analogicky sa definuje medzný produkt pre faktor práce ako prírastok produktu vyvolaný prírastkom práce o jednotku za nemennosti množstva kapitálu:

$$MP_L = \Delta Q / \Delta L, \quad (4.8)$$

kde  $MP_L$  je medzný produkt práce,  $\Delta Q$  je prírastok produktu a  $\Delta L$  je prírastok práce.

## 5 Analýzy a výsledky

### 5.1 DEA modely

Všetky výpočty ktoré boli v rámci tejto podkapitoly uskutočnené, boli vypočítané pomocou softwarového systému Frontier Analyst Professional, ktorý je priamo určený pre analýzu DEA modelov. Autorom tohto konkrétneho systému je firma Banxia Software ([www.banxia.com](http://www.banxia.com)).

#### 5.1.1 DEA modely pre nemocnice

Výsledky analýz jednotlivých DEA modelov (CCR a BCC, modely orientované na vstupy i výstupy) pre súbor nemocníc, sú uvedené v tabuľkách č. 2 a č. 4.

V prípade modelov orientovaných na výstupy (tabuľka č. 2) je ohodnotenie pre neefektívne jednotky väčšie ako 1, ale z dôvodu prehľadnosti bola táto hodnota transformovaná tak, aby miera efektívnosti neefektívnych jednotiek bola nižšia ako 100 % a pre jednotky efektívne bola rovná 100 %. Ako bolo spomenuté, efektívnosť v prípade CCR modelov je rovnaká bez ohľadu na orientáciu modelu, takže nie je zarážajúce, že pre oba typy modelu – pri orientácii na vstupy, tak i pri orientácii na výstupy - je v prípade CCR modelov efektívnosť jednotiek rovnaká. Rovnako sa potvrdil aj predpoklad, že efektívnosť v prípade variabilných výnosov z rozsahu (modely BCC) je vyššia oproti modelom CCR. Takisto sa v týchto modeloch zvýšil počet efektívnych jednotiek. Zatiaľ čo v prípade CCR modelov je efektívna len jedna jednotka, Nemocnice Písek, v prípade BCC modelov je už efektívnych celkovo 6 jednotiek. Ide o Nemocnicu Písek, Fakultní nemocnicu Na Bulovce, Nemocnice Jihlava, Nemocnice na Žižkově, Nemocnice Rokycany a Nemocnice Stodůlky.

Tabuľky sú rozdelené podľa orientácie modelov. Obe obsahujú údaje o hodnotení miery efektívnosti jednotiek, cieľové hodnoty ktorých by neefektívna jednotka mala dosiahnuť, aby dosiahla efektívnu hranicu (v prípade efektívnych jednotiek sa tieto hodnoty zhodujú so vstupnými údajmi z tabuľky č. 1) a vzory.

Konkrétne cieľové hodnoty zodpovedajú modelu s variabilnými výnosmi z výnosu (model BCC). Vzory uvedené v tabuľkách zodpovedajú, tak ako aj uvedené cieľové hodnoty, vzorom ktoré boli určené pre model s variabilnými výnosmi z rozsahu. V prípade CCR modelu by vzorom pre všetky neefektívne jednotky bola Nemocnica Písek, pretože je jedinou efektívnou jednotkou. Táto situácia ale nie je z interpretačného hľadiska zaujímavá. V prípade väčšieho počtu efektívnych jednotiek v modeli je typické, že pre neefektívne jednotky je určený väčší počet vzorov. Existuje však nekonečne mnoho kombinácií vstupov a výstupov, ktoré umožnia neefektívnej jednotke hranicu efektívnosti dosiahnuť. Uvedené cieľové hodnoty sú jedným z možných spôsobov, avšak určite nie jediným možným. Dá sa povedať, že ak pre neefektívnu jednotku existuje viacero vzorov, tak cieľové hodnoty tejto jednotky by mali byť lineárnou kombináciou vzorov.

**Tabuľka č. 2:** Modely orientované na vstupy, pre nemocnice

Č.	Nemocnice	CCR-I	BCC-I	počet lôžok	počet lekárov v lôž. časti	Počet hospitalizovaných	Vzory
1	Fakultní nemocnice Na Bulovce	78,70	<b>100,00</b>	1199	160,39	38697	1
2	Nemocnice Blansko	56,40	72,10	153	18,05	4925	9, 11
3	Nemocnice Český Krumlov	82,00	86,80	245	28,42	9482	11, 14
4	Nemocnice Domažlice	82,30	97,10	200	21,56	6950	9, 11, 14
5	Nemocnice Havlíčkův Brod	87,50	97,10	558	66,02	20637	1, 11
6	Nemocnice Chrudim	73,60	74,70	329	35,54	13270	11, 14
7	Nemocnice Jihlava	79,40	<b>100,00</b>	740	80,87	24090	7
8	Nemocnice Kroměříž	92,10	96,10	441	48,74	17329	1, 11
9	Nemocnice Na Žižkově	55,10	<b>100,00</b>	106	8,90	1919	9
10	Nemocnice Pelhřimov	85,00	87,10	306	33,59	12234	11, 14
11	Nemocnice Písek	<b>100,00</b>	<b>100,00</b>	382	40,04	15664	11
12	Nemocnice Prachatice	86,10	96,10	184	23,26	6740	11, 14
13	Nemocnice Prostějov	84,70	96,30	414	44,75	16566	1, 11
14	Nemocnice Rokycany	88,70	<b>100,00</b>	178	22,77	6477	14
15	Nemocnice Benešov	82,90	83,30	363	38,46	14823	11, 14
16	Nemocnice Stodůlky	90,60	<b>100,00</b>	212	19,40	6874	16
17	Nemocnice Tábor	73,20	75,60	427	46,62	16923	1, 11
18	Nemocnice Třebíč	77,80	87,50	566	65,63	20638	1, 7, 11
19	Nemocnice Vyškov	77,50	77,90	388	40,89	15826	1, 11
20	Nemocnice Znojmo	74,80	85,30	598	70,03	21498	1, 7, 11
21	Oblastní nemocnice Jičín	84,2	86,3	306	33,57	12224	11, 14
22	Okresní nemocnice J.Hradec	82,1	82,5	361	38,24	14704	11, 14

Pre konkrétnu interpretáciu som si zvolila Nemocnicu Prachatice z Jihočeského kraja. Táto nemocnica nie je efektívna ani v jednom z modelov. Podľa modelu CCR dosahuje 86,1%-nej efektívnosti, v modeli BCC-O 95,4 % a podľa modelu BCC-I efektívnosť o niečo

vyššiu, konkrétne 96,5 %. Ako už vieme, medzi veľké prednosti DEA analýzy patrí možnosť identifikácie zdroju neefektívnosti a efektívne vzory pre jednotku neefektívnu. Aby sa jednotka dostala na efektívnu hranicu a stala sa tak „100%-ne efektívnou“ je v prípade modelu BCC-I potrebná redukcia počtu lekárov v lôžkovej časti o 4,24 (z pôvodných 27,5 na 23,26), pri súčasnom znížení počtu lôžok z 191 na 184 a pri zachovaní počtu hospitalizovaných.

V modeli BCC-I by sa stala jednotka efektívnou v prípade navýšenia počtu hospitalizovaných z 6740 na 7062 pri nezmenenom počte lôžok a redukcii počtu lekárov v lôžkovej časti z 27,5 na 23,87. Za predpokladu konštantných výnosov z rozsahu, model CCR-I, by jednotka 100%-nej efektívnosti dosiahla znížením počtu lôžok na 164 a súčasným znížením počtu lekárov o 10,27 na 17,23. Je logické, že redukcia vstupov v CCR modeloch je „náročnejšia“ oproti modelom BCC, a to z dôvodu, že vzdialenosť jednotky od efektívnej hranice je v tomto modeli väčšia ako v modeloch BCC. Pre sprehľadnenie výsledkov a cieľových hodnôt pre dosiahnutie efektívnej hranice pre Nemocnicu Prachatice som údaje zhrnula do nasledujúcej Tabuľky č. 3.

**Tabuľka č. 3:** Pôvodné a cieľové hodnoty pre dosiahnutie efektívnosti v jednotlivých modeloch pre Nemocnicu Prachatice

Ukazovateľ	pôvodné hodnoty	hodnoty potrebné pre dosiahnutie efektívnosti		
		CCR-I	BCC-O	BCC-I
počet lôžok	191	164	184	bez zmeny
Počet hospitalizovaných	6740	bez zmeny	bez zmeny	7062
počet lekárov	27,5	17,23	23,26	23,87

Efektívnym vzorom pre Nemocnicu Prachatice je v CCR modeli Nemocnica Písek, čo je logické, pretože táto nemocnica je v tomto modeli jediná 100%-ne efektívna jednotka. V modeloch BCC je pre prachatickú nemocnicu vzorom okrem píseckej nemocnice aj Nemocnice Rokycany.

Tabuľka č. 4: Modely orientované na výstupy, pre nemocnice

Č.	Nemocnice	CCR-O	BCC-O	počet lôžok	počet lekárov v lôž. časti	počet hospitalizovaných	Vzory
1	Fakultní nemocnice Na Bulovce	78,70	<b>100,00</b>	1199	160,39	38697	1
2	Nemocnice Blansko	56,40	61,20	213	25,73	8053	11, 14
3	Nemocnice Český Krumlov	82,00	85,00	282	31,57	11161	11, 14
4	Nemocnice Domažlice	82,30	96,00	206	22,20	7240	9, 11, 14
5	Nemocnice Havlíčkův Brod	87,50	97,80	575	68,47	21105	1, 11
6	Nemocnice Chrudim	73,60	76,70	440	48,58	17299	1, 11
7	Nemocnice Jihlava	79,40	<b>100,00</b>	740	80,87	24090	7
8	Nemocnice Kroměříž	92,10	97,20	459	51,38	17835	1, 11
9	Nemocnice Na Žižkově	55,10	<b>100,00</b>	106	8,90	1919	9
10	Nemocnice Pelhřimov	85,00	85,70	351	37,42	14268	11, 14
11	Nemocnice Písek	<b>100,00</b>	<b>100,00</b>	382	40,04	15664	11
12	Nemocnice Prachatice	86,10	95,40	191	23,87	7062	11, 14
13	Nemocnice Prostějov	84,70	90,30	477	54,03	18342	1, 11
14	Nemocnice Rokycany	88,70	<b>100,00</b>	178	22,77	6477	14
15	Nemocnice Benešov	82,90	86,20	436	47,99	17186	1, 11
16	Nemocnice Stodůlky	90,60	<b>100,00</b>	212	19,40	6874	16
17	Nemocnice Tábor	73,20	81,40	564	66,85	20795	1, 11
18	Nemocnice Třebíč	77,80	91,40	647	75,05	22571	1, 7, 11
19	Nemocnice Vyškov	77,50	86,10	498	53,25	18390	11, 7
20	Nemocnice Znojmo	74,80	89,70	701	82,08	23964	1, 7, 11
21	Oblastní nemocnice Jičín	84,20	84,90	354	37,67	14403	11, 14
22	Okresní nemocnice J.Hradec	82,10	85,40	437	48,14	17215	1, 11

Za povšimnutie určite stojí aj výkonnosť, resp. miera efektívnosti fakultnej nemocnice. Fakultná nemocnica Na Bulovce dosiahla 100%-nej efektívnosti v prípade BCC modelu, zatiaľ čo v modeloch CCR je až ôsma v poradí. Jej postavenie v rámci BCC modelov je dané tým, že je to najväčšia nemocnica zo všetkých, ktoré boli do analýzy zahrnuté a ako taká by sa mohla považovať za tzv. odľahlé pozorovanie. Je logické, že táto jednotka bude v modeli s variabilnými výnosmi z rozsahu ležať na efektívnej hranici. Analýza obalu dát je metóda, ktorá sa snaží nájsť nejaký „obal dát“ a tak všetky jednotky určitým spôsobom „obaliť“. Obdobne je tomu tak aj v prípade opačného „odľahlého pozorovania“. Ide o nemocnicu, ktorá je v rámci analyzovaného súboru jednotiek najmenšia. Je ňou nemocnica na Žižkově. Tá sa stáva v modeli BCC jednou zo šiestich 100%-ne efektívnych, pričom pri uvažovaní konštantných výnosov z rozsahu je najneefektívnejšou jednotkou z celého súboru.



### 5.1.2 DEA modely pre LDN

Rovnaké analýzy, aké boli uskutočnené pre súbor nemocníc, som spravila pre súbor liečební dlhodobo chorých. Analýzy sú, obdobne ako pre nemocnice, rozdelené na modely orientované na vstupy (tabuľka č. 5) a modely orientované na výstupy (tabuľka č. 6). Rovnako, ako pri analýzach nemocníc, aj v prípade LDN sa nám potvrdili všetky základné teoretické predpoklady: efektívnosť jednotiek v CCR-I a efektívnosť v CCR-O je rovnaká, počet efektívnych jednotiek v modeloch s variabilnými výnosmi z rozsahu rastie a efektívnosť v modeloch BCC je vyššia ako v modeloch CCR.

**Tabuľka č. 5:** Modely orientované na vstupy, pre LDN

Č.	LDN	CCR-I	BCC-I	počet lôžok	počet lekárov v lôž. časti	počet hospitalizovaných	Vzory
1	LDN Horažďovice	40,23	45,02	63	2	293	2, 4, 8
2	LDN Hradec Králové	<b>100,00</b>	<b>100</b>	99	2	332	2
3	LDN Chitussiho 1	24,94	54,25	54	1,47	169	5, 8
4	LDN Jablonec nad Nisou	<b>100,00</b>	<b>100</b>	67	2,79	454	4
5	LDN Lomnice nad Popelkou	86,80	<b>100</b>	59	2,5	347	5
6	LDN na Vojkově, Tehovec	69,51	<b>100</b>	100	6,5	471	6
7	LDN Nejdek	32,52	60,4	54	1,49	173	5, 8
8	LDN Opočno	61,50	<b>100</b>	53	1,2	122	8

V prípade modelov s konštantnými výnosmi z rozsahu vyšli ako 100%-ne efektívne liečebne: liečebne v Hradci Králové a v Jablonci nad Nisou. Naopak, za predpokladu variabilných výnosov z rozsahu, vzrástol počet efektívnych jednotiek na 5 z celkového počtu 8. K jednotkám efektívnym z modelu CCR: LDN Hradec Králové a LDN Jablonec nad Nisou, sa pridáva LDN Lomnice nad Popelkou, LDN na Vojkově a LDN Opočno. Vzory uvedené v oboch tabuľkách sú opäť vzory určené pre model s variabilnými výnosmi z rozsahu. Je zaujímavé, že pre modely orientované na vstupy sú vzormi pre neefektívne jednotky, vo všetkých prípadoch, LDN Jablonec nad Nisou a LDN na Vojkově.

**Tabuľka č. 6:** Modely orientované na výstupy, pre LDN

Č.	LDN	CCR-O	BCC-O	počet lôžok	počet lekárov v lôž.časti	počet hospitalizovaných	Vzory
1	LDN Horažďovice	40,23	63,47	82	4,45	462	4, 6
2	LDN Hradec Králové	<b>100</b>	<b>100,00</b>	99	2,00	332	2
3	LDN Chitussiho 1	24,94	36,71	79	4,18	460	4, 6
4	LDN Jablonec nad Nisou	<b>100</b>	<b>100,00</b>	67	2,79	454	4
5	LDN Lomnice nad Popelkou	86,8	<b>100,00</b>	59	2,50	347	5
6	LDN na Vojkově, Tehovec	69,51	<b>100,00</b>	100	6,50	471	6
7	LDN Nejdeč	32,52	37,93	71	3,26	456	4, 6
8	LDN Opočno	61,5	<b>100,00</b>	53	1,20	122	8

Pre interpretáciu som si vybrala LDN Horažďovice, z Klatovského okresu v Plzeňskom kraji. Táto liečebňa nie je efektívna ani v jednom z modelov. V modeli CCR-I (a ako už vieme aj v modeli CCR-O) je jednotka efektívna na 40,23 %. Pre dosiahnutie hranice efektívnosti by mala jednotka pri zachovaní počtu pacientov znížiť počet lôžok o 60 % na konečných 56. Rovnako by malo dôjsť k redukcii počtu lekárov na 1,79. Vzorom v tomto modeli pre túto jednotku sú LDN v Jablonci nad Nisou a v Hradci Králové. V modeli orientovanom na výstupy jednotka dosahuje pre model BCC-O efektívnosti 63,47 % a je jednou z troch neefektívnych liečební, pričom z týchto neefektívnych jednotiek vychádza „najlepšie“. Vzorom pre túto jednotku sú LDN na Vojkově a LDN v Jablonci nad Nisou. Tento model navrhol najmenšiu redukcii lôžok na konečných 82 pri súčasnom výraznom navýšení počtu hospitalizovaných pacientov o viac ako 1,5-krát na 462, pričom počet lekárov zostáva bezo zmeny. V BCC modeli orientovanom na vstupy (BCC-I) došlo takisto k redukcii počtu lôžok. Celková redukcia je viac ako 50%-ná, na konečných 63 lôžok pri nezmenenom počte hospitalizovaných. Tento model súčasne navrhuje zníženie počtu lekárov na dvoch. Vzorom v tomto prípade je LDN v Hradci Králové, LDN v Jablonci nad Nisou a LDN Opočno. Neefektívnosť tejto jednotky nie je žiadnym prekvapením. Je to liečebňa s najväčším počtom lôžok z celého analyzovaného súboru, ale čo sa počtu hospitalizovaných pacientov týka, je až štvrtá v poradí. Preto jednotlivé modely tak výrazne redukujú počet lôžok pri zachovaní počtu hospitalizovaných. tabuľka č. 7 prehľadne sumarizuje mieru redukcie, resp. navýšenia jednotlivých vstupov a výstupov pre použité modely.

**Tabuľka č. 7:** Pôvodné a cieľové hodnoty pre dosiahnutie efektívnosti v jednotlivých modeloch pre LDN Horažďovice

Ukazovateľ	pôvodné hodnoty	hodnoty potrebné pre dosiahnutie efektívnosti		
		CCR-I	BCC-I	BCC-O
počet lôžok	140	56	63	82
počet hospitalizovaných	293	bez zmeny	bez zmeny	462
počet lekárov	4,45	1,79	2,00	bez zmeny

Pre ďalšie analýzy a nájdenie vysvetlenia neefektívnosti by bolo určite vhodné porovnať prípadový mix konkrétnych liečební. Za povšimnutie stojí, že LDN v Jablonci je vzorom pre túto jednotku vo všetkých typoch modelov a rovnako je vzorom aj pre ostatné neefektívne jednotky. Opäť pripomínam, že existuje možnosť zoradenia efektívnych jednotiek pomocou modelov super efektívnosti, ktorú ale v tejto práci neuskutočňujem.

## 5.2 Odhad Cobb-Douglasovej produkčnej funkcie

Tak ako pri aplikácii analýzy obalu dát, aj v prípade odhadu produkčnej funkcie rozdeľujem pôvodný súbor na samostatný súbor pre nemocnice a súbor, v ktorom sú zahrnuté len liečebne. Odhad konštánt ( $A$ ,  $a$ ,  $b$ ) Cobb-Douglasovej produkčnej funkcie je, ako už bolo názorne uvedené v kapitole číslo 4, primárne nelineárna funkcia. Napriek faktu, že Cobb-Douglasova produkčná funkcia vo svojej základnej definícii je funkcia nelineárna, môžeme pre odhad hodnôt konštánt použiť lineárny regresný model. Pomocou logaritmickej transformácie a matematických úprav je možné primárne nelineárnu úlohu previesť na úlohu lineárnu (spôsob a matematický postup bol názorne uvedený v kapitole štyri – Ekonometrická analýza produkčných funkcií). Vysvetľovanou premennou sa tak stáva logaritmus počtu hospitalizovaných (v ďalšom texte už len premenná  $Y$ ) a vysvetľujúcimi premennými sú logaritmus počtu lekárov v lôžkovej časti nemocnice (premenná  $X_1$ ) a logaritmus počtu lôžok (premenná  $X_2$ ). Tabuľka č. 8 obsahuje hodnoty logaritmov premenných použitých pre ďalšie analýzy, opäť agregovane, pre nemocnice i LDN. Pôvodné hodnoty sú uvedené v tabuľke č. 1. Hodnotenie kvality modelu a jednotlivých výsledkov je veľmi zložitá. Neexistuje žiadne jednoznačné kritérium, pomocou ktorého by sa dalo povedať „Áno, model nám vyšiel správne a je to ten *pravý*, ktorý vystihuje a popisuje skúmaný jav“. Kritérií je mnoho a nie je zriedkavé, že výsledky vychádzajú protikladne.

**Tabuľka č. 8:** Pôvodné a upravené dáta pre nemocnice a LDN

Č.	Nemocnice	ln(Y)	Ln(X1)	ln(X2)
		ln(počet hosp.)	ln(počet lekárov v lôž.časti)	ln(počet lôžok)
1	Nemocnice na Žižkově	7,55956	2,18605	4,66344
2	Nemocnice Stodůlky	8,83550	2,96527	5,35659
3	Nemocnice Domažlice	8,84650	3,10009	5,32788
4	Nemocnice Rokycany	8,77601	3,12544	5,18178
5	Nemocnica Prachatice	8,81582	3,31419	5,25227
6	Nemocnice Český Krumlov	9,15715	3,55220	5,64191
7	Nemocnice Písek	9,65912	3,68988	5,94542
8	Nemocnice Pelhřimov	9,41197	3,85630	5,86079
9	Nemocnice Benešov	9,60394	3,91402	6,07764
10	Nemocnice Kroměříž	9,76014	3,93944	6,12905
11	Oblastní nemocnice Jičín	9,41116	3,94700	5,86930
12	Nemocnice Chrudim	9,49326	3,96348	6,08677
13	Nemocnice Vyškov	9,66941	3,97500	6,21060
14	Nemocnice Prostějov	9,71511	4,09983	6,16752
15	Nemocnice Tábor	9,73643	4,21509	6,33505
16	Nemocnice Blansko	8,50208	4,30784	5,36129
17	Nemocnice Havlíčkův Brod	9,93484	4,31735	6,35437
18	Nemocnice Třebíč	9,93489	4,31815	6,47235
19	Okresní nemocnice J.Hradec	9,59587	4,33755	6,07993
20	Nemocnice Jihlava	10,08955	4,39284	6,60665
21	Nemocnice Znojmo	9,97572	4,40769	6,55251
22	Fakultní nemocnice Na Bulovce	10,56352	5,07761	7,08924
23	LDN Opočno	4,80402	0,18232	3,97029
24	LDN Hradec Králové	5,80513	0,69315	4,59512
25	LDN Lomnice nad Popelkou	5,84932	0,91629	4,07754
26	LDN Jablonec nad Nisou	6,11810	1,02604	4,20469
27	LDN Nejedek	5,15329	1,18173	4,49981
28	LDN Chitussiho 1	5,12990	1,43031	4,60517
29	LDN Horažďovice	5,68017	1,49290	4,94164
30	LDN na Vojkově, Tehovec	6,15486	1,87180	4,60517

### 5.2.1 Odhad Cobb-Douglasovej produkčnej funkcie pre nemocnice

Pomocou štatistického softwaru: Statgraphics Plus for Windows 3.11, GiveWin, SPSS a niektorých dopočtov v tabuľkovom kalkulátore MS Excel, som odhadovala hodnoty jednotlivých parametrov modelu, ako aj určovala hodnoty testových kritérií a k nim prislúchajúce p-value<sup>1</sup> pre *t-testy*, celkový *F test* a ďalšie charakteristiky potrebné k hodnoteniu a tvorbe záverov o vhodnosti modelu ako celku. Všetky hypotézy v ďalších analýzach sú testované na 5-%nej hladine významnosti ( $\alpha = 0,05$ ). Pre analýzy nebol použitý len jeden software, a to z dôvodu, že Statgraphics Plus for Windows 3.11 neposkytuje možnosť uloženia napr. Jackknife reziduí. Preto bol pre odhad charakteristík pri analýze odľahlých pozorovaní použitý SPSS v kombinácii so Statgraphicsom. GiveWin bol použitý pri testovaní charakteristík klasických reziduí z dôvodu, že tieto testy nie sú v Statgraphicse možné, resp. sú komplikované a možnosť týchto testov je v rámci GiveWinu užívateľsky prístupnejšia. Presné definície nulových hypotéz jednotlivých testov, spôsob konštrukcie a pravdepodobnostné rozdelenia testových kritérií, sú uvedené v každej štatisticky alebo ekonometricky zameranej knihe, takže ich presné definície nebudem uvádzať.

Do pôvodného modelu boli zahrnuté obe premenné, počet lekárov (premenná X1) i počet lôžok (premenná X2). Hodnoty odhadnutých parametrov premenných výsledného modelu sú po spätnej transformácii logaritmicke upravených premenných odhadnuté hodnoty konštánt z rovnice 4.2 resp. 4.5. Všetky potrebné informácie a komentované hodnoty sú postupne uvedené vo výstupoch a tabuľkách. Odhad základného (primárneho) modelu je uvedený vo výstupe č. 1. Ďalej sa budem venovať komentárom jednotlivých parametrov a charakteristík, poprípade ďalším analýzám, ktoré budú potrebné na potvrdenie, resp. vyvrátenie určitých domnienok.

#### **Interpretácia a komentár charakteristík modelu**

Celkovým F testom<sup>2</sup> bola zamietnutá nulová hypotéza o „nehodnosti“ modelu ako celku. Avšak na konečné závery o „vhodnosti“ modelu je ešte príliš skoro. Nesmieme

---

<sup>1</sup> P-value je kritická hodnota, ktorá sa porovnáva s hladinou významnosti  $\alpha$ . V prípade že p-value je väčšia ako  $\alpha$ , znamená to, že na  $\alpha$  hladine významnosti sme nepotvrdila nulovú hypotézu. Tvar nulovej hypotézy je daný druhom testu. V mojich analýzach je  $\alpha = 0,05$  (tj. 5%).

<sup>2</sup> Pomocou F testu testujeme hypotézu o nulových hodnotách všetkých regresných koeficientov, tj.  $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ , kde  $\beta_i$  sú jednotlivé regresné koeficienty.

zabudnúť na výsledky jednotlivých t-testov<sup>3</sup> a ďalšie charakteristiky. Rovnako je dôležité zistiť, či sú splnené predpoklady napr. o vlastnostiach reziduálnej zložky atď.

### ***t-testy, $R^2$ , upravený $R^2$***

V odhadnutom modeli, vyšla odhadnutá hodnota parametru pre premennú X1 (počet lekárov) záporne. Táto skutočnosť je v rozpore s teoretickými predpokladmi o hodnotách jednotlivých konštánt Cobb-Douglasovej produkčnej funkcie, tak ako boli uvedené v teoretickej časti. Výsledok t-testu naznačuje štatistickú nevýznamnosť danej premennej v modeli (hodnota testového kritéria je nízka a hodnota p-value pre tento test a túto premennú je vysoká). Zvyšné premenné modelu, konštanta a premenná X2 (počet lôžok), spĺňajú podmienky kladené teóriou a aj podľa štatistických kritérií vychádzajú významne: hodnoty testových kritérií pre t-testy poukazujú na ich významnosť a teda opodstatnenosť v modeli.

Za povšimnutie stojí hodnota kritéria  $R^2$ , tzv. index determinácie<sup>4</sup>. Jeho hodnota je vypočítaná ako pomer variability, ktorú dokážeme modelom popísať  $Q(\hat{y})$  k variabilite celkovej  $Q(y)$ . Jeho hodnota je 94,3226 %. Jeho hodnota je vysoká a značí dobrú zhodu modelu s dátami. Môžeme povedať, že tento model, s pomocou oboch vysvetľujúcich premenných X1, X2, vysvetľuje približne 94,32 % celkovej variability dát. Avšak tento záver nie je konečný, vzhľadom k výsledku t-testu pre premennú X1 sa budeme ďalej zaoberať a skúmať jej významnosť pre model. Ekvivalentnou charakteristikou je hodnota indexu korelácie ako charakteristiky sily lineárnej závislosti. Jej hodnota je 97,1198 %. Takto vysoká hodnota ale nie je zarážajúca, pretože korelačný pomer je druhou odmocninou pomeru determinácie ( $P = (R^2)^{1/2}$ ). Pre porovnanie s hodnotou  $R^2$ , upravený  $R^2$  má hodnotu 93,725 %. Táto charakteristika je o niečo „presnejšia“ a to z toho dôvodu, že upravuje pôvodný  $R^2$  a zohľadňuje počet parametrov modelu a následne tak umožňuje porovnávanie modelov s rôznym počtom parametrov. Ako vidíme, ich hodnoty sa líšia len minimálne, čo je ale spôsobené tým, že model obsahuje len dve vysvetľujúce premenné, takže počet parametrov sa výrazne nepremieta do hodnoty  $R^2$ .

<sup>3</sup> Nulová hypotéza má pre t-test tvar:  $H_0: \beta_j = c_j, j=0, \dots, K$ , K je počet parametrov modelu a  $c_j$  je ľubovoľne zvolené číslo. Význam má testovanie hypotézy:  $H_0: \beta_j = 0$ . Zamietnutím nulovej hypotézy sme štatisticky potvrdili oprávnenosť j-tej vysvetľujúcej premennej v regresnej rovnici.

<sup>4</sup> Index determinácie je v podstate viacnásobný korelačný koeficient  $r(y, x_1, x_2, \dots, x_k)$ . Základnou myšlienkou výpočtu jeho hodnoty je rozklad celkovej variability  $Q(y)$  na variabilitu vysvetlenú modelom  $Q(\hat{y})$  a variabilitu modelom nevysvetlenú  $Q(e)$ . Matematicky zapísané:  $Q(y) = Q(\hat{y}) + Q(e)$ .

**Výstup č. 1: Regresná analýza pre nemocnice, zahrnutie oboch premenných**

Multiple Regression Analysis

Dependent variable: hospitalizovany

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2,67	0,454679	5,87227	0,0000
X1	-0,071382	0,128742	-0,554456	0,5857
X2	1,18184	0,142061	8,31925	0,0000

## Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	8,40322	2	4,20161	157,83	0,0000
Residual	0,5058	19	0,026621		
Total (Corr.)	8,90902	21			

R-squared = 94,3226 percent

R-squared (adjusted for d.f.) = 93,725 percent

Standard Error of Est. = 0,16316

Mean absolute error = 0,12033

Durbin-Watson statistic = 1,07061

**Korelačná analýza, multikolarita**

Vzhľadom k predchádzajúcim výsledkom som sa rozhodla podrobnejšie sa zaoberať vzájomným vzťahom oboch vysvetľujúcich premenných X1, X2. Všebečným problémom ekonometrických analýz (a nielen ekonometrických) je problém multikolarity. V štatistickom kontexte je multikolarita definovaná ako „silná vzájomná lineárna závislosť vysvetľujúcich premenných, ktorej dôsledkom je nedostatočne podmienená matica  $\mathbf{X}^{5c}$  (Hebák, Hustopecský, Malá, 2005). Z definície je jasné, že problém multikolarity je zviazaný s maticou  $\mathbf{X}$  a nie s typom regresného modelu. Preto sa multikolarita označuje ako problém dát.

Dôsledkom multikolarity sú napr. nepresné a nestabilné odhady, nahodnotenie súčtu štvorcov regresných koeficientov (a na ich základe sa niektoré premenné môžu zdať významnejšie ako v skutočnosti sú). Multikolarita komplikuje rozumnú interpretáciu výsledkov modelu. Existuje mnoho charakteristík a kritérií, pomocou ktorých sa multikolarita dá identifikovať. V rámci tejto práce sa jej identifikáciou zaoberať nebudem, a to z dôvodu, že v prípade dvoch vysvetľujúcich premenných hovoriť o multikolarite nemá zmysel. V prípade dvoch premenných, je dostatočné hovoriť o vzájomnej lineárnej závislosti vysvetľujúcich premenných. Ako sa uvádza (Hušek, 1999), multikolarita, resp.

<sup>5</sup> Nedostatočná podmienenosť matice  $\mathbf{X}$  znamená, že determinant matice  $\mathbf{X}$ ,  $|\mathbf{X}|$ , je blízky nule.



vzájomná lineárna závislosť premenných, sa považuje za neúnosnú v prípade, že niektorý z jednoduchých korelačných koeficientov je v absolútnej hodnote väčší ako 0,8, resp. 0,9. Ako je to v prípade vysvetľujúcich premenných analyzovaného modelu?

Domnievam sa, že každý, i ten kto nevie nič o štatistike, bude intuitívne predpokladať existenciu určitého vzťahu medzi počtom lôžok (premenná  $X_2$ ) a počtom lekárov (premenná  $X_1$ ). V štatistickej terminológii by sme tento vzťah označili ako závislosť premenných. Konkrétnejšie ide o vzájomnú koreláciu premenných. Jedným z dôvodov korelácie je fakt, že vysvetľujúce premenné sa vzťahujú k tej istej vysvetľovanej premennej. V ekonomickej teórii ide o komplementaritu vstupov (možnosť ich vzájomnej substitúcie).

Rozlišujeme niekoľko typov korelačných koeficientov:

- *jednoduché (párové) korelačné koeficienty*  $r_{yxi}$  (merajú intenzitu lineárnej závislosti konkrétnych dvoch premenných pri neuvažovaní vplyvu všetkých ostatných veličín).
- *dielčie koeficienty korelácie*  $r_{yx1, x2...xp}$  (merajú intenzitu lineárnej závislosti medzi jednou premennou a konkrétnych dvoch premenných za predpokladu, že ostatné premenné, tie ktoré sú uvedené za bodkou, sú konštantné).
- *viacnásobný (mnohonásobný) korelačný koeficient*  $r_{y, x1x2xp}$  (meria tesnosť lineárnej závislosti vysvetľovanej premennej  $Y$  na všetkých vysvetľujúcich premenných).

V nasledujúcej tabuľke uvádzam hodnoty jednoduchých a dielčích korelačných koeficientov. V poslednom riadku tabuľky je uvedená i hodnota viacnásobného korelačného koeficientu (ktorého hodnota zodpovedá hodnote indexu determinácie).

**Tabuľka č. 9:** Korelačné koeficienty pre nemocnice, model s oboma vysvetľujúcim premennými, 22 pozorovaní

$\Gamma_{yx1}$	0,8581
$\Gamma_{yx2}$	0,9707
$\Gamma_{x1x2}$	0,8978
$\Gamma_{yx1.x2}$	-0,1264
$\Gamma_{yx2.x1}$	0,8857
$\Gamma_{y.x1x2}$	0,9433

Je evidentné, že intuitívny predpoklad o lineárnej závislosti oboch vysvetľujúcich premenných sa potvrdil. Rovnako sa potvrdil aj predpoklad o lineárnej závislosti jednotlivých premenných s vysvetľovanou premennou Y. Všetky jednoduché korelačné koeficienty sú väčšie ako kritická hodnota 0,8. Táto vzájomná lineárna závislosť premenných nie je žiaduca a budem sa s ňou musieť nejakým spôsobom vyrovnat'. Možno sa v analyzovanom súbore nachádzajú niektoré odľahlé pozorovania, ktorých odstránením by došlo k poklesu hodnôt jednoduchých korelačných koeficientov.

### ***Odľahlé a vplyvné pozorovania***

Dôležitým kritériom pri tvorbe a formulácii záverov o vhodnosti modelu, je analýza odľahlých a vplyvných pozorovaní. Za odľahlé pozorovanie sa považuje pozorovanie, ktorého hodnoty premennej Y, sa zásadne líšia od ostatných hodnôt vysvetľovanej premennej Y (vybočujúce odľahlé pozorovanie), alebo také pozorovanie, ktorého vektor  $\mathbf{x}$  (z matice X, vysvetľujúcich premenných) predstavuje takú kombináciu hodnôt vysvetľujúcich premenných, ktoré sa značne líšia od ostatných (extrémne odľahlé pozorovanie). V tabuľke č. 10 sú uvedené hodnoty klasických reziduí, reziduá štandardizované, ich vzájomný rozdiel, Jackknife reziduá a v poslednom stĺpci tabuľky i prvok projekčnej matice  $\mathbf{H}$ ,  $h_{ii}$ . Vplyvným pozorovaním sa považuje pozorovanie, pri vynechaní ktorého dochádza k zásadným zmenám regresných charakteristík.

*Klasické reziduá*, predstavujú odhad náhodnej zložky a sú rozdielom hodnoty pôvodnej, a hodnoty vyrovnanej - získanej odhadom modelu ( $e_i = y_i - \hat{y}_i$ ). Predpoklady platné pre klasické reziduá vypočítané zo správneho modelu sú vševšeobecne známe a ich overením sa budem zaoberať ďalej.

*Predikované reziduá*. Je to reziduum, ktoré je vypočítané ako rozdiel skutočnej hodnoty  $y_i$  a odhadnutej vyrovnanej hodnoty  $\hat{y}_{i(-i)}$ , ktorú získame na základe n-1 pozorovaní pri

vynechaní i-teho pozorovania, resp. hodnotu predikovaného rezidua je možné vypočítať aj na základe vzťahu k projekčnému prvku  $h_{ii}$ . V matematickom vyjadrení teda platí:

$e_{i(-i)} = y_i - \hat{y}_{i(-i)}$ , resp.  $e_{i(-i)} = e_i/(1-h_{ii})$ . Veľký rozdiel medzi klasickými a predikovanými reziduami nastáva pri vysokej hodnote projekčného prvku  $h_{ii}$  alebo pri výskyte odľahlého pozorovania.

*Jackknife reziduá*, ako ďalšia z charakteristík slúžiacich k odhaleniu vplyvných pozorovaní. Hodnota veličiny, ktorú vráti software, sa porovnávajú s kvantilom Studentovho rozdelenia t.j. s  $t_{1-\alpha/2}$  (n-p-1) stupňami voľnosti. Hodnota určená pomocou softwaru sa berie v absolútnej hodnote a následne sa porovnáva s kvantilom (t.j. s kvantilom  $t_{0,975}(22-3-1)=t_{0,975}(18)=2,445006$  pre primárny model).

*Projekčný prvok matice* - prvok  $h_{ii}$ , meria vzdialenosť i-teho pozorovania od centra ostatných bodov z hľadiska bodov vysvetľujúcich premenných. Prvok  $h_{ii}$ , sa k identifikácii používa z dôvodu, že každá vyrovnaná hodnota  $\hat{y}_i$  je lineárnou kombináciou  $y_i$ , pri váhach  $h_{ii}$  (dôkaz je uvedený v Hebák, Hustopecký, Malá, 2005). Takže platí, že čím je hodnota prvku  $h_{ii}$  väčšia, tým je váha daného pozorovania vzhľadom k vyrovnaným hodnotám väčšia. Hodnota prvku sa porovnáva s hodnotou  $(2p/n)$ , čo v našom modeli predstavuje hodnotu 0,27272. (Hodnoty prvkov  $h_{ii}$  sú uvedené tak, ako ich počíta Statgraphics, takže sú to prvky necentované.)

Použitie reziduí k identifikácii vplyvných pozorovaní má nevýhodu v tom, že niektoré pozorovania maskujú samé seba tým, že prispôsobujú odhady parametrov regresnej funkcie.

Ako vidieť z tabuľky (hodnoty poukazujúce na možnosť existencie vplyvného pozorovania, sú v tabuľke č. 10 uvedené tučne), za vplyvné pozorovania boli určené pozorovania č. 1 a č. 16.

**Tabuľka č. 10:** Analýza reziduí a vplyvných pozorovaní pre súbor

nemocníc, model s oboma vysvetľujúcimi premennými

Č.	Klasické reziduá $e_i$ (1)	Predikované reziduá $e_{i(-i)}$ (2)	Rozdiel $ (1)-(2) $	Jackknife reziduá	Leverage $h_{ii}$
1	-0,46585	-0,65074	<b>0,18489</b>	<b>-2,85518</b>	<b>0,28413</b>
2	0,04652	0,05140	0,00489	0,28510	0,09506
3	0,10107	0,11231	0,01124	0,61947	0,10009
4	0,20506	0,23554	0,03048	1,25680	0,12942
5	0,17503	0,19766	0,02263	1,07277	0,11448
6	0,07286	0,07737	0,00451	0,44655	0,05829
7	0,22596	0,23672	0,01076	1,38487	0,04546
8	0,09070	0,09511	0,00441	0,55592	0,04632
9	0,03051	0,03206	0,00155	0,18700	0,04835
10	0,12777	0,13461	0,00685	0,78308	0,05086
11	0,08631	0,09048	0,00417	0,52899	0,04614
12	-0,08743	-0,09191	0,00448	-0,53585	0,04874
13	-0,05680	-0,06020	0,00340	-0,34815	0,05643
14	0,04872	0,05146	0,00274	0,29860	0,05324
15	-0,11973	-0,12856	0,00883	-0,73381	0,06871
16	-0,19663	-0,21709	0,02046	-1,20511	0,09426
17	0,06315	0,06798	0,00483	0,38704	0,07103
18	-0,07618	-0,08349	0,00731	-0,46689	0,08754
19	0,04997	0,05251	0,00254	0,30624	0,04844
20	-0,07491	-0,08429	0,00938	-0,45911	0,11133
21	-0,12369	-0,13761	0,01391	-0,75811	0,10110
22	-0,12241	-0,16118	0,03878	-0,75022	<b>0,24058</b>

Žiadna charakteristika ale nie je univerzálna a 100%-ne platná. Preto tabuľke č. 11 sú uvedené hodnoty ďalších charakteristík, ktoré by mali pomôcť pri identifikácii vplyvných pozorovaní: Mahalanobisova vzdialenosť, Cookova vzdialenosť a veličinu DFFIT štandardizovaný. Tabuľka obsahuje hodnoty charakteristík, ktoré boli vypočítané pomocou programu SPSS. Identifikované, vplyvné pozorovania sú v tabuľke zvýraznené tučne.

**Tabuľka č. 11:** Ďalšie charakteristiky slúžiace k analýze vplyvných pozorovaní, model s oboma vysvetľujúcimi premennými

Č.	Mahalanobisova vzdialenosť	Prepočet pre Mahalanobisovu vzdialenosť	Cookova vzdialenosť	Covratio	Štandardizovaný DFFIT <sup>(-i)</sup>
1	7,27102	2,09751	<b>2,87639</b>	<b>0,04947</b>	<b>-5,26713</b>
2	2,40673	0,69428	0,00615	1,37893	0,13252
3	1,48135	0,42733	0,01899	1,24130	0,23500
4	1,76460	0,50904	0,08996	0,99975	0,53168
5	1,66771	0,48109	0,06255	1,08373	0,43702
6	0,27426	0,07912	0,00439	1,20788	0,11230
7	0,43246	0,12475	0,04841	0,89351	0,39275
8	0,07935	0,02289	0,00561	1,17459	0,12737
9	0,16356	0,04718	0,00069	1,23501	0,04440
10	0,28392	0,08190	0,01361	1,12557	0,20016
11	0,31126	0,08979	0,00637	1,19361	0,13559
12	0,09937	0,02867	0,00532	1,18007	-0,12401
13	0,56279	0,16235	0,00339	1,24172	-0,09852
14	0,16443	0,04743	0,00177	1,22391	0,07104
15	0,51213	0,14774	0,01449	1,15232	-0,20609
16	14,48982	<b>4,17995</b>	<b>5,08694</b>	<b>1,59835</b>	<b>-4,50911</b>
17	0,55846	0,16110	0,00418	1,23538	0,10943
18	0,95852	0,27651	0,00801	1,24559	-0,15187
19	1,50904	0,43532	0,00471	1,31018	0,11599
20	1,61784	0,46671	0,01118	1,29008	-0,17937
21	1,22457	0,35326	0,02475	1,18383	-0,26981
22	4,16680	1,20202	0,08003	1,37968	-0,48653

*Mahalanobisova vzdialenosť* je jednou z ďalších charakteristík slúžiacich k analýze a identifikácii vplyvných pozorovaní. Hodnoty vypočítané štatistickým programom sa musia ešte prepočítať podľa vzorca (5.1.1) a následne sa otestujú.

$$F = \frac{(n-p)n}{(n^2-1)p} D_i^2 \quad \text{má rozdelenie } F(p, n-p), \quad (5.1.1)$$

kde  $n$  je počet pozorovaní,  $p$  je počet parametrov,  $D_i^2$  je hodnota charakteristiky určená pomocou softwaru. Pre primárny model sa hodnoty  $D_i^2$  pre násobili konštantou 0,288 (pre násobené hodnoty sú uvedené v druhom stĺpci tabuľky) a následne sa porovnali s kvantilom Fisherovho rozdelenia:  $f_{0,95}(p, n-p)$ , tj. s  $f_{0,95}(3, 22-3)=3,127$ .

*Cookova vzdialenosť*. Jej hodnota kombinuje projekčný prvok  $h_{ii}$  a hodnoty štandardizovaných reziduí. Táto charakteristika indikuje vplyvné pozorovanie v prípade, že je jej hodnota väčšia ako 1.

$Covratio_{(-i)}$  je charakteristika, ktorá vyjadruje dôsledky vypustenia  $i$ -teho pozorovania na kovariančnú maticu.  $Covratio$  má pre vplyvné pozorovania hodnoty väčšie než  $1 + (3p/n)$ , resp. hodnoty menšie ako  $1 - (3p/n)$ .

Ako posledná charakteristika použitá pre identifikáciu vplyvných pozorovaní, bola použitá charakteristika *standardizovaný DFFIT*,  $DFFIT_{(-i)}$ . Táto veličina vyjadruje relatívny rozdiel  $i$ -tej vyrovnanej hodnoty pri vynechaní  $i$ -teho bodu. V prípade vplyvného pozorovania veličina má hodnoty väčšie ako  $2\sqrt{(p/n)}$ .

Nasledujúca tabuľka č. 12 prehľadne zobrazuje kritické hodnoty použitých charakteristík použitých k identifikácii vplyvných pozorovaní. Ku konkrétnym kritickým hodnotám sú uvedené aj čísla pozorovaní, ktoré boli podľa konkrétnych charakteristík označené ako pozorovania vplyvné.

**Tabuľka č. 12:** Hodnoty charakteristík použitých k identifikácii vplyvných pozorovaní a identifikované vplyvné pozorovania, model s oboma vysvetľujúcimi premennými

Charakteristika	kritické hodnoty	Identifikované vplyvné pozorovania
Mahalanobisova vzdialenosť (po prepočte)	$D_i^2 > 3,127$	č. 16
Cookova vzdialenosť	$C_i > 1$	č. 1, č. 16
$Covratio_{(-i)}$	$covratio_{(-i)} > 1,049$ $covratio_{(-i)} < 0,591$	č. 1, č. 16
$DFFIT_{(-i)}$	$ DFFIT_{(-i)}  > 0,736$	č. 1, č. 16

Ako vidieť z tabuľky č. 12, ako vplyvné pozorovanie boli vo všetkých prípadoch označené pozorovanie č. 16. Pozorovanie č. 1 bolo taktiež identifikované podľa všetkých charakteristík s výnimkou Mahalanobisovej vzdialenosti. Na základe vypočítaných charakteristík (tabuľka č. 11 a č. 12) môžeme považovať za odľahlé pozorovanie č. 1 a pozorovanie č. 16. Ide o Nemocnicu na Žižkove a Nemocnicu Blansko. Ako sme už pri analýze v rámci DEA modelov uviedli Nemocnica na Žižkove sa za odľahlé, resp. vplyvné pozorovania dá považovať. Dôvodom je fakt, že ide o najmenšiu nemocnicu (Nemocnica na Žižkove) v rámci analyzovaného súboru. Neskôr sa zameriam na zmeny regresných charakteristík po vylúčení pozorovaní č. 1 a č. 16. Predtým sa pozriem na výsledky Durbinovho-Watsonovho testu.

**Durbinov-Watsonov test**

Pomocou Durbinovho-Watsonovho<sup>6</sup> testu testujeme hypotézu o neexistencii autokorelácie reziduí. Testové kritérium nadobúda hodnôt z intervalu  $\langle 0, 4 \rangle$ . Tento interval je na základe znalostí kritických medzí ( $d_d$  – dolná medza testu,  $d_h$  – horná medza testu) rozdelený na podintervaly:

- $\langle -2, 4-d_h \rangle$  resp.  $\langle d_h, 2 \rangle$ , v ktorých testovaná  $H_0$  o nekorelovanosti reziduí platí,
- $\langle d_d, d_h \rangle$  resp.  $\langle 4-d_h, 4-d_d \rangle$ , v ktorých nedokážeme o platnosti  $H_0$  rozhodnúť,
- $\langle 0, d_d \rangle$  resp.  $\langle 4-d_d, 4 \rangle$ , v ktorých  $H_0$  o nekorelovanosti reziduí neplatí.

Vybrané kritické hodnoty pre tento test potrebné k rozhodnutiu o korelovanosti či nekorelovanosti reziduí zobrazuje tabuľka č. 13. Na základe hodnoty Durbinovej-Watsonovej charakteristiky (v primárnom modeli má hodnotu 1,07061 – vid' výstup č. 1) a porovnaním s kritickými medzami (tabuľka č. 13) zamietame  $H_0$  o nekorelovanosti reziduí. Reziduá sú teda podľa Durbinovho-Watsonovho testu autokorelované.

**Tabuľka č. 13:** Kritické hodnoty pre Durbinov-Watsonov test

počet pozorovaní	počet vysvetľujúcich premenných	kritické hodnoty DW	
		$d_d$	$d_h$
22	2	1,23949	1,42888
20	2	1,20149	1,41073

Zdroj: [http://nb.vse.cz/~arltova/vyuka/DW\\_krithod.xls](http://nb.vse.cz/~arltova/vyuka/DW_krithod.xls)

**Testy reziduí**

Ako už bolo spomínané pri identifikácii vplyvných pozorovaní, pre správny model je dôležité, aby reziduá (ako odhad náhodnej zložky) spĺňali určité podmienky. Ide o podmienky:

- nekorelovanosti (vid' predchádzajúci odsek – Durbinov-Watsonov test),
- homoskedasticity, tj. konštantný rozptyl reziduí a
- normálne rozdelenie reziduí.

Či sú splnené uvedené podmienky pre reziduá v primárnom modeli budem skúmať v nasledujúcom kroku. Pre tieto testy bol použitý program GiveWin z dôvodu, že Statgraphics tieto testy nepočíta priamo ako súčasť regresnej analýzy. V prípade preukázania akejkoľvek

<sup>6</sup> Pre Durbinov-Watsonov test je definovaná  $H_0: \rho=0$ ; alternatíva ako  $H_1: \rho \neq 0$  ( $\rho$  je autokorelačný koeficient).

„nenáhodnosti“ reziduí, indikuje táto skutočnosť nejakú komplikáciu alebo určitý nedostatok modelu.

Nasledujúci výstup (výstup č. 2) obsahuje test reziduí o homoskedasticite (ARCH test) a test o normálnom rozdelení reziduí (Asymptotic test). Výsledok *ARCH*<sup>7</sup> testu rozhodol o prijatí nulovej hypotézy o podmienenej homoskedasticite reziduí. Testom bol potvrdený predpoklad o konštantnom rozptyle reziduí.

Pre testovanie normálneho rozdelenia reziduí, používa GiveWin, *Jarqueov-Berov test*<sup>8</sup> (v ďalšom texte už len JB test). Testové kritérium JB testu je založené na súčasnom výpočte šikmosti a špicatosti reziduí a následným porovnaním hodnôt s hodnotami charakteristík normálneho rozdelenia. (Vieme, že šikmost' - tretí normovaný moment normálneho rozdelenia má hodnotu 0, štvrtý normovaný moment tohto rozdelenia má hodnotu 3.) Testové kritérium má  $\chi^2$  rozdelenie s dvoma stupňami voľnosti ( $\chi^2(2)$ ). Kritická hodnota sa porovnáva s príslušným kvantilom, tj. s hodnotou  $\chi^2_{0,95}(2)$ . Pre náš primárny model ide o kvantil, ktorého hodnota je 5,991. Týmto testom nebolo preukázané normálne rozdelenie reziduí (Testové kritérium má hodnotu 7,2979).

### Výstup č. 2: Testy reziduí, model s oboma vysvetľujúcimi premennými, 22 pozorovaní

```
ARCH coefficients:
  Lag Coefficient  Std.Error
    1   -0.053796   0.07408
RSS = 0.0040644  sigma = 0.0154623

Testing for error ARCH from lags 1 to 1
ARCH 1-1 test:    F(1,17) = 0.52730 [0.4776]

Normality test for Residuals
Observations      22
Mean              0.00000
Std.Devn.        0.15163
Skewness         -1.1059
Excess Kurtosis  1.7518
Minimum          -0.46585
Maximum          0.22595
Asymptotic test: Chi^2(2) = 7.2979 [0.0260]*
Normality test:  Chi^2(2) = 5.9343 [0.0515]
```

<sup>7</sup> ARCH test (AutoRegressive Conditional Heteroskedasticity Test) je založený na tom, že sa vytvorí „umelá regresia“ kde vysvetľovanou premennou je kvadrát reziduí  $e_i^2$  a vysvetľujúcou premennou je kvadrát reziduí v oneskorení o jedno obdobie, tj.  $e_{i-1}^2$ . Parametre takto definovaného modelu sa odhadnú pomocou metódy najmenších štvorcov. Za predpokladu nulovej hypotézy o podmienenej homoskedasticite nesystematickej zložky má testové kritérium rozdelenie  $\chi^2(1)$ . (Počet stupňov voľnosti je daný počtom oneskorení v rámci modelu „umelej regresie“.)

<sup>8</sup> Testové kritérium má formu  $JB=SK^2+K^2$ , kde SK je charakteristika šikmosti, K je charakteristika špicatosti.



Na základe výsledkov testov reziduí sa domnievam, že vylúčenie niektorých pozorovaní bude nutné. Pretože ako bolo testami otestované bez vylúčenia pozorovaní reziduá nespĺňajú podmienky, ktoré by pre správny model splnené mali byť – reziduá sú autokorelované a nemajú normálne rozdelenie.

V nasledujúcom kroku sa zameriam na zmenu charakteristík po vylúčení pozorovaní č. 1 a č. 22 a samostatné vylúčenie pozorovaní č. 1 a č. 16.

### **Zmeny po vylúčení pozorovaní č. 1 a č. 16**

Po vylúčení pozorovaní č. 1 a č. 16 (Nemocnice na Žižkově a Nemocnice Blansko) z pôvodného dátového súboru sa hodnoty niektorých charakteristiky v porovnaní s predchádzajúcou analýzou zmenili výraznejšie. Výsledok regresnej analýzy zobrazuje výstup č. 3. Zmenila sa predovšetkým štatistická významnosť premennej X1 pre model. Pomocou t-testu sme ako platnú hypotézu prijali nulovú hypotézu o nevýznamnosti premennej pre model. Zvýšil sa index determinácie a výrazne vzrástla aj hodnota testového kritéria F testu. V pôvodných dátach má testové kritérium F testu hodnotu 153,83, v analýze pri vylúčení pozorovaní č. 1 a 16 táto hodnota vzrástla o viac ako 100 %, na hodnotu 366,13. P-value v oboch analýzach je 0,0000. Treba mať stále na zreteli výsledok t-testu pre premennú X1 a výsledky zvyšných charakteristík treba brať s „rezervou“.

### Výstup č. 3: Regresná analýza po vylúčení pozorovania č. 1 a č. 16, analýza pre nemocnice

Multiple Regression Analysis

-----  
Dependent variable: Y  
-----

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	3,94517	0,34074	11,5783	0,0000
X1	0,043287	0,122021	0,354749	0,7271
X2	0,901213	0,128602	7,00779	0,0000

## Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	4,17561	2	2,0878	366,13	0,0000
Residual	0,0969388	17	0,00570229		
Total (Corr.)	4,27255	19			

R-squared = 97,7311 percent

R-squared (adjusted for d.f.) = 97,4642 percent

Standard Error of Est. = 0,0755135

Mean absolute error = 0,0498704

Durbin-Watson statistic = 2,13406

Zlepšila sa Durbinova-Watsonova charakteristika popisujúca autokorelovanosť reziduí. Na základe hodnoty testového kritéria Durbinovho-Watsonovho testu a príslušným medziam (tabuľka č. 13) je prijatá hypotéza o nekorelovanosti reziduí. Výsledky testov reziduí naznačujú, že podmienky kladené na klasické reziduá sú splnené. Pre testovanie normality a heteroskedasticity boli použité rovnaké testy ako v predchádzajúcich analýzách, tj. JB test a ARCH test. (Výsledky sú zobrazené vo výstupe č. 4.) Na základe týchto výsledkov testov a hodnôt p-value k daným testom prislúchajúcim, bolo rozhodnuté o podmienenej homoskedasticite reziduí (tj. o konštantnom rozptyle). Po vylúčení pozorovaní č. 1 a č. 16 už nezamietame  $H_0$  o normálnom rozdelení ich reziduí.

**Výstup č. 4:** Analýza reziduí, obe vysvetľujúce premenné, vylúčenie pozorovaní č. 1 a č. 16

ARCH coefficients:

```
Lag Coefficient  Std.Error
1      -0.21138    0.2544
RSS = 0.00147692  sigma = 0.00992278
```

```
Testing for error ARCH from lags 1 to 1
ARCH 1-1 test:   F(1,15) = 0.69041 [0.4191]
```

```
Normality test for Residuals
Observations      20
Mean              0.00000
Std.Devn.         0.069620
Skewness          1.0031
Excess Kurtosis   1.2887
Minimum           -0.10895
Maximum           0.19614
Asymptotic test:  Chi^2(2) = 4.7379 [0.0936]
Normality test:   Chi^2(2) = 4.8278 [0.0895]
```

Nasledujúca tabuľka zobrazuje hodnoty jednoduchých, dielčích korelačných koeficientov, ako aj hodnotu viacnásobného korelačného koeficientu pre súbor nemocníc, po vylúčení pozorovania č.1 a č.16 zo základného súboru. Ako vidieť, hodnoty koeficientov vzrástli. Potvrďuje sa tým správnosť predpokladu o tom, že tieto pozorovania sú ako vplyvné označené správne.

**Tabuľka č. 14:** Korelačné koeficienty, pre nemocnice, model s oboma vysvetľujúcimi premennými, po vylúčení pozorovaní č. 1 a č. 16

$\Gamma_{yx1}$	0,9549
$\Gamma_{yx2}$	0,9885
$\Gamma_{x1x2}$	0,9624
$\Gamma_{yx1.x2}$	0,0857
$\Gamma_{yx2.x1}$	0,8619
$\Gamma_{y.x1x2}$	0,9773

**Obe alebo len jedna vysvetľujúca premenná?**

Ako bolo ukázané, hodnoty jednoduchých korelačných koeficientov dokazujú silnú závislosť vysvetľujúcich premenných. Ak by v modeli boli ponechané obe, došlo by ku skresleniu odhadov. Výsledky regresnej analýzy a pomocných analýz (analýza reziduí, vplyvných pozorovaní a testy reziduí) v prípade ponechania oboch premenných ako vysvetľujúcich premenných zobrazujú predchádzajúce výstupy. Bolo by vhodnejšie ponechať v modeli namiesto dvoch vysvetľujúcich premenných len jednu? Vyriešil by sa tým problém silnej lineárnej závislosti vysvetľujúcich premenných medzi sebou a s vysvetľovanou premennou? Ktorá z premenných by bola vhodnejšia pre ponechanie v modeli?

Odpoveď na poslednú otázku je viac-menej jasná. Štatistická nevýznamnosť premennej X1, bola viackrát naznačená výsledkom t-testu. I na základe hodnôt jednoduchých korelačných koeficientov je zrejmé, že premennou, ktorá by mala byť ponechaná v modeli, je premenná X2. (Pomocou samotnej premennej X2 je vysvetlené väčšie percento variability dát.) Dali by sa aplikovať špeciálne metódy slúžiace ku určaniu vhodnej podmnožiny vysvetľujúcich premenných: metóda Forward a metóda Backward. Ale v prípade dvoch premenných a znalosti hodnôt jednoduchých korelačných koeficientov sa mi aplikácia týchto metód zdá byť zbytočná. (Výsledkom oboch metód by bolo ponechanie premennej X2 ako jedinej vysvetľujúcej premennej.)

### **Definícia modelu**

Na základe uskutočnených analýz usudzujem, že model, ktorý bol zostavený na začiatku tejto analýzy, nie je „tým pravým“ modelom. V prípade premennej X1 bolo t-testom rozhodnuté o štatistickej nevýznamnosti pre model (čo bolo z časti určite spôsobené silnou lineárnou závislosťou oboch vysvetľujúcich premenných). V modeli sa vyskytli odľahlé pozorovania (pozorovanie č. 1, č. 16). Po ich vylúčení sa ale nezmenil výsledok t-testu premennej X1, hoci sa zlepšili charakteristiky reziduí. Preto ďalším krokom bude odhad modelu, v ktorom vysvetľujúcou premennou bude premenná X2 (počet lôžok). Opäť sa zameriam na vlastnosti reziduí a zmeny charakteristík, ktoré v predchádzajúcich analýzach nevychádzali v súlade s teoretickými predpokladmi.

### **Model po vylúčení premennej X1 (počet lekárov)**

Pristúpme k definícii regresnému modelu s vysvetľovanou premennou Y (počet hospitalizovaných) a jednou vysvetľujúcou premennou X2 (počet lôžok). Ako sa zmenia charakteristiky modelu, v ktorom ponechám len jednu vysvetľujúcu premennú? Je tento model „dobrý“ alebo naopak? V modeli ponechávam všetky pôvodné pozorovania. Neskôr sa budem zaoberať analýzou odľahlých pozorovaní. V prípade ich existencie ich z analýzy vylúčim a následne preskúmam zmeny charakteristík. Výstup č. 5 zobrazuje odhad parametrov pomocou regresnej analýzy pri pôvodnom počte pozorovaní s jednou vysvetľovanou premennou, premennou X2.

**Výstup č. 5: Regresná analýza, model s jednou vysvetľujúcou premenou (premenná X2) pre nemocnice**

Multiple Regression Analysis

Dependent variable: hospitalizovany

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2,81406	0,366609	7,67591	0,0000
X2	1,11113	0,061477	18,0739	0,0000

## Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	8,39503	1	8,39503	326,67	0,0000
Residual	0,513984	20	0,0256992		
Total (Corr.)	8,90902	21			

R-squared = 94,2307 percent

R-squared (adjusted for d.f.) = 93,9423 percent

Standard Error of Est. = 0,16031

Mean absolute error = 0,121282

Durbin-Watson statistic = 1,11841

Hodnota testového kritéria F testu sa výrazne zvýšila (kritérium má hodnotu 326,67) a p-value je menšia ako  $\alpha$  hladina významnosti. Obdobne sú na tom oba t-testy o významnosti jednotlivých parametrov pre model. Ani jedným t-testom sme nepreukázali štatistickú nevýznamnosť premennej pre model. Hodnota indexu  $R^2$  sa nám v porovnaní s predchádzajúcim modelom výrazne nezmenila (upravený  $R^2$  dokonca nepatrne vzrástol). Prvý pohľad na výsledky regresnej analýzy hovorí, že vylúčenie premennej X1 z modelu bol správny. Zmenili sa ostatné charakteristiky? Hodnota jednoduchého korelačného koeficientu  $r_{yx2}$  je 0,9707 (pre model zahŕňajúci všetkých 22 pozorovaní). Nasledujúce tabuľky (tabuľka č. 15 a č. 16) obsahujú hodnoty charakteristík potrebných k analýze a následnej identifikácii vplyvných pozorovaní. Boli použité rovnaké veličiny ako pre model s oboma vysvetľujúcimi premennými. (Preto nebudem popisovať spôsob ich konštrukcie.) Testami vlastností klasických reziduí sa budem zaoberať neskôr. Za zmienku stojí výrazný rozdiel klasických a predikovaných reziduí pre prvé pozorovanie.

Jackknife reziduá sa porovnávajú s hodnotou kvantilu  $t_{0,975}(22-2-1) = t_{0,95}(19) 2,4344$  (ak je veličina väčšia, resp. menšia ako uvedené hodnoty, pozorovanie sa považuje za vplyvné). Na základe tejto charakteristiky bolo identifikované jedno pozorovanie.

Projekčný prvok  $h_{ii}$  sa v tomto modeli porovnával s hodnotou 0,1818. Opäť bolo ako vplyvné označené pozorovanie č. 1, ale aj pozorovanie č. 22.

**Tabuľka č. 15:** Analýza reziduí a vplyvných pozorovaní pre súbor nemocníc, model s jednou vysvetľujúcou premennou

Č.	Klasické reziduá $e_i$	Predikované reziduá $e_{i(-i)}$	Rozdiel $ (1)-(2) $	Jackknife reziduá	Leverage $h_{ii}$
1	-0,43618	-0,60930	<b>0,17312</b>	<b>-2,72086</b>	<b>0,28413</b>
2	0,06958	0,07689	0,00731	0,43404	0,09506
3	0,11248	0,12499	0,01251	0,70165	0,10009
4	0,20433	0,23470	0,03037	1,27458	0,12942
5	0,16581	0,18725	0,02144	1,03434	0,11448
6	0,07420	0,07880	0,00459	0,46287	0,05829
7	0,23893	0,25031	0,01138	1,49046	0,04546
8	0,08582	0,08999	0,00417	0,53533	0,04632
9	0,03684	0,03871	0,00187	0,22981	0,04835
10	0,13592	0,14320	0,00728	0,84784	0,05086
11	0,07555	0,07921	0,00365	0,4713	0,04614
12	-0,08398	-0,08829	0,00430	-0,52389	0,04874
13	-0,04543	-0,04814	0,00272	-0,28336	0,05643
14	0,04814	0,05085	0,00271	0,30031	0,05324
15	-0,11669	-0,12530	0,00861	-0,72788	0,06871
16	-0,26906	-0,29706	0,02800	-1,67839	0,09426
17	0,06026	0,06486	0,00461	0,37588	0,07103
18	-0,07078	-0,07758	0,00679	-0,44155	0,08754
19	0,02623	0,02756	0,00134	0,16359	0,04844
20	-0,06535	-0,07354	0,00819	-0,40764	0,11133
21	-0,11902	-0,13241	0,01339	-0,74245	0,10110
22	-0,12760	-0,16802	0,04042	-0,79596	<b>0,24058</b>

Ako ďalšie charakteristiky pre identifikáciu vplyvných pozorovaní boli opäť použité charakteristiky ako Mahalanobisova vzdialenosť, Cookova vzdialenosť, Covratio a veličina štandardizovaný DFFIT<sub>(-i)</sub>. Kritické hodnoty pre tieto charakteristiky uvádzam v tabuľke č. 16.

**Tabuľka č. 16:** Ďalšie charakteristiky slúžiace k analýze vplyvných pozorovaní, model s jednou vysvetľujúcou premennou

Č.	Mahalanobisova vzdialenosť	Prepočet pre Mahalanobisovu vzdialenosť	Cookova vzdialenosť	Covratio	Štandardizovaný DFFIT <sub>(-i)</sub>
1	5,01215	2,28297	<b>2,05224</b>	<b>0,36098</b>	<b>-2,84150</b>
2	1,04175	0,47450	0,01093	1,19907	0,14489
3	1,14729	0,52258	0,03042	1,16483	0,24378
4	1,76323	0,80313	0,13870	1,04633	0,53912
5	1,44959	0,66027	0,07810	1,10467	0,39740
6	0,26962	0,12281	0,00704	1,15001	0,11633
7	0,00020	0,00009	0,05542	0,90638	0,34521
8	0,01812	0,00825	0,00730	1,12720	0,11864
9	0,06076	0,02768	0,00141	1,15787	0,05182
10	0,11346	0,05168	0,02029	1,08066	0,20017
11	0,01431	0,00652	0,00563	1,13473	0,10405
12	0,06892	0,03139	0,00739	1,13144	-0,11937
13	0,23054	0,10501	0,00254	1,16433	-0,06968
14	0,16357	0,07450	0,00268	1,15923	0,07151
15	0,48840	0,22246	0,02099	1,12306	-0,20259
16	1,02496	0,46686	0,16184	0,87245	<b>-0,60343</b>
17	0,53701	0,24460	0,00581	1,17468	0,10551
18	0,88386	0,40259	0,01025	1,18853	-0,14030
19	0,06276	0,02859	0,00072	1,16117	0,03691
20	1,38334	0,63009	0,01171	1,22364	-0,14988
21	1,16859	0,53228	0,03449	1,15822	-0,25999
22	4,09757	1,86639	0,13214	<b>1,33986</b>	-0,51185

**Tabuľka č. 17:** Hodnoty charakteristík použitých k identifikácii vplyvných pozorovaní a identifikované vplyvné pozorovania, model s premennou X2 ako vysvetľujúcou

Charakteristika	kritické hodnoty	Identifikované vplyvné pozorovania
Mahalanobisova vzdialenosť (po prepočte)	$D_i^2 > 3,493$	žiadne pozorovanie
Cookova vzdialenosť	$C_i > 1$	č. 1
Covratio	$covratio_{(-i)} > 1,2727$ $covratio_{(-i)} < 0,7272$	č. 1, č. 22
DFFIT <sub>(-i)</sub>	$ DFFIT_{(-i)}  > 0,603$	č. 1, č. 16

Podľa posledných analýz boli v modeli identifikované niektoré pozorovania ako odľahlé. Vo väčšine charakteristík bolo ako vplyvné pozorovanie určené pozorovanie č. 1 a č. 16. Výsledok je rovnaký ako pri analýze odľahlých pozorovaní pre primárny model (tabuľka č. 10 a č. 11). Pozorovanie č. 22 bolo označené ako vplyvné jednou veličinou. Nebudem ho ale z modelu vylučovať. Zmenia sa výrazne regresné charakteristiky po ich

vylúčení, tak ako tomu bolo v modeli zahrňujúcom obe premenné, alebo nie? Hodnota jednoduchého korelačného koeficientu, po vylúčení pozorovania č. 1 z analyzovaného súboru vzrástla na  $r_{yx2}=0,9885$ .

**Výstup č. 6:** Regresná analýza po vylúčení pozorovaní č. 1 a č. 16, model s jednou vysvetľujúcou premennou, analýza pre nemocnice

Multiple Regression Analysis

-----  
Dependent variable: Col\_1  
-----

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	3,85034	0,2061	18,6819	0,0000
X2	0,945119	0,0340705	27,7401	0,0000

-----  
Analysis of Variance  
-----

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	4,17489	1	4,17489	769,51	0,0000
Residual	0,0976565	18	0,00542536		
Total (Corr.)	4,27255	19			

R-squared = 97,7143 percent

R-squared (adjusted for d.f.) = 97,5873 percent

Standard Error of Est. = 0,073657

Mean absolute error = 0,0509721

Durbin-Watson statistic = 2,10366

Vidíme, že hodnoty odhadnutých parametrov v prípade konštanty vzrástli (z hodnoty 2,81406 v modeli ktorý zahŕňal 22 pozorovaní) na hodnotu 3,85034 po vylúčení vplyvných pozorovaní. V prípade odhadu parametru premennej X2 sa hodnota znížila z pôvodnej hodnoty 1,11113 na 0,945119. Konštanta i premenná zostávajú podľa jednotlivých t-testov štatisticky významné, testové kritéria majú ale výrazne vyššie hodnoty. Túto skutočnosť je možné interpretovať ako ešte jednoznačnejšie zamietnutie nulovej hypotézy  $H_0$  o nulovej hodnote jednotlivých parametrov. Už pri analýze, ktorá zahrňovala všetky pozorovania, bolo zamietnutie  $H_0$  „presvedčivé“. Za pozornosť stojí zmena vo výsledku celkového F testu. Nezmenil sa výsledok, ale rovnako, ako pri t-testoch, došlo k zvýšeniu hodnoty testového



kritéria. Je logické predpokladať, že pri zmenách hodnôt testových kritérií, ku ktorým došlo, dôjde aj k rastu percentuálnej zhody modelu s dátami. Tento predpoklad potvrdzuje index determinácie,  $R^2$ . V modeli, ktorý zahŕňa všetky pozorovania, je jeho hodnota 94,2307 %, čo je približne na úrovni hodnoty  $R^2$  v primárnom modeli. Po vylúčení pozorovania č. 1 a č. 16 došlo k rastu jeho hodnoty na hodnotu 97,7143 %. Hodnota Durbinovej-Watsonovej štatistiky je po vylúčení pozorovaní 2,10366. Pred vylúčením pozorovaní bola jej hodnota relatívne nízka, 1,11841. Netreba zabudnúť na testy (výstup č. 7) reziduálnej zložky a *predčasne* sa tešiť z predchádzajúcich výsledkov.

### Výstup č. 7: Analýza reziduí, model s jednou vysvetľujúcou premennou, po vylúčení pozorovaní č. 1 a č. 16, analýza pre nemocnice

Testing for error ARCH from lags 1 to 1  
ARCH 1-1 test: F(1,16) = 0.79185 [0.3867]

Normality test for Residuals  
Observations 20  
Mean 0.00000  
Std.Devn. 0.069877  
Skewness 0.82439  
Excess Kurtosis 0.86082  
Minimum -0.10980  
Maximum 0.18965  
Asymptotic test: Chi<sup>2</sup>(2) = 2.8829 [0.2366]  
Normality test: Chi<sup>2</sup>(2) = 3.7695 [0.1519]

Výsledky naznačujú, že predpoklady kladené teóriou na vlastnosti reziduí sú, po vylúčení dvoch vplyvných pozorovania, splnené. ARCH testom bola prijatá nulová hypotéza o homoskedastickom rozptyle reziduí a JB test potvrdil predpoklad normálneho rozdelenia reziduálnej zložky.

### Konečný model

Na základe uskutočnených analýz bolo rozhodnuté, že ako konečný model, bude označený model, ktorý popisuje väzbu vysvetľovanej premennej Y (počet hospitalizovaných) len pomocou premennej X2 (počet lôžok). Na základe analýz reziduí a odlahlých pozorovaní považujem pozorovanie č. 1 za pozorovania vplyvné a z konečného modelu ho vylučujem.

Môžeme teda tento model charakterizovať rovnicou:

$$\hat{Y} = 3,85034 + 0,945119 \cdot X_2, \quad (5.1)$$

kde  $\hat{Y}$  je logaritmicke transformovaná premenná predstavujúca počet hospitalizovaných ( $\ln Y$ ) a  $X_2$  logaritmicke upravenú premennú počet lôžok ( $\ln X_2$ ).

Ako vieme, premenné boli pre možnosť použitia lineárnej regresnej analýzy logaritmicky transformované. Preto z modelu tak ako bol uvedený v (5.1) je možné spätne transformovať výsledky a dostať tak odhad Cobb-Douglasovej produkčnej funkcie pre súbor nemocníc. Výsledkom je rovnica:

$$\hat{Y} = 3,85034 X_2^{0,945199},$$

kde  $Y$  a  $X_2$  sú už hodnoty pôvodných, nelogaritmicky transformovaných veličín. Táto rovnica popisuje závislosť počtu hospitalizovaných (premenná  $Y$ ) na počte lôžok (premenná  $X_2$ ).

### 5.2.2 Odhad Cobb-Douglasovej produkčnej funkcie pre LDN

Rovnako ako som postupovala v prípade odhadu produkčnej funkcie pre súbor nemocníc, rozhodla som sa postupovať aj pre súbor LDN. Zostavila som základný model s oboma vysvetľujúcimi premennými. Výsledok regresnej analýzy zobrazuje výstup č. 8.

#### Výstup č. 8: Regresná analýza pre súbor LDN, obe vysvetľujúce premenné

Multiple Regression Analysis

-----  
Dependent variable: Y  
-----

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	6,60893	3,35261	1,97128	0,1058
X1	0,599957	0,522689	1,14783	0,3030
X2	-0,378964	0,839672	-0,451324	0,6707

#### Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0,414801	2	0,207401	0,78	0,5061
Residual	1,32491	5	0,264981		
Total (Corr.)	1,73971	7			

R-squared = 23,8432 percent

R-squared (adjusted for d.f.) = 0,0 percent

Standard Error of Est. = 0,514763

Mean absolute error = 0,365256

Durbin-Watson statistic = 1,77354

Ako vidieť z výstupu, hneď podľa prvej analýzy sa model javí ako absolútne nevhodný. Individuálne t-testy vyšli pre obe vysvetľujúce premenné i pre konštantu nevýznamne: nízke hodnoty testových kritérií a vysoké hodnoty p-value svedčia v prospech nulových hypotéz. Výsledok priaznivý pre nulovú hypotézu sme dostali aj v prípade celkového F testu. V súlade s výsledkami testov o významnosti jednotlivých premenných, ako aj modelu ako celku, je aj hodnota  $R^2$ , ktorá je veľmi nízka. Ako vidieť z rozkladu rozptylu na variabilitu vysvetlenú modelom a variabilitu reziduálnu, model vystihuje nízke percento celkovej variability. Takto nízke percento charakteristiky  $R^2$  je skôr výnimkou ako nejakým štatistickým dôsledkom. Nepriaznivé výsledky sa nezmenia ani v prípade, že by sme do modelu zaradili len jednu vysvetľujúcu premennú. Výsledky regresnej analýzy v prípade vysvetľujúcej premennej reprezentujúcej počet lekárov zobrazuje výstup č. 9. Následne bol model odhadnutý pre prípad jedinej vysvetľujúcej premennej, premennej predstavujúcej počet lôžok.

**Výstup č. 9: Regresná analýza pre súbor LDN, jedna vysvetľujúca premenná (počet lekárov)**

Multiple Regression Analysis

-----  
Dependent variable: Y  
-----

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	5,10938	0,417048	12,2513	0,0000
X1	0,434335	0,346628	1,25303	0,2568

## Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0,360826	1	0,360826	1,57	0,2568
Residual	1,37888	6	0,229814		
Total (Corr.)	1,73971	7			

R-squared = 20,7406 percent

R-squared (adjusted for d.f.) = 7,53073 percent

Standard Error of Est. = 0,479389

Mean absolute error = 0,383058

Durbin-Watson statistic = 1,53156

Ako vidieť, výsledky analýz len s jednou vysvetľovanou premennou v modeli sa nezmenili. Jednou z možných príčin nevhodnosti modelu je určite aj malé množstvo pozorovaní. Ale pretože ide o dáta, ktoré neboli získané experimentom, nie je možné dáta nejakým spôsobom upraviť, resp. uskutočniť nový experiment.

**Výstup č. 10:** Regresná analýza pre súbor LDN, jedna vysvetľujúca premenná (počet lôžok)

Multiple Regression Analysis

-----  
Dependent variable: Y  
-----

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	4,26583	2,72896	1,56317	0,1690
X2	0,297699	0,613546	0,485211	0,6447

## Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0,0656859	1	0,0656859	0,24	0,6447
Residual	1,67402	6	0,279004		
Total (Corr.)	1,73971	7			

R-squared = 3,77568 percent

R-squared (adjusted for d.f.) = 0,0 percent

Standard Error of Est. = 0,528208

Mean absolute error = 0,41489

Durbin-Watson statistic = 1,43438

Záverom je teda skutočnosť, že na základe dát, ktoré boli k dispozícii, nie je pre súbor liečební (LDN) možné odhadnúť Cobb-Douglasovu produkčnú funkciu.

## 6 Závěry

Analýza obalu dát bola aplikovaná bez významnejších komplikácií, najprv na súbor nemocníc a následne na súbor liečební dlhodobo chorých. Prípadné obmedzenia a limity analýz boli zmienené v prvej kapitole. V rámci analyzovaného súboru pre nemocnice, boli identifikované tzv. 100%-ne efektívne jednotky. Pre model CCR (orientovaný ako na vstupy, tak na výstupy) efektívnou jednotkou bola jediná jednotka, Nemocnice Písek. Pre modely BCC orientované na výstupe a vstupe boli ako efektívne jednotky označené Nemocnice v Jihlavě, Nemocnice na Žižkově, v Stodůlkách, v Rokycanech, Fakultní Nemocnice Na Bulovce a takisto už spomínaná písecká nemocnica. V súbore LDN boli v rámci CCR modelov ako efektívne jednotky označené jednotky dve, a to LDN v Hradci Králové a LDN Jablonec nad Nisou. K týmto efektívnym jednotkám sa po zmene predpokladu o charaktere výnosov z rozsahu, z konštantného na variabilný, pridali ďalšie: LDN Lomnice nad Popelkou, LDN na Vojkově a LDN Opočno. Pre obe analýzy, pre súbor nemocníc i súbor LDN, boli identifikované vzorové jednotky (jednotky ktoré v danom modeli sú efektívne) a cieľové hodnoty ktorých by neefektívne jednotky mali dosiahnuť, aby sa stali jednotkami efektívnymi. Cieľové hodnoty boli určené pre model s variabilnými výnosmi z rozsahu.

Na rovnaký súbor dát bola po analýze obalu dát aplikovaná ekonometrická analýza – odhad Cobb-Douglasovej produkčnej funkcie. Prvým krokom bola logaritmická transformácia dát, a to z dôvodu použitia lineárnej regresnej analýzy. Komplikácie pri praktickom odhade boli spôsobené problémami, ktoré sú typické pre odhadovanie ekonomických situácií. Vysvetľujúce premenné boli vzájomne silne lineárne závislé. V súbore sa vyskytli niektoré vplyvné pozorovania. Pre súbor nemocníc sa model s oboma vysvetľujúcimi premennými nezdal vhodný z dôvodu výsledku t-testu pre túto premennú. Medzi vysvetľujúcimi premennými bola silná lineárna závislosť, čo takisto skresľovalo výsledky jednotlivých analýz. Analýzou vplyvných pozorovaní sa ako vplyvné ukázali pozorovania č. 1 (Nemocnice na Žižkově) a pozorovanie č. 16 (Nemocnice Blansko). Po ich vylúčení zo základného súboru došlo k zlepšeniu niektorých charakteristík (testy o predpokladoch reziduí, Durbinov-Watsonov test), avšak nedošlo k zmene výsledku t-testu pre predstavujúcu počet lekárov (tak, aby nebola v rozpore s teoretickými požiadavkami ohľadne kladnej hodnoty). Nakoniec bol primárny model modifikovaný tak, že bola z modelu vypustená premenná počet lekárov. Počet hospitalizovaných som sa pokúsila popísať

pomocou premennej počet lôžok. Tak ako v primárnom modeli, aj tu sa analýzou vplyvných pozorovaní ukázali pozorovania č. 1 a č. 16 ako vplyvné. Následne boli z modelu vylúčené. Po testovaní ďalších charakteristík – opätovný odhad regresnej funkcie pri zmenšenom počte pozorovaní a nové testy reziduí - bolo rozhodnuté o konečnom tvare regresného modelu. Modelom je model popisujúci počet hospitalizovaných pacientov na základe údajov o počte lôžok, s počtom pozorovaní 20 (z pôvodného počtu 22). Obdobný postup, použitý pre analýzu súboru nemocníc, som chcela aplikovať aj na súbor liečební dlhodobo chorých. Tu sa vyskytli ale významné problémy hneď od začiatku. Primárny model, tak ako bol formulovaný pre súbor nemocnice, nebol vôbec štatisticky významný. Výsledky sa nemenili ani v prípade zže v modeli zostala len jedna z vysvetľujúcich premenných. Bolo teda konštatované, že odhad Cobb-Douglasovej produkčnej funkcie pre súbor liečební dlhodobo chorých nebol úspešný a nie je z daného dátového súboru možný. Analyzovaný súbor obsahoval len 8 pozorovaní, čo sa v rámci DEA modelov nezdalo byť prekážkou, ale na aplikáciu regresnej analýzy je to súbor veľmi malý, čo mohlo byť príčinou tak nepriaznivých výsledkov, akých bolo dosiahnutých.

V tejto práci som sa snažila použiť rôzne prístupy k hodnoteniu efektívnosti jednotiek: neparametrické metódy a ekonometrickú analýzu. Každá z týchto metód má svoje výhody a nevýhody. Nespornou výhodou analýzy obalu dát je možnosť práce s viacrozmernými vstupmi a výstupmi, to, že pre každú neefektívnu jednotku je možné identifikovať vzor a v závislosti od modelu aj cieľové hodnoty. Ďalšou, predovšetkým užívateľskou, výhodou je možnosť softwarového riešenia úloh. Na druhej strane nevýhodou tohto spôsobu skúmania efektívnosti jednotiek je jeho deterministický prístup a nie tak prepracovaná metodika ohľadne testovania významnosti vstupov a výstupov. Posledný nedostatok sa časom určite zmenší, resp. vyrieši. Ekonometrická analýza je na rozdiel od DEA „staršia“ a tým aj po štatistickej stránke prepracovanejšia. Najmä v oblasti testovania významnosti jednotlivých premenných a v oblasti voľby funkčného vzťahu zvolených premenných (samotná formulácia modelu). Všeobecným problémom ekonomických situácií je často silná závislosť jednotlivých premenných čo spôsobuje problémy pri overovaní základných predpokladov, ktorých splnenie je nutné pre relevantnosť ďalších analýz. Všetko má svoje pre i proti a preto sa nedá povedať, ktorý spôsob skúmania je „lepší“.

Oblasť zdravotníctva vždy bola a stále je oblasťou, v ktorej je otázka efektívnosti, či už zariadení alebo jednotlivých osôb (lekári, pomocný zdravotný personál) „nadčasová“.

Efektívnosť v oblastiach, ktoré nie sú riadené „neviditeľnou rukou trhu“, je komplikovaná o to viac, že tu nefunguje mechanizmus samonastavenia rovnováhy. Preto je riešenie neefektívnosti „behom na dlhú trať“ a nie je v silách jednotlivca, aby ju vyriešil. V rámci tejto diplomovej práce sa nesnažím neefektívnosť analyzovaných jednotiek riešiť, ale na ňu poukázať. Pri analýze pomocou DEA modelov je načrtnutý možný spôsob redukcie neefektívnosti jednotiek.

V rámci písania mojej diplomovej práce som mohla spolupracovať na riešení grantového projektu na katedre ekonometrie. Časti mojej diplomovej práce, ktoré vznikli v rámci tohto grantu sú využité v článku predloženom do časopisu *Politická ekonomie* (Jablonský, J., Dlouhý, M., Novosádová, I.: Využití analýzy obalu dat pro hodnocení efektívnosti českých nemocnic), ktorý je v tejto dobe v recenznom procese.



## 7 Použitá literatúra

- [1] Coelli, T., Rao, D.S.P., Battese, G.E.: *An introduction to efficiency and productivity*, Kluwer Academic Publishers, Boston/Dordrecht/London, 1998
- [2] Cooper, W. W., Seiford, L. M., Charnes, A., Lewin, A.Y.: *Data Envelopment Analysis – Theory, Methodology and Applications*, 4th Printing, Kluwer Academic Publishers, Boston/Dordrecht/London, 1998
- [3] Cooper, W. W., Seiford, L. M., Zhu, J.: *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, Boston/Dordrecht/London, 2004
- [4] Farrell, M. J.: The measurement of productive efficiency, *Journal of Royal Statistical Society A* 120:253-281, 1957
- [5] Hebák, P., Hustopecký, J., Malá, I.: *Vícerozměrné statistické metody (2)*, Informatorium, Praha, 2005
- [6] Hollingsworth B.: Non-parametric and parametric applications measuring efficiency in health care, *Health Care Management Science*, č. 6: 203-218, 2003
- [7] Hušek, R.: *Ekonometrická analýza*, Ekopress, Praha, 1999
- [8] Jablonský, J., Dlouhý, M.: *Modely hodnocení produkčních jednotek*, 1. vydanie, Professional Publishing, Praha 2004
- [9] Jablonský, J., Lagová, M.: *Lineární modely*, VŠE, Praha, 2004
- [10] Jablonský, J., Dlouhý, M., Novosádová, I.: Využití analýzy obalu dat pro hodnocení efektívnosti českých nemocnic, *Politická ekonomie*, v recenznom procese
- [11] Nunamaker, T. R.: Measuring routine nursing service efficiency: A comparison of cost per patient day and data envelopment analysis models, *Health Services Research*, č. 18: 183-205, 1983
- [12] Sherman, H. D.: Hospital efficiency measurement and evaluation. Empirical test of a new technique, *Medical Care*, č. 10: 922-938, 1984
- [13] Soukupová, J., Horejší, B., Macáková, L., Soukup, J.: *Mikroekonomie*, 3. doplnené vydání, Management Press, Praha, 2004

**Internetové zdroje:**

- Ústav zdravotnických informací a statistiky ČR: [www.uzis.cz](http://www.uzis.cz)