VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ
FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY
INSTITUTE OF MATHEMATICS

REGRESNÍ ANALÝZA PROSTOROVĚ A ČASOVĚ DISTRIBUOVANÝCH DAT
REGRESSION ANALYSIS OF SPATIALLY AND TIME DISTRIBUTED DATA

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **Martin Rosecký**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Josef Bednář, Ph.D.**
SUPERVISOR

BRNO 2016

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA STROJNÍHO INŽENÝRSTVÍ

Zadání bakalářské práce

Ústav:	Ústav matematiky
Student:	**Martin Rosecký**
Studijní program:	Aplikované vědy v inženýrství
Studijní obor:	Matematické inženýrství
Vedoucí práce:	**Ing. Josef Bednář, Ph.D.**
Akademický rok:	2015/16

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Regresní analýza prostorově a časově distribuovaných dat

Stručná charakteristika problematiky úkolu:

Práce prohlubuje znalosti studenta v oblasti regresní analýzy a klade důraz na jejich praktické využití. Práce podpoří další vývoj nástroje JUSTÝNA, což je rekurzivně použitý stochastický matematický model, který je aplikován na území (ČR) rozděleném na menší správní jednotky (obce s rozšířenou působností). Nástroj zpracovává dostupná statistická data z různých zdrojů informací, kombinuje je s obecně platnými modely a na mikroregionální úrovni vyhodnocuje kredibilitu těchto modelů. Úkolem a hlavním přínosem práce bude aplikace poznatků regresní analýzy při přípravě modelů pro nástroj JUSTÝNA. Student bude pracovat s reálnými daty.

Cíle bakalářské práce:

Studium teoretických poznatků v oblasti regresní analýzy.

Rešerše v oblasti přístupu prognózování komunálních odpadů

Formulace vhodných modelů pro vybranou aplikaci

Seznam literatury:

Zvára, K. (2008): Regrese. MatfyzPress, Praha.

Anděl, J. (2011): Základy matematické statistiky. MatfyzPress, Praha.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2015/16

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc. ředitel ústavu	doc. Ing. Jaroslav Katolický, Ph.D. děkan fakulty

ABSTRAKT

V práci byly shrnuty poznatky z oblasti prognózování komunálního odpadu (KO). Byly popsány základní informace týkající se lineární regrese a korelační analýzy. Byla provedena analýza vlivů dostupných faktorů na úrovni obcí s rozšířenou působností (ORP). Výsledné modely objasňují až 99 % variability. Modely pro množství odpadu na osobu vysvětlují 12 až 75 % variability. Variabilita KO na osobu vysvětlená modelem je cca o 20% menší, než u srovnatelné studie, která však používá běžně nedostupná data. Modely jsou pro oblast odpadového hospodářství (OH) použitelné a jejich zdánlivá jednoduchost je v praxi výhodou.

KLÍČOVÁ SLOVA

lineární regrese, lineární regresní model, korelační analýza,komunální odpad

ABSTRACT

This thesis sumarizes findings about municipal solid waste (MSW) forecasting. Basic information about linear regression and correlation analysis were described. Analysis of influencing factors was realized on municipality with extended competence level. The resulting models explain up to 99 % of variability. Final models of MSW per capita explain between 12 and 75 % of variability. Variability explained by model of MSW per capita is lower by 20% than comparable study which however uses data that are not usually available. Models can be used in waste management and their simplicity is benefit for real usage.

KEYWORDS

linear regression, linear regression model, correlation analysis, municipal solid waste

ROSECKÝ, Martin *Regresní analýza prostorově a časově distribuovaných dat*: bakalářská práce. BRNO: Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav Matematiky, 2016. 44 s. Vedoucí práce Ing. Josef Bednář, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Regresní analýza prostorově a časově distribuovaných dat“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

BRNO ....

podpis autora

PODĚKOVÁNÍ

Tímto bych rád poděkoval vedoucímu bakalářské práce panu Ing. Josefu Bednářovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci. Dále bych chtěl poděkovat Ing. Radovanu Šomplákovi, Ph.D. za nespočet konzultací, odborných rad a užitečných podnětů.

BRNO ....

podpis autora

OBSAH

ÚVOD

Lidé už ode dávna mají potřebu a snahu předpovídat vývoj budoucích událostí. To souvisí jak s osobním prospěchem, tak i s kolektivní potřebou. V dávných dobách byli jedinou možností věštci a proroci, jejichž následovníci se o totéž snaží dodnes. Tento fakt je potvrzením toho, že lidskou rasu nepřestává ovládat potřeba vidět do budoucnosti, ba naopak, tato potřeba stále roste a stává se nutností. Kdo z nás nikdy nezatoužil po tom, znát nějakou informaci z budoucnosti? Tím spíše, že v dnešní době jsou informace možná tou nejcennější komoditou. Jen si představte, jaké by to bylo znát vývoj akciových trhů, politických a vojenských manévrů, sportovních utkání, klimatických změn atd. Dnes už jsou ale k dispozici i poněkud exaktnější přístupy, které jsou založené na pevných matematických základech a pokud jsou použity správným způsobem, mohou nám pomoci i v oblastech, které jsou důležité pro společnost, respektive pro celé lidstvo. Za všechny jmenujme předpovědi přírodních katastrof a jejich vývoje, změn klimatu, růstu a složení obyvatelstva či chování kosmických těles.

V neposlední řadě lze do této kategorie zařadit i predikování odpadu, který lidstvo produkuje a musí s ním také nakládat. Tento problém je globální, protože lidé produkují odpady ve všech částech světa, ve kterých se vyskytují. S rostoucím počtem obyvatel Země pak logicky roste i množství odpadu, který jako lidstvo produkujeme. Jde navíc o velice aktuální problém, jelikož v současné době nic nenapovídá tomu, že by se zmíněné trendy měly zásadně měnit. Hlavním cílem práce tak bude zanalyzování dostupných dat, která by mohla mít vliv na produkci odpadu a posléze vytvoření regresních modelů, které se pokusí kvalitativně i kvantitativně popsat produkovaný odpad.

Neméně podstatným problémem je, že s růstem populace roste i spotřeba komodit, kterých máme ale většinou pouze omezené množství. Prioritou je samozřejmě minimalizování množství odpadu, posléze přichází na řadu jeho recyklace. Pokud již není možné materiální využití odpadu, přichází na řadu využití energetické, znamenající úsporu primárních zdrojů. K tomu však musí být vybudována odpovídající infrastruktura. Zde se opět vynořuje potřeba co nejpřesnější predikce odpadu a jeho složení. S využitím těchto informací může být posouzeno zda, a kde je potřeba vybudovat nové zařízení, rozšířit nebo zredukovat stávající či je výhodnější jiný způsob zpracování odpadu. Nezanedbatelnou roli hraje predikce odpadu i v oblasti optimalizace sběru a transportování odpadu. Všechny zmíněné (a mnohé další) aspekty mohou přinést nezanedbatelné úspory peněz, energií a také zmenšení dopadu na životní prostředí. Jedná se tedy o cíle, které by se neměly měnit v závislosti na vládnoucí garnituře, tak abychom naši zemi, potažmo celou planetu, zanechali budoucím generacím v co možná nejlepším stavu.

1 KLASIFIKACE MODELŮ

Odpadové hospodářství (OH) je považováno za veřejnou službu zprostředkovávající občanům systém nakládání s odpady ekologickou a ekonomickou cestou. Základní informace, tj. množství a složení odpadu, jsou v OH využívány pro plánování týkající se svozu odpadu, infrastruktury a zpracovatelských zařízení. Požadavky na výše popsaná data rostou jak přirozenou cestou (snaha o ekonomičtější a ekologičtější nakládání s odpady), tak i cestou zákonnou (nařízení EU či jednotlivých států)[1]. Plánování má zásadní vliv na využití personálních zdrojů, svozové techniky a provozní náklady s ohledem na sběr a přepravu odpadu, což je předmětem vývoje a optimalizace systémů odpadového hospodářství [2]. Neméně důležité je také sledování a vyhodnocování systémových opatření, která cílí na snižování produkce odpadu a efektivnější recyklaci [3], jak ukazuje obrázek 1.1. Pokud ale mají data sloužit k čemukoliv spojenému s plánováním, vzniká potřeba nejen data shromažďovat a vyhodnocovat, ale také předpovídat jejich další vývoj, zvláště v dnešní rychle se měnící době. Z tohoto důvodu během posledních 40 let vznikly desítky studií, které se méně či více úspěšně pokusily o předpovědi množství a složení produkovaného odpadu pomocí matematických modelů. Mezi největší problémy při jejich návrhu patří nedostatečné množství kvalitních dat a problematické měření produkovaného odpadu [1]. Kvalitními daty se zde rozumí souhrn socioekonomických, demografických případně jiných vhodných faktorů, které by mohly mít vliv na produkci odpadu, a to z pohledu kvality i kvantity. Mělo by se jednat o data konzistentní, která by navíc měla být shromažďována dostatečně dlouhou dobu. Měření odpadu je problematické v tom, že odpad lze přímo měřit pouze výjimečně na úrovni domácností díky tzv. pay as you throw systémům [1]. Téměř vždy je tedy odpad měřen nepřímo, na rozdíl například od elektřiny, což nutně vede k nepřesnostem. Velmi cenný souhrn přístupů k predikování v oblasti OH přináší [1], tyto poznatky byly doplněny o závěry z novějších studií [2]-[11].

Obr. 1.1: Hierarchie nakládání s odpady[12]

1.1 Toky odpadu

Většina modelů popisuje produkci komunálního odpadu (KO) jako jednu závisle proměnnou [1]. Rozložení na jednotlivé složky odpadu je většinou dosti nákladné a nepřináší užitečné informace kvůli nemožnosti identifikace vlivů faktorů na změny v množství jednotlivých složek [1]. Zkoumané toky odpadu často nejsou jasně definovány, takže bývá problémem zjistit, které z nich byly zahrnuty. Zkreslení toků KO spojené s dalšími zdroji v podobě komerčního odpadu a turismu nebo aktivity spojené s jinou formou využití odpadu (použití odpadu na vytápění v domácnosti, nelegální skládkování) zůstávají skryté, leč mohou být úspěšně odhadnuty užitím vhodných zástupných proměnných [4].

Nejkomplexnějším přístupem je sledování toků materiálu, při němž se adresují všechny odpady pocházející od zpracovatelů a to prostřednictvím input-output analýzy. Tato metoda vzhledem ke své povaze neuvažuje použitý proces sběru odpadu a záznamy o sběru jsou využívány pouze pro ověřování [1]. Dalším možným přístupem je sledování shromažďovacích toků. Oficiální odpadové statistiky jsou používány převážně pro modelování celkového množství KO nebo pro modelování jednotlivých toků odpadu [1]. Případně pro modelování celkového množství recyklovatelných odpadů či množství jednotlivých materiálů (papír a karton, sklo, plasty, kovy)[1]. Tento přístup však nedokáže zachytit ostatní možnosti nakládání s odpadem jako spalování odpadu a nelegální skládkování [1]. Posledním uváděným konceptem jsou složky odpadu z domácností. Takovéto modely jsou založené na analýze třídění směsného nebo zbytkového odpadu, respektive analyzují složení odpadu posbíraného z popelnic a kontejnerů.

1.2 Územní členění

Modely jsou také charakterizovány použitým uzemním členěním, které odkazuje na nejmenší identifikovatelnou územní jednotku. Typicky se používají existující administrativní jednotky (domácnost, místní část, obec, okres, kraj, stát), výjimku tvoří případy, kdy autor usoudí, že tyto jednotky nejsou z pohledu dat homogenní [1]. Při volbě územního členění působí dva protichůdné vlivy, na jednu stranu je snaha o co nejjemnější rozlišovací schopnost modelu, na stranu druhou však je třeba udržet rozumný počet sledovaných objektů. Výběr několika nadměrně velkých nebo mnoha malých územních jednotek může významně ovlivnit užitečnost a efektivitu nákladů na výzkum. V přehledu [1] používá 31, z celkového počtu 45, studií členění, jehož základní jednotkou jsou okresy, případně menší územní jednotky. V kontrastu s tímto faktem působí diskutabilně využívání modelů na úrovni států zodpovědnými lidmi

[1]. Nalezení nejvýznamnější nezávisle proměnné je často realizováno skrze vztahy mezi časovými řadami produkce odpadu a daným faktorem, ačkoliv nebyl tento fakt potvrzen průřezovými analýzami [5], [6].

Domácnosti: Při této volbě hrají nikoliv nedůležitou roli náklady, na úrovni domácností může oproti ostatním dojít k výraznému nárůstu nákladů, které nejsou úměrné získaným informacím. Na druhou stranu studie týkající se domácností mohou odhalit vztahy mezi produkcí odpadu a rozsáhlým souborem individuálních charakteristik či zvyků jednotlivých domácností. Domácnosti jsou typicky rozděleny podle příjmů, věku nebo úrovně vzdělání. Kvůli velké náročnosti trvá sledování většinou 3 týdny až 6 měsíců [1] a data jsou získávána formou rozhovorů, případně dotazníků.

Místní části: Studie na této úrovni vykazují pozitivní zkušenosti s ohledem na vztahy mezi složením obyvatelstva a charakteristikami produkce odpadu, což hovoří pro výběr homogenních částí obcí jako základní územní jednotky. Homogenita hustoty osídlení a typů příbytků jsou většinou považovány za dostatečnou záruku proměnných typu příjmu, zaměstnaneckého statusu a velikosti domácnosti. Vybraná oblast často koresponduje s nejmenšími administrativními jednotkami, kterými bývají volební okresky s několika stovkami obyvatel, větší jednotky se používají pouze zřídka. Analýza odpadu z domácností typicky zahrnuje dokumentaci shromážděného odpadu a třídící analýzy vzorků z vybraných okruhů sběru.

Regiony: Do této kategorie spadají studie, jejichž základními územními jednotkami jsou veškeré jednotky od úrovně obce až po kraje. I díky takto širokému záběru se jedná o nejrozšířenější typ územního členění. Dalšími důvody častého využití jsou použitelnost pro regionální plánování a možnost okamžitého použití dat. Tato volba také umožňuje pokrytí velkého území prostřednictvím malých či středně velkých obcí. Některé studie na této úrovni využívají časových řad na měsíční nebo denní bázi. Jako data popisující produkci odpadu jsou používány statistiky o množství odpadu, sporadicky pak analýzy třídění odpadu. Pro modelování nezávislých proměnných slouží data ze sčítání lidu, ekonomická data ve spojení s daty z OH a také názory expertů [7].

Státy: Modely na této, nejvyšší používané, úrovni se dělí na 3 základní typy: input-output analýzy, průřezové analýzy a analýzy časových řad. První zmíněný typ cílí na odhady toků hlavních složek odpadu (například plasty, papír, dřevo) v dané zemi. Další dvě metody jsou metody regresní a zaměřují se na porovnání mezi zeměmi, případně jednotlivými roky. Mezi hlavní zdroje dat se řadí souhrny z OH, data

ze sčítání lidu, ekonomická data shromažďovaná statistickými agenturami a data poskytovaná asociacemi z oblasti průmyslu a obchodu [1].

1.3 Nezávisle proměnné

Již ve fázi návrhu je často možno odhadnout, zda navrhovaný model s charakteristickou skupinou faktorů bude schopen uspokojit základní informační potřeby, tedy dostupnost dat ve formě časových řad, aplikovatelnost pro predikce a udržitelnou kvalitu dat [1]. Schopnost rychlé reakce na nové trendy v produkci odpadu vyžaduje použití modelů založených na aktuálních datech [1]. Modely používající rozsáhlé databáze s velkým počtem navrhovaných faktorů (input-output analýzy, vícenásobné regresní modely) nejsou schopny podávat relevantní aktuální informace, především z důvodu zdlouhavého procesu získávání potřebných dat [1]. Hlavním cílem modelů našeho typu je zprostředkování nástroje pro předpovídání produkce odpadu, často ale naráží na problém nedostatku zásadních dat pro vytvořený model. K tomuto dochází v případě modelů, jejichž nezávisle proměnné jsou sbírány například pouze při sčítání lidu. Velký počet nezávisle proměnných také přináší problém s garancí úrovně kvality dat.

Jedno z možných dělení nezávisle proměnných, neboli faktorů ovlivňujících závisle proměnnou v podobě produkce odpadu, uváděných v literatuře, je dělení na tzv. horizontální a vertikální faktory. Horizontální faktory popisují procesy změn mezi jednotlivými druhy odpadu. Například přesuny mezi zbytkovým, objemným, recyklovaným a ilegálně likvidovaným odpadem jsou většinou způsobeny různými způsoby sběru jednotlivých druhů odpadů a neovlivňují celkové množství odpadu. Naproti tomu vertikální faktory popisují změny celkového množství všech toků odpadu v závislosti na demografickém, ekonomickém, technickém a sociálním vývoji [1].

Jak již bylo zmíněno, při hledání závislostí mezi faktory a produkcí či složením KO jsou velmi užitečné studie na nejmenších územních jednotkách. Na jejich základě je tok odpadu (ať už jako celku nebo jednotlivých součástí) rozdělen podle životního cyklu výrobku na odpady z výrobní a prodejní, spotřební a likvidační části [1]. Data ohledně výroby a prodeje obsahují přímé i nepřímé informace o množství výrobků a tocích odpadu při jejich zpracování. Z této oblasti ale obvykle nejsou dostupná data o množství, převážně tak dochází ke konverzi peněžních údajů na fyzická data formou dotazníků, předpokladů nebo statistických odhadů [1].

Skupina proměnných spojená se spotřebou odráží vztahy mezi životními podmín-

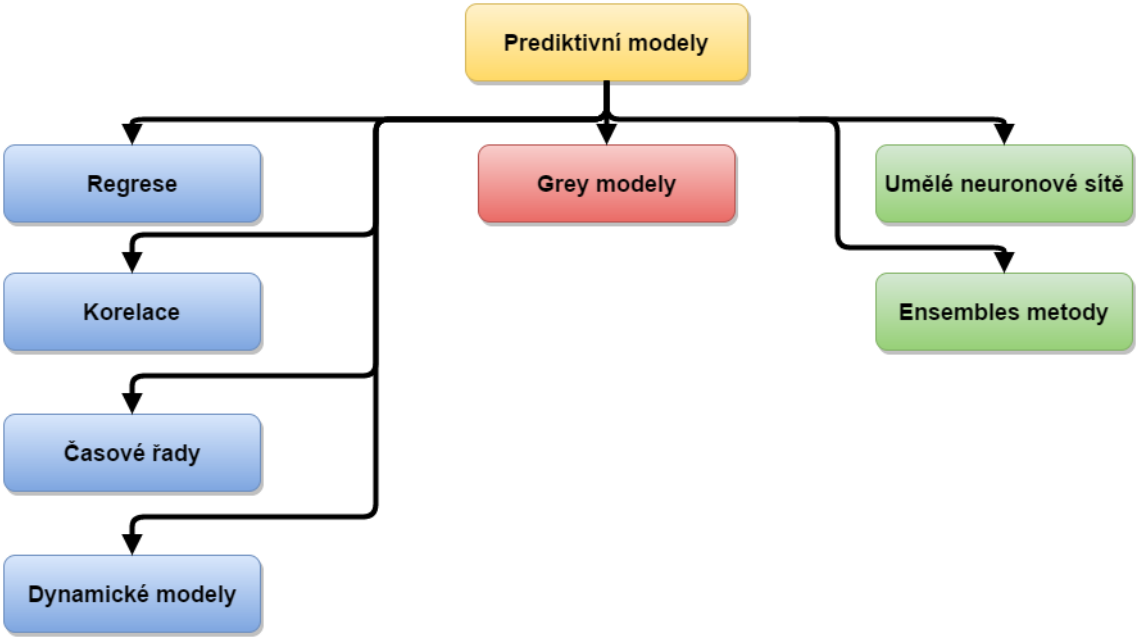
kami a strukturami produkce odpadu. Dobře jsou zdokumentovány vlivy místního obyvatelstva, hůře už pak vliv turismu, tyto faktory tak slouží jako obecné ukazatele životní úrovně (především příjem, velikost majetku, rychlost jeho růstu a životní náklady)[1]. Mezi další faktory patřící do této kategorie řadíme druh obydlí, postavení v zaměstnání, hustota zalidnění, úroveň urbanizace, délka života a dětská úmrtnost. Významné nezávisle proměnné na úrovni domácností jsou velikost, věkové složení, životní etapa domácnosti a spotřební návyky, tyto ukazatele jsou získávány prostřednictvím dotazníků [1].

Poslední kategorie nezávisle proměnných (spojených s likvidací výrobku) by měla ovlivňovat závisle proměnnou hlavně v horizontálním smyslu. Jako náhrada pro podíl komerčního odpadu byly v minulosti úspěšně použity zaměstnanost podle sektoru a data o prodeji z jednotlivých oborů. Významný vliv na množství recyklovaného odpadu mají způsob vytápění domácností, osvěta v oblasti recyklování, velikost kontejnerů, hustota sběrných míst a poplatky za KO [1].

1.4 Modelovací metody

Navzdory tomu, že řešené problémy jsou si v jednotlivých studiích velmi podobné, pro jejich řešení se používá pestrá škála modelovacích technik na různých úrovních komplexnosti. Přehled [1] uvádí 7 skupin modelů, od té doby ale došlo k výraznému rozvoji stávajících technik i použití zcela nových. Tyto modely se liší především v počtu nezávisle proměnných, způsobu ověřování modelu a jejich použitelnosti pro predikce. Modelovací metody je možné dělit jak podle počtu nezávisle proměnných, tak i podle míry čitelnosti vnitřního fungování modelu, viz obrázek 1.2. Modely uvažující pouze jednu nezávisle proměnnou mají hlavní výhodu v tom, že mohou být snadno ověřovány na reálných datech. Některé z nich je také možné rozšířit na více (většinou 2 až 5) nezávisle proměnných, tyto modely jsou komplexnější, jejich nevýhodou je však obtížnost jejich ověřování, které je v některých případech zcela nemožné [1].

Četné využití regresních modelů je dáno vyspělým teoretickým základem a jednoduchostí algoritmu [8], má však také několik nevýhod, mezi které patří chybějící schopnost učení se z nových dat, stejně tak adaptace na nové situace, špatné výsledky při použití nepřesných dat a také to, že nebere v potaz všechny faktory ovlivňující produkci odpadu [2]. Lepší výsledky zprostředkovává analýza časových řad, jejíž hlavní doménou je uvažování sezónních vlivů na produkci odpadu. Na druhou stranu pro predikce v krátkodobém horizontu potřebuje velké množství dat a na



Obr. 1.2: Rozdělení modelů podle míry čitelnosti vnitřního fungování modelu

základě výsledků této metody nelze vyvozovat obecné závěry [2]. Poměrně novým přístupem je používání umělých neuronových sítí, ty obecně vykazují lepší výsledky než regresní modely i modely používající časové řady a to díky schopnosti se učit a konstruovat komplexní nelineární systémy. Hrozba přeučení, obtížné rozeznávání stavby sítě, lokální minima a obtížná generalizace však zůstávají hlavními překážkami v aplikování umělých neuronových sítí pro praktické problémy [2].

Dalším krokem ve vývoji modelů je tzv. Grey model (GM), který řeší problémy s nedostatkem dat a komplexností prediktivních modelů. Tyto modely bývají zapisovány jako GM (m, n), kde m je řád diferenciální rovnice a n počet proměnných. GM byly úspěšně implementovány pro dlouhodobé předpovědi a dosahují zde lepší přesnosti, než modely využívající časové řady nebo umělé neuronové sítě. Výhodou GM oproti regresnímu je schopnost akceptovat menší množství významných odchylek, s výjimkou modelů typu GM (1, n) a GMC [1] (n, 1) [2]. K vytvoření přesného modelu je nutné nejen vybrat správné faktory, ale také je přesně odhadnout.

[1] GM s konvolučním integrálem

Tab. 1.1: Přehled vybraných studií

Země	Rok	Územní členění	Modely	Nezávisle proměnné
Írán[7]	2015	města	ANN, MLR	počet obyvatel
				frekvence sběru odpadu
				maximální teplota
				nadmořská výška
Thajsko[3]	2014	stát	GM, GMC	spotřební výdaje na domácnost
				počet členů domácnosti
				zaměstnanost
				hustota zalidnění
				úroveň urbanizace
Čína[8]	2013	město	GM+SARIMA	historická data o produkci odpadu
Turecko[6]	2011	provincie	SAR	nezaměstnanost
				podíl asfaltovaných cest
			GWR	nezaměstnanost
				teplota
				podíl vysokoškolsky vzdělaných
				hodnota zemědělské produkce
Rakousko[2]	2010	města	LR	počet členů domácnosti
				podíl staveb s vytápěním na tuhá paliva
				daňové příjmy obce
Španělsko[4]	2010	města	GLM, BR	počet imigrantů na 100 obyvatel
				podíl vysokoškolsky vzdělaných
				HDP
				index spotřebitelských cen
				životní výdaje na osobu
				počet přespání turistů na 100 obyvatel

ANN-Umělé neuronové sítě, MLR-Vícenásobná lineární regrese, LR-Lineární regrese, GM-Grey model, GMC-Grey model s konvolučním integrálem, SARIMA-seasonal autoregressive integrated moving average, SAR-Simultaneous spatial autoregression, GWR-Geographically weighted regression, GLM-Zobecněné lineární modely, BR-β regrese

2 MATEMATICKÝ APARÁT

V [14] se uvádí, že význam samotného slova regrese se dá vyložit mj. jako zpětný pochod, postup, ústup či návrat k některé z předešlých vývojových forem související se zjednodušením namísto očekávaného zdokonalení (zejména v biologii). Tento popis může působit mírně dehonestujícím dojmem, ale jak nám ukazuje život nebo matematika sama, například zmíněné zjednodušení je někdy naopak velmi vítaným jevem. Paradoxem je také fakt, že použití regrese v matematice a zejména v jejích aplikacích může znamenat značný progres neboli pokrok. Následující řádky jsou výtahem, který vychází převážně z poznatků [15] a [16], ty jsou dále doplněny o informace z [17].

2.1 Lineární model

Předpokládejme, že střední hodnoty nekorelovaných náhodných veličin $Y_1, ..., Y_n$ je možné popsat lineární funkcí s pomocí $k+1$ neznámých parametrů

$$\mathsf{E}\, Y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}, \tag{2.1}$$

kde x_{ij} jsou známé konstanty. Dále budeme předpokládat $varY_i = \sigma^2$ pro všechna i, σ je dalším, zpravidla neznámým, parametrem. Známé konstanty x_{ij} uspořádáme do matice o n řádcích a $k+1$ sloupcích

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \tag{2.2}$$

tuto matici nazýváme *regresní maticí* či *maticí modelu*. Pro hodnost matice $\mathbf{X}$ platí $h(\mathbf{X}) = r > 0$ a $n > r$. Náhodný vektor $\mathbf{Y}$ má pak střední hodnotu $\mathbf{X}\boldsymbol{\beta}$ a varianční matici $\sigma^2\mathbf{I}$. Uvedené předpoklady budeme dále zapisovat jako $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. V případě $k = 1$ hovoříme o *jednoduché* lineární regresi, o *mnohonásobnou* lineární regresi by se pak jednalo v případě $k > 1$. Pokud se snažíme vysvětlit chování několika složek vektorové veličiny $\mathbf{Y}_i$, místo skalární vysvětlované veličiny Y_i, mluvíme o *mnohorozměrné* regresi.

Dále bude používáno následující označení. Nechť sloupce matice $\mathbf{Q}$ tvoří nějakou ortonormální bázi *regresního prostoru* $\mathcal{M}(\mathbf{X})$, nechť sloupce matice $\mathbf{N}$ doplní tuto bázi na ortonormální bázi prostoru $\mathbb{R}^n$. Obdržíme tak ortonormální matici

$\mathbf{P} = (\mathbf{Q},\mathbf{N})$ takovou, že $\mathcal{M}(\mathbf{X}) = \mathcal{M}(\mathbf{Q})$, $\mathbf{PP}' = \mathbf{I}_n$. Z ortonormálnosti sloupců matice $\mathbf{P}$ plynou vztahy

$$\mathbf{QQ}' + \mathbf{NN}' = \mathbf{I}_n, \qquad \mathbf{QQ}' = \mathbf{I}_r, \qquad \mathbf{NN}' = \mathbf{I}_{n-r}, \qquad \mathbf{Q}'\mathbf{N} = \mathbf{O}.$$

Označme $\mathbf{H} = \mathbf{QQ}'$ a $\mathbf{M} = \mathbf{NN}'$. Takto zavedené matice jsou symetrické a idempotentní. Jelikož platí $\mathbf{HM} = \mathbf{O}$, jsou vektory na pravé straně vztahu

$$\mathbf{y} = \mathbf{Hy} + \mathbf{My}$$

navzájem ortogonální, jde tedy o průměty obecného vektoru $\mathbf{y} \in \mathbb{R}^n$ do regresního prostoru $\mathcal{M}(\mathbf{X})$ a na něj kolmého *reziduálního prostoru* $\mathcal{M}(\mathbf{X})^\perp$. Z vlastností projekce plyne jednoznačnost těchto průmětů a tím také jednoznačnost projekčních matic $\mathbf{H}$ a $\mathbf{M}$.

V dalším textu bude praktické znát vztahy

$$\mathbf{HX} = \mathbf{X} \tag{2.3}$$

a také

$$\mathbf{MX} = \mathbf{O}. \tag{2.4}$$

2.2 Odhad vektoru středních hodnot

Nejprve se budeme zabývat odhadem vektoru středních hodnot $\mu = \mathbf{X}\boldsymbol{\beta}$. V prostoru $\mathcal{M}(\mathbf{X})$ najdeme nejbližší vektor k náhodnému vektoru $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$ a označíme jej $\hat{\mathbf{Y}}$.

Věta 1 (**Gaussova-Markovova**)**.** *V Modelu $\boldsymbol{Y} \sim (\boldsymbol{X\beta},\sigma^2\boldsymbol{I})$ je $\hat{\boldsymbol{Y}}$ nejlepším nestranným lineárním odhadem vektoru $\boldsymbol{X\beta}$.*

Důkaz uvádí [16].

2.3 Rezidua

Nyní se budeme věnovat průmětu vektoru $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$ do prostoru reziduí $\mathcal{M}(\mathbf{X})^\perp$ a bude zaveden nestranný odhad rozptylu σ^2. *Vektor reziduí* definovaný vztahem $\mathbf{u} = \mathbf{Y}$ - $\hat{\mathbf{Y}}$ porovnává napozorované hodnoty vysvětlované proměnné s odhadem jejích středních hodnot. *Reziduální součet čtverců* $RSS = \|\mathbf{u}\|^2 = \sum_{i=1}^{n} \left(\mathbf{Y}_i - \hat{\mathbf{Y}}_i\right)^2$ udává čtverec vzdálenosti vektorů $\mathbf{Y}$ a $\hat{\mathbf{Y}}$ a tím popisuje jejich odlišnost. *Reziduální rozptyl* pak zavedeme jako $S^2 = RSS/(n-r)$.

Věta 2 (**O reziduích**). *V lineárním modelu* $\boldsymbol{Y} \sim (\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ *platí*

$$\boldsymbol{u} = \boldsymbol{MY} = \boldsymbol{Me}, \tag{2.5}$$

$$\boldsymbol{u} \sim (\boldsymbol{0}, \sigma^2\boldsymbol{M}), \tag{2.6}$$

$$RSS = \boldsymbol{e}'\boldsymbol{Me}, \tag{2.7}$$

$$\mathsf{E}RSS = (n-r)\sigma^2, \tag{2.8}$$

$$\mathsf{E}\, S^2 = \sigma^2, \tag{2.9}$$

$$\boldsymbol{X}'\boldsymbol{u} = \boldsymbol{0}. \tag{2.10}$$

Důkaz tvrzení předcházející věty je možno nalézt v [16]. Poznamenejme, že vektor $\mathbf{u}$ je možno interpretovat jako odhad náhodné složky vektoru $\mathbf{e} = \mathbf{Y}$ - $\mathbf{X}\boldsymbol{\beta}$. Podle tvrzení (2.9) předchozí věty je reziduální rozptyl S^2 nestranným odhadem rozptylu σ^2.

2.4 Normální rovnice

Dosud nebyl řešen odhad vektoru $\boldsymbol{\beta}$ vyjadřujícího střední hodnotu náhodného vektoru $\mathbf{Y}$ formou lineární kombinace sloupců regresní matice $\mathbf{X}$. Dále předpokládejme lineární nezávislost sloupců matice $\mathbf{X}$. Symbolem $\mathbf{b}$ označme řešení soustavy

$$\mathbf{Xb} = \hat{\mathbf{Y}},$$

vektor $\mathbf{b}$ pak tvoří hledané koeficienty lineární kombinace. Skutečnost, že

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{u}$$

je ortogonálním rozkladem, je ekvivalentní s požadavkem ortogonality $\mathbf{u}$ vůči regresnímu prostoru $\mathcal{M}(\mathbf{X})^{\perp}$, který lze zapsat jako

$$\mathbf{X}'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{0},$$

to je ekvivalentní s *normální rovnicí* pro $\mathbf{b}$

$$\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{Y}. \tag{2.11}$$

Poznamenejme, že existence řešení normální rovnice je zaručena přítomností lineární kombinace řádků matice $\mathbf{X}$ na obou stranách vztahu (2.11), jednoznačnost řešení normální rovnice však není zaručena.

2.5 Normální lineární model s plnou hodností

Předpokládejme, že má náhodný vektor $\mathbf{Y}$ normální rozdělení, což lze zapsat jako $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$. Tento model pak nazveme *normálním lineárním modelem*. Platí-li navíc $r = \mathbf{h(X)} = k+1$ (lineární nezávislost sloupců matice $\mathbf{X}$), hovoříme o *regulárním lineárním modelu*. Pokud je lineární model regulární, je zaručena jednoznačnost řešení normální rovnice (2.11).

Věta 3 (**Klasický model mnohonásobné lineární regrese**). *Má-li matice $\boldsymbol{X}$ v normálním modelu $\boldsymbol{Y} \sim (\boldsymbol{X\beta},\sigma^2\boldsymbol{I})$ hodnost rovnu počtu jejích sloupců, pak platí*
a) řešením normální rovnice je statistika

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}; \tag{2.12}$$

b) $\boldsymbol{b}$ je nejlepší nestranný lineární odhad vektoru $\boldsymbol{\beta}$;
c)označíme-li $\boldsymbol{V} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$ (s indexy $0 \leq i,j \leq k$), pak platí

$$\boldsymbol{b} \sim (\boldsymbol{\beta},\sigma^2\,\boldsymbol{V});$$

d) náhodné vektory $\boldsymbol{b}$ a $\boldsymbol{u}$ jsou nezávislé;
e) statistiky $\boldsymbol{b}$ a S^2 jsou nezávislé;
f) pro j=0,1,...,k platí

$$T_j = \frac{b_j - \beta_j}{S\sqrt{v_{jj}}} \sim t_{n-k-1}; \tag{2.13}$$

g) interval

$$(b_j - S\sqrt{v_{jj}}t_{n-k-1}(\alpha); b_j + S\sqrt{v_{jj}}t_{n-k-1}(\alpha)) \tag{2.14}$$

tvoří interval spolehlivosti pro β_j se spolehlivostí 1 - α;
h)množina

$$\mathcal{K} = \{\boldsymbol{\beta} \in \mathbb{R}^{k+1} : (\boldsymbol{\beta} - \boldsymbol{b})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{b}) < (k+1)S^2F_{k+1,n-k-1(\alpha)}\} \tag{2.15}$$

tvoří konfidenční množinu pro $\boldsymbol{\beta}$ se spolehlivostí 1 - α.

Důkaz je možno nalézt v [16].

2.6 Koeficient determinace

Koeficient determinace R^2 je nejčastěji definován vztahem

$$R^2 = 1 - \frac{RSS}{\sum(Y_i - \bar{Y})^2}. \tag{2.16}$$

Hodnota R^2 je v případě lineárního modelu shodná se čtvercem výběrového koeficientu mnohonásobné korelace spočítaného z vektoru $\mathbf{Y}$ a odpovídajících netriviálních (nekonstantních) sloupců matice $\mathbf{X}$. Koeficient determinace je ukazatelem vyjadřujícím velikost dílu výchozí variability hodnot závisle proměnné charakterizované výrazem

$$SS_T = \sum(Y_i - \bar{Y})^2 = \|\mathbf{Y} - \bar{Y}\mathbf{1}\|^2 = \| \mathbf{u_0} \|^2.$$

Variabilita vysvětlená modelem je dána vztahem

$$SS_R = \sum(\hat{Y}_i - \bar{Y})^2 = \|\hat{\mathbf{Y}} - \bar{Y}\mathbf{1}\|^2 = \|\mathbf{d}\|^2.$$

2.6.1 Korigovaný koeficient determinace

Hodnota R^2 roste s počtem regresorů, což uměle zvyšuje přesnost modelu. Z tohoto důvodu se zavádí *korigovaný koeficient determinace*

$$\bar{R}^2 = 1 - \frac{n-1}{n-r}(1 - R^2), \tag{2.17}$$

kde $r = k - 1$, připomeňme, že se jedná o rozměry regresní matice. Pro oba zmíněné koeficienty platí, že čím více se blíží jedné, tím těsnější je regresní závislost. Z hlediska aplikace R^2 a $\bar{R}^2$ je důležité si povšimnout, že uvedená tvrzení nepožadují předpoklad normality.

2.7 Korelace

Nechť X a Y jsou náhodné veličiny s konečnými druhými momenty a kladnými rozptyly. Vzájemná závislost X a Y se často měří pomocí *korelačního koeficientu*

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{(var\ X)(var\ Y)}}. \tag{2.18}$$

Přesněji řečeno se jedná o tzv. *Pearsonův korelační koeficient* popisující lineární závislost náhodných veličin. Snadno se nahlédne, že díky Schwarzově nerovnosti platí $-1 \leq \rho_{X,Y} \leq 1$. Přičemž kladné hodnoty označují přímou lineární závislost, záporné pak nepřímou. Pokud nastane případ $\rho_{X,Y} = 0$, říkáme, že X a Y jsou lineárně nezávislé.

2.7.1 Výběrový korelační koeficient

Mějme náhodný výběr

$$(X_1, Y_1)', \cdots, (X_n, Y_n)' \tag{2.19}$$

z nějakého dvojrozměrného rozdělení. Charakteristiky výběru X_1,... ,X_n označme jako $\bar{X}$ a S_X^2, obdobně charakteristiky výběru Y_1,... ,Y_n označme $\bar{Y}$ a S_Y^2. Dále definujme

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}). \tag{2.20}$$

Je-li $S_X^2 > 0$ a $S_Y^2 > 0$, zavedeme *výběrový korelační koeficient* r vzorcem

$$r = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}. \tag{2.21}$$

Věta 4. *Nechť (2.19) je výběrem z dvojrozměrného normálního rozdělení s kladnými rozptyly a korelačním koeficientem $\rho = 0$. Pak pro $n \geq 3$ platí*

$$T = \frac{r}{\sqrt{1 - r^2}} \sqrt{n-2} \sim t_{n-2}.$$

Důkaz uvedené věty je uveden například v [17]. Pokud chceme testovat hypotézu, že $\rho_{X,Y} = 0$, vypočteme nejprve r a pak T podle uvedeného vzorce. Kritické hodnoty pro r jsou tabelovány, ale dnešní statistický software uvádí přímo dosaženou hladinu testu.

2.7.2 Spearmanův korelační koeficient

Mezi předpoklady předchozí věty je důležité zdůraznit požadavek na normální rozdělení výběrů, ten však v praxi bývá často porušen (značně problematické je také jeho ověřování) a uvedenou větu tedy nelze použít. Je tedy nezbytné použít jinou formu ověření vzájemné závislosti náhodných veličin.

Předpokládejme, že $(X_1, Y_1)', ..., (X_n, Y_n)'$ je výběrem ze spojitého dvojrozměrného rozdělení. Označme dále $R_1, ..., R_n$ pořadí veličin $X_1, ..., X_n$ a $Q_1, ..., Q_n$ pořadí veličin $Y_1, ..., Y_n$. *Spearmanův korelační koeficient* r_S je pak definován jako výběrový korelační koeficient spočítaný z dvojic $(R_1, Q_1)', ..., (R_n, Q_n)'$ a je roven

$$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2. \tag{2.22}$$

Kritické hodnoty $r_S(\alpha)$ je možno nalézt ve statistických tabulkách. Pro $n > 30$ se pak využívá asymptotická normalita koeficientu r_S. Je dána vztahem

$$r_S(\alpha) = \frac{u\left(\frac{\alpha}{2}\right)}{\sqrt{n-1}}. \tag{2.23}$$

Hypotéza nezávislosti se pak zamítá pro $|r_S| \geq r_S^*(\alpha)$. Je také důležité si povšimnout, že Spearmanův korelační koeficient není přímou náhradou Pearsonova, nepopisuje totiž linearitu závislosti náhodných veličin, ale monotonii.

3 NÁVRH MODELU

V této kapitole bude popsán samotný proces tvorby regresních modelů. Konkrétně půjde o lineární regresní modely, ty byly vybrány především pro, již dříve, zmíněné výhody v podobě jednoduchosti a pevných matematických základů. Při tvorbě modelu budou využity hluboké poznatky z oblasti OH navržené [1], doplněné o postupy uvedené v [2]-[11]. Je vhodné zdůraznit důležitost [3], která využívá také lineární regresi, srovnatelné územní členění (úroveň obcí) a zkoumaná oblast (Štýrsko) je navíc i geograficky blízká České republice. Navíc budou využity i teoretické poznatky z [16] a v neposlední řadě také informace o vhodných statistických nástrojích, integrovaných do softwaru [18]-[20]. Nejprve bude provedena analýza dat, která bude mít za úkol konfrontovat intuitivní předpoklady o vzájemných závislostech jednotlivých proměnných. Poté budou za pomocí softwaru vytvořeny jednotlivé modely, které budou dále zhodnoceny.

3.1 Analýza dat

Na základě poznatků shrnutých v prvé kapitole a dostupnosti dat byla zvolena práce s daty na úrovni obcí s rozšířenou působností (ORP). Veškerá data o produkci odpadu pocházejí z Informačního systému odpadového hospodářství (ISOH), všechna ostatní pak z Českého statistického úřadu (ČSÚ). ISOH zprostředkoval data za roky 2009 až 2013. Data pocházející z ČSÚ byla dostupná za roky 1993-2013, v roce 2012 byla provedena změna v jejich publikaci, která způsobila absenci některých dat pro rok 2013 a naopak přidání některých, které k dispozici nebyly (pro rok 2012 jsou dostupné obě varianty).

Tento fakt má za následek, že po celé zkoumané období (2009-2013) jsou dostupné pouze tyto faktory: počet obyvatel (OBYV), počet obyvatel ve věku 0-14 let (0-14), počet obyvatel ve věku 15-64 let (15-64), počet obyvatel ve věku 65 a více let (65+), počet dokončených bytů(BYT). Z OBYV a rozlohy jednotlivých ORP pak byla vypočtena ještě hustota zalidnění (HZ). Dále byla použita data o počtu osob žijících na jedné adrese (OBADR) v jednotlivých ORP. Ta jsou však, z pochopitelných důvodů, dostupná pouze za rok 2011 a to díky sčítání lidu, domů a bytů. Problémem s OBADR bylo, že zde bylo zastoupeno celkem 551 kategorií[1]. Takovýto počet je zřejmě pro další analyzování nevhodný a bylo tedy nezbytné jej vhodným způsobem zredukovat. Toho bylo docíleno prostřednictvím shlukové analýzy. Shluková analýza

[1]od kategorie zahrnující neobydlené adresy a adresy s jedinou osobou až po kategorii s 1101 a 1102 obyvateli

byla provedena v programu STATISTICA 12, dle pokynů [18], [19] a její výsledky shrnuje tabulka 3.1 .

Tab. 3.1: Výsledky shlukové analýzy OBADR

Skupina	Zahrnuté intervaly
A	[0,2[
B	[2,4[
C	[4,6[
D	[24,26[∪ [46,48[∪ [48,50[
E	ostatní

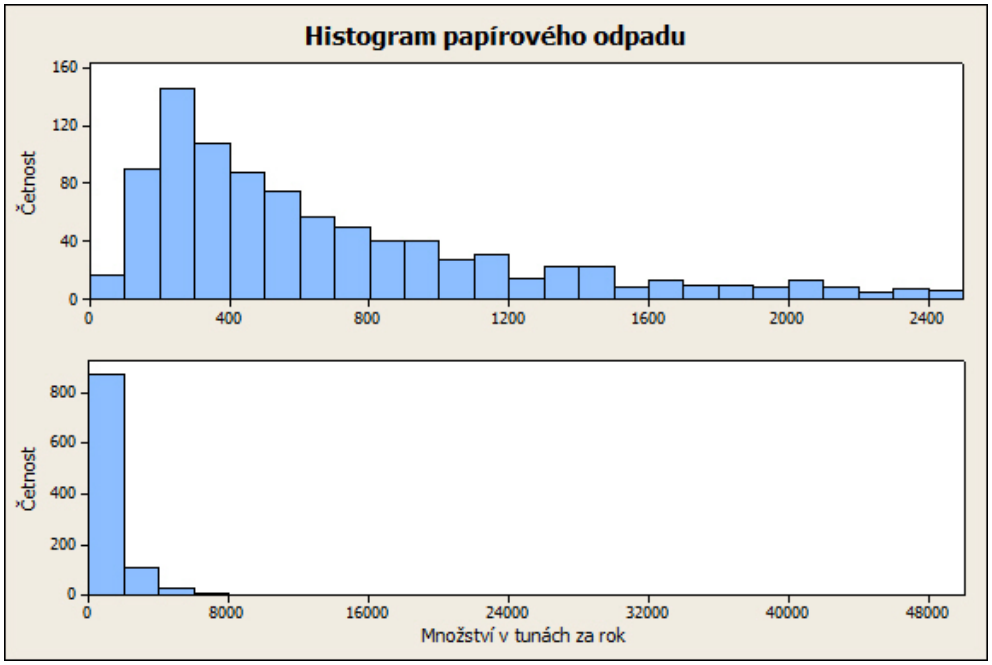
Údaje z tabulky 3.1 a data o věkovém složení obyvatelstva byly přepočteny z absolutních hodnot na relativní, tak aby se předešlo možné multikolinearitě s počtem obyvatel [1]. Data z ISOH zahrnují informace o množství jednotlivých druhů odpadu, ze kterých bylo vypočteno i celkové množství KO. Z těchto dat pak byly odvozeny přepočty na osobu. Výčet všech nezávisle proměnných je uveden v tabulce 3.2. Autoři [1] uvádějí, že rozložení KO na jednotlivé složky je nákladné a nepřináší cenné informace. Toto však provedeno bude, mj. kvůli požadavku na možnost analyzování potenciálu třídění jednotlivých ORP.

Tab. 3.2: Přehled závisle proměnných

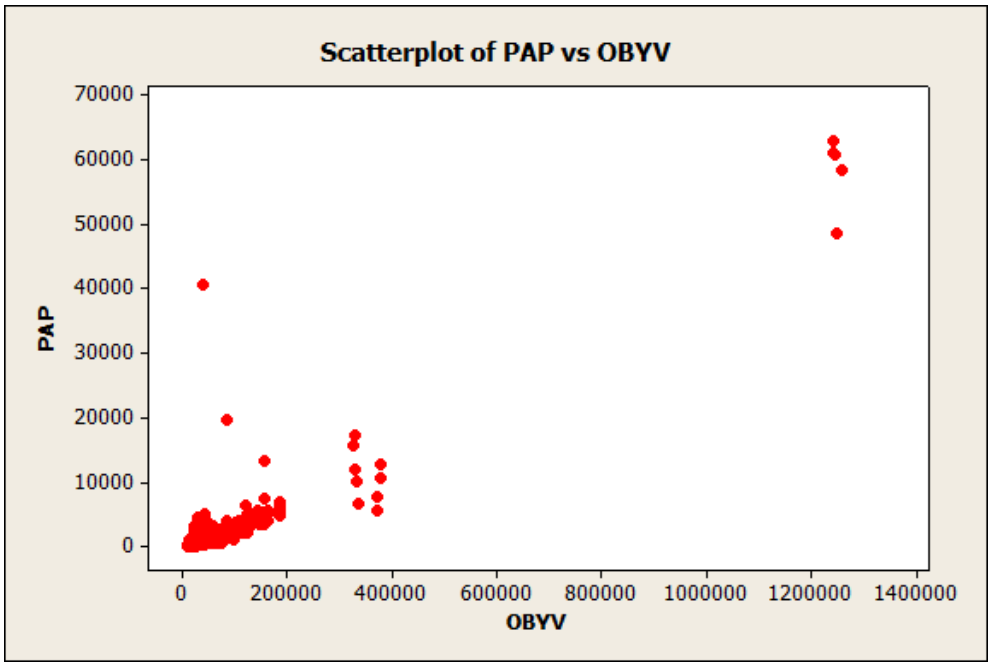
	Zkratky	
Druh odpadu	Celkové [t/rok]	Na osobu [kg/rok]
Papír	PAP	papOs
Plast	PLST	plastOs
Sklo	SKLO	sklOs
Kovy	KOV	kovOs
Bioodpad	BIO	biOs
Směsný a komunální	SKO	SKOs
Objemný	OBJ	objOs
Komunální	KO	KOs

Dalším krokem bylo prozkoumání histogramů všech dostupných dat, především za účelem vytvoření představy o rozdělení dat. Zkoumaná data většinou, podle očekávání, nevykazovala normální rozdělení, což je dokumentováno na obrázku 3.1. Na obrázku 3.2 se pak snadno vidí, že např. výběr PAP a OBYV není výběrem z dvojrozměrného normálního rozdělení (očividně je zde porušen předpoklad symetrie). Použití Pearsonova korelačního koeficientu, by tedy nebylo korektní, vzhledem k předpokladům věty 4. Pro korelační analýzu tak byl podobně jako v [3] využit

Spearmanův pořadový korelační koeficient. Korelace byla vypočtena užitím softwaru STATISTICA 12 na hladině významnosti α=0,05 a její výsledky jsou zobrazeny v tabulce 3.3.



Obr. 3.1: Histogram



Obr. 3.2: Závislost PAP na OBYV

Korelace významné na hladině α=0,05	HZ	OBYV	0-14	15-64	65+	BYTY	A	B	C	D	E	PAP	PLST	SKLO	KOV	BIO	SKO	OBJ	Papos	Plastos	Sklos	Kovos	Bios	SKOs	Objos
HZ	1,00	0,51	0,14	0,03	-0,13	0,39	-0,71	-0,01	0,25	0,25	0,34	0,43	0,38	0,38	0,36	0,45	0,46	0,54	0,21	-0,11	-0,25	0,14	0,22	-0,11	0,30
OBYV	0,51	1,00	0,05	0,08	-0,10	0,76	-0,34	-0,18	-0,08	0,21	0,45	0,87	0,85	0,89	0,59	0,54	0,93	0,74	0,44	-0,06	-0,20	0,18	0,15	-0,09	0,20
0-14	0,14	0,05	1,00	-0,12	-0,47	0,09	-0,19	0,11	0,22	0,00	0,09	0,13	0,13	0,03	0,01	0,02	0,05	-0,02	0,19	0,19	-0,05	0,02	0,01	0,02	-0,11
15-64	0,03	0,08	-0,12	1,00	-0,77	0,02	-0,27	-0,07	-0,11	-0,04	0,27	-0,03	-0,07	-0,08	-0,09	-0,17	0,06	0,09	-0,18	-0,29	-0,35	-0,13	-0,25	0,02	0,11
65+	-0,13	-0,10	-0,47	-0,77	1,00	-0,11	0,36	-0,03	-0,01	0,06	-0,28	-0,04	-0,02	0,04	0,09	0,12	-0,10	-0,07	0,04	0,12	0,32	0,12	0,19	-0,05	-0,03
BYTY	0,39	0,76	0,09	0,02	-0,11	1,00	-0,24	0,04	0,01	0,07	0,18	0,70	0,72	0,74	0,41	0,40	0,73	0,56	0,40	0,09	-0,04	0,08	0,13	-0,01	0,13
A	-0,71	-0,34	-0,19	-0,27	0,36	-0,24	1,00	-0,23	-0,35	-0,02	-0,36	-0,20	-0,23	-0,18	-0,30	-0,27	-0,30	-0,38	0,02	0,12	0,35	-0,17	-0,10	0,08	-0,23
B	-0,01	-0,18	0,11	-0,07	-0,03	0,04	-0,23	1,00	0,42	-0,49	-0,62	-0,23	-0,11	-0,18	-0,18	-0,08	-0,17	-0,20	-0,21	0,11	0,01	-0,08	0,05	0,09	-0,14
C	0,25	-0,08	0,22	-0,11	-0,01	0,01	-0,35	0,42	1,00	-0,23	-0,23	-0,13	0,02	0,01	0,06	0,09	-0,13	-0,16	-0,13	0,22	0,19	0,16	0,15	-0,13	-0,20
D	0,25	0,21	0,00	-0,04	0,06	0,07	-0,02	-0,49	-0,23	1,00	0,36	0,28	0,20	0,21	0,11	0,14	0,23	0,26	0,29	-0,01	-0,03	-0,02	-0,02	-0,01	0,18
E	0,34	0,45	0,09	0,27	-0,28	0,18	-0,36	-0,62	-0,23	0,36	1,00	0,44	0,33	0,32	0,32	0,23	0,41	0,43	0,29	-0,14	-0,29	0,13	0,00	-0,12	0,23
PAP	0,43	0,87	0,13	-0,03	-0,04	0,70	-0,20	-0,23	-0,13	0,28	0,44	1,00	0,84	0,84	0,52	0,50	0,82	0,64	0,80	0,10	-0,08	0,15	0,15	-0,08	0,17
PLST	0,38	0,85	0,13	-0,07	-0,02	0,72	-0,23	-0,11	0,02	0,20	0,33	0,84	1,00	0,89	0,51	0,55	0,80	0,60	0,55	0,42	0,07	0,15	0,23	-0,08	0,10
SKLO	0,38	0,89	0,03	-0,08	0,04	0,74	-0,18	-0,18	0,01	0,21	0,32	0,84	0,89	1,00	0,58	0,51	0,83	0,64	0,49	0,16	0,21	0,22	0,18	-0,11	0,13
KOV	0,36	0,59	0,01	-0,09	0,09	0,41	-0,30	-0,18	0,06	0,11	0,32	0,52	0,51	0,58	1,00	0,39	0,55	0,48	0,26	-0,06	-0,06	0,88	0,12	-0,14	0,14
BIO	0,45	0,54	0,02	-0,17	0,12	0,40	-0,27	-0,08	0,09	0,14	0,23	0,50	0,55	0,51	0,39	1,00	0,51	0,46	0,30	0,15	-0,06	0,17	0,86	0,00	0,18
SKO	0,46	0,93	0,05	0,06	-0,10	0,73	-0,30	-0,17	-0,13	0,23	0,41	0,82	0,80	0,83	0,55	0,51	1,00	0,68	0,41	-0,05	-0,21	0,17	0,16	0,22	0,16
OBJ	0,54	0,74	-0,02	0,09	-0,07	0,56	-0,38	-0,20	-0,16	0,26	0,43	0,64	0,60	0,64	0,48	0,46	0,68	1,00	0,31	-0,14	-0,22	0,15	0,14	-0,16	0,78
Papos	0,21	0,44	0,19	-0,18	0,04	0,40	0,02	-0,21	-0,13	0,29	0,29	0,80	0,55	0,49	0,26	0,30	0,41	0,31	1,00	0,29	0,12	0,07	0,11	-0,05	0,06
Plastos	-0,11	-0,06	0,19	-0,29	0,12	0,09	0,12	0,11	0,22	-0,01	-0,14	0,10	0,42	0,16	-0,06	0,15	-0,05	-0,14	0,29	1,00	0,53	-0,04	0,23	0,04	-0,18
Sklos	-0,25	-0,20	-0,05	-0,35	0,32	-0,04	0,35	0,01	0,19	-0,03	-0,29	-0,08	0,07	0,21	-0,06	-0,06	-0,21	-0,22	0,12	0,53	1,00	0,02	0,10	-0,07	-0,18
Kovos	0,14	0,18	0,02	-0,13	0,12	0,08	-0,17	-0,08	0,16	-0,02	0,13	0,15	0,15	0,22	0,88	0,17	0,17	0,15	0,07	-0,04	0,02	1,00	0,08	-0,09	0,03
Bios	0,22	0,15	0,01	-0,25	0,19	0,13	-0,10	0,05	0,15	-0,02	0,00	0,15	0,23	0,18	0,12	0,86	0,16	0,14	0,11	0,23	0,10	0,08	1,00	0,07	0,06
SKOs	-0,11	-0,09	0,02	0,02	-0,05	-0,01	0,08	0,09	-0,13	-0,01	-0,12	-0,08	-0,08	-0,11	-0,14	0,00	0,22	-0,16	-0,05	0,04	-0,07	-0,09	0,07	1,00	-0,14
Objos	0,30	0,20	-0,11	0,11	-0,03	0,13	-0,23	-0,14	-0,20	0,18	0,23	0,17	0,10	0,13	0,14	0,18	0,16	0,78	0,06	-0,18	-0,18	0,03	0,06	-0,14	1,00

Tab. 3.3: Spearmanovy korelační koeficienty

3.1.1 Modely 1

Dalším krokem v analýze bylo sestavení modelů, jejichž nezávisle proměnnými byly pouze HZ, OBYV, BYT, podíly věkových skupin a skupiny A-E. Zmíněné modely, pro jednotlivé složky odpadu i KO, slouží především pro dokreslení korelační analýzy a získání představy o tom, které druhy odpadů budou modelovatelné lépe a které hůře. Byly sestaveny tzv. stepwise metodou[2] s pomocí Minitabu 15, tato metoda byla zvolena především kvůli problému se „skorosingularitou“ matice plánu. Z předchozích zkušeností kolegů z Ústavu procesního inženýrství (ÚPI) jsme věděli, že tyto modely nemají velkou vypovídající hodnotu. Stojí však za povšimnutí, že modely s odpady na osobu mají velmi malé koeficienty determinace (do 15%), zkušenosti tak byly potvrzeny.

3.1.2 Modely 2

Poté už byly modely tvořeny výhradně na datech z předchozího roku (proměnné A - E byly zahrnuté také, ale pochopitelně nepopisují časový vývoj dat). Konkrétně tedy byly analyzovány roky 2010-2012, rok 2013 byl vynechán pro pozdější možnost kontroly modelů na reálných datech. Na těchto datech byly nejdříve vytvořeny modely založené na všech dostupných datech, za podobným účelem, jako Modely 1. Tyto modely měly za cíl stanovení mezí, kam až se lze, s maticí plánu obsahující pouze lineární členy, dostat. Budeme-li brát jako kritérium hodnocení kvality jen hodnotu R^2, jedná se o velice dobré modely. U celkových množství odpadů se R^2 pohybuje (kromě KOV a BIO) nad hranicí 95 %. U množství na osobu pak (kromě plastOs a sklOs) dosahoval koeficient determinace alespoň 40 %, což sice není mnoho, ale jedná se o velký progres oproti předchozím modelům. Jedním dechem je ale nutné dodat, že modely obsahovaly běžně kolem 7 regresorů, což obecně nemusí být problém, ale v našem případě by to problém být mohl a to kvůli vzájemné provázanosti některých dat (multikolinearita nezávisle proměnných).

3.1.3 Modely 3

Z předchozího plyne, že pro vytvoření použitelných modelů bylo nezbytné, snížit počet regresorů. Některé dosud vytvořené modely se totiž neučily z dat, ale „naučily se“ data, což by při jejich reálném využití mohlo mít velmi nepříjemné následky. Kvůli tomu byly vytvořeny skupiny proměnných, které by, při současném zahrnutí dvou a více z nich, mohly způsobovat multikolinearitu. Zároveň je nutné mít na paměti, že tímto v žádném případě multikolinearita nebyla úplně vyloučena. V první fázi bylo vytvořeno 5 skupin nezávisle proměnných, jejich souhrn uvádí tabulka 3.4.

[2] podrobněji viz [16]

Tab. 3.4: Skupiny nezávisle proměnných

Skupina	Zahrnuté proměnné
Demografické	HZ, OBYV, 0-14, 15-64, 65+, BYT
OBADR	A, B, C, D, E viz tabulka 3.1
Odpady	PAP, PLST, SKLO, KOV, BIO, SKO, OBJ, KO
Interakce	0-14+65+
Mocniny	druhé mocniny nezávisle proměnných

Interakcemi se zde rozumí situace, kdy zvolené proměnné nemají samy o sobě velký vliv na závisle proměnnou, v případě jejich kombinace ale může být tento vliv významný. Tento případ si lze představit třeba na příkladu modelování pravděpodobnosti poruchy chladničky v závislosti na nastavené teplotě a tlaku. Pokud bude jedna z těchto hodnot nastavena na nějaké „rozumné“ hodnotě, lze tu druhou měnit prakticky libovolně a vliv na závisle proměnnou by neměl být velký. Pokud se ale obě tyto hodnoty budou blížit ke svým extrémům, pravděpodobnost poruchy se tím výrazně zvýší.

Pochopitelně nebyly zkoušeny všechny interakce. Byla použita např. interakce založená na myšlence, že s narozením vnoučat (0-14) se může zlepšit morálka třídění odpadu u osob důchodového věku (65+).

Při výběru regresorů pro tyto modely už byly použity výsledky další korelační analýzy (viz příloha A). Konkrétně tím způsobem, že z každé skupiny proměnných byly vybrány ty s největšími hodnotami korelačních koeficientů. Poté byl, v Minitabu, stepwise metodou sestaven model, který byl dále upravován tak, aby žádný model neobsahoval více než jeden regresor z každé skupiny. Velmi pozitivní zprávou je, že se pro jednotlivé druhy odpadu i KO podařilo sestavit modely objasňující velký podíl variability.

Kromě KOV a BIO se hodnota R^2 u všech modelů pohybovala nad hranicí 90%, což jsou velmi dobré výsledky. Nižší hodnota R^2 (69%) u KOV byla očekávaná již od počátku, na základě výsledků korelační analýzy. U BIO (83%) se pravděpodobně projevilo to, že jeho sběr ještě nemá v našich luzích a hájích tak bohatou tradici. Z dat je patrné, že se sběrem BIO se ve zkoumaném období teprve začínalo a tak je možné, že bude BIO v budoucnu lépe popsatelný. Všechny takto vytvořené modely tak můžeme považovat, z hlediska koeficientu determinace, za kvalitní.

Modely popisující odpady na osobu však dopadly hůře. Modely pro SKOs a objOs jsou ještě koeficientem determinace srovnatelné s předchozí kategorií. Problematická situace u plastOs byla nastíněna již při tvorbě předchozích modelů. K výraznému zlepšení nedošlo ani u sklOs. Pro papOs, kovOs, biOs a KOs se pak míra vysvětlené variability pohybovala v rozmezí 45 až 55%. KOs lze srovnat s výsledky [3],

Tab. 3.5: Přehled výsledných modelů celkových množství odpadu

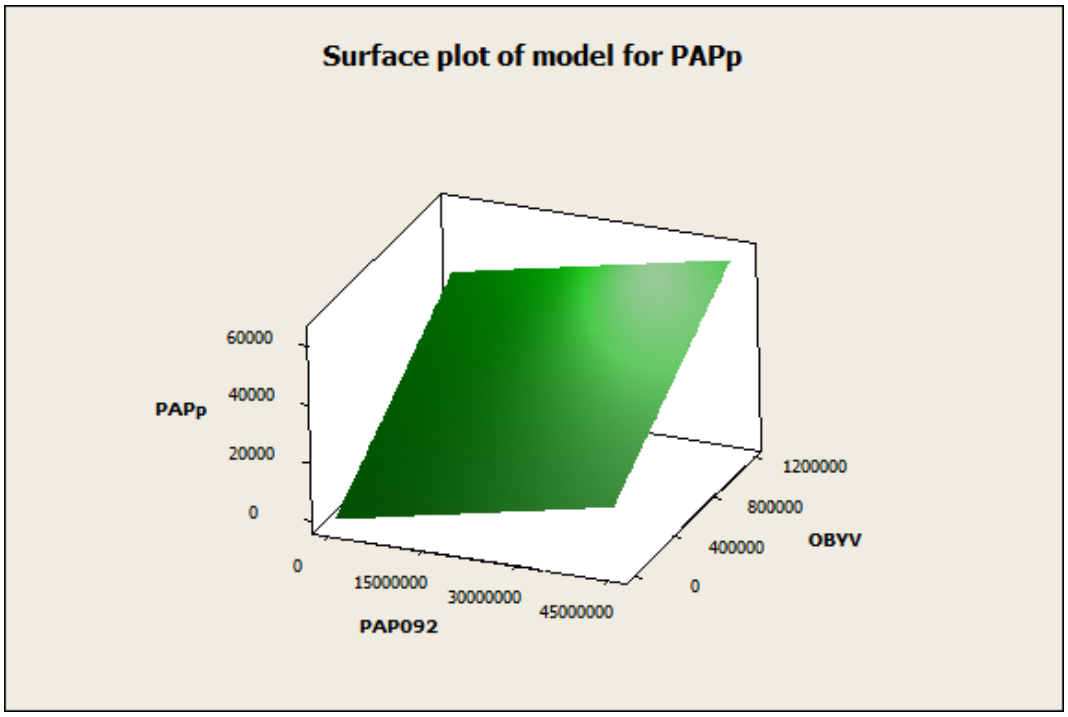
Y	OBYV09	BYT09	SKLO09	KOV09	BIO09	SKO09	OBJ09	PAP092	R^2
PAP	•							•	**96,7**
PLST	•		•						92,9
SKLO			•						95,3
KOV	•			•					69,6
BIO	•				•				83,7
SKO						•			99,0
OBJ		•					•		97,6
KO	•								98,4

xxx09 - údaj xxx z předchozího roku, PAP092 - druhá mocnina PAP z předchozího roku

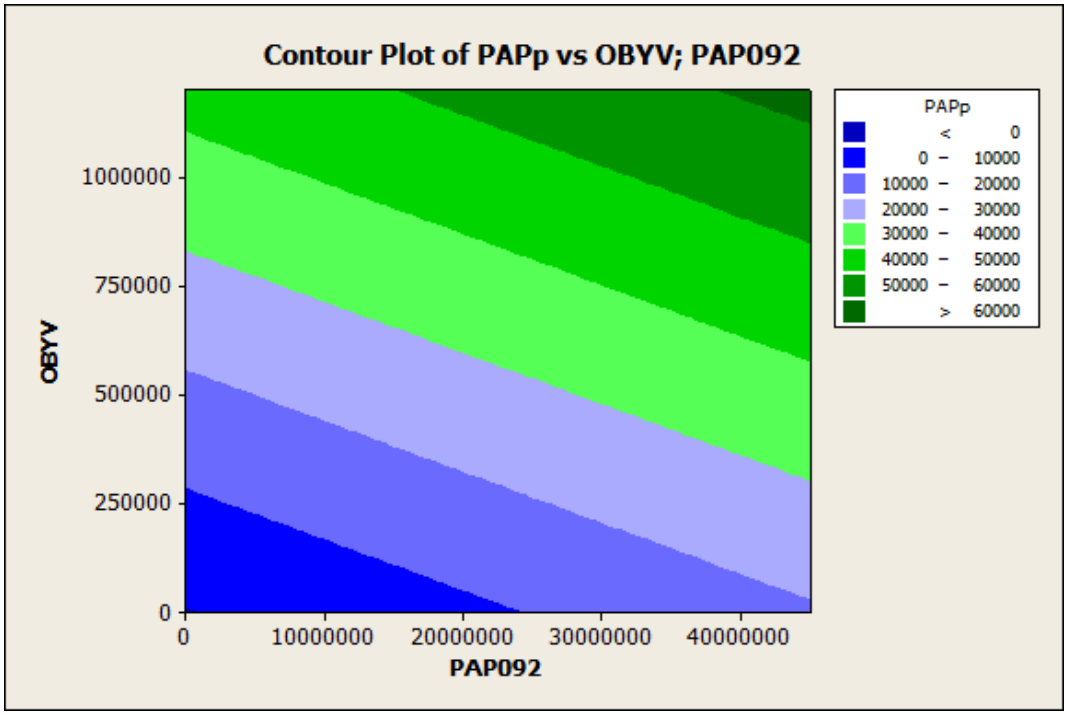
v našem případě je hodnota koeficientu determinace cca o 20% horší, to vzhledem k užšímu spektru nezávisle proměnných můžeme považovat za povzbuzující. Jako velký problém se však jeví, že většina z těchto modelů používá jako regresor pouze údaj o produkci daného odpadu na osobu v předchozím roce (polovina z nich přitom v první i druhé mocnině). Pokud se u takovýchto modelů nebude nijak zásadně měnit jejich časový vývoj, nemuselo by jít o fatální problém. Změní-li se však tyto trendy dramaticky během krátkého časového období, mohou být výsledky velmi nepřesné.

Celkově jsou tyto hodnoty srovnatelné nebo dokonce lepší, než u Modely 1 a Modely 2, hlavně ale s lepšími vlastnostmi, což byl hlavní důvod jejich tvorby.

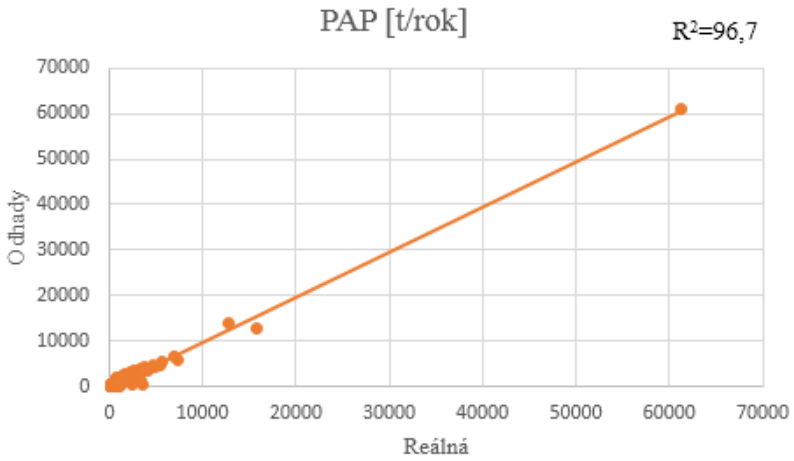
Na obrázcích 3.3 a 3.4 je vidět ukázka grafické interpretace regresního modelu, konkrétně modelu pro PAP. Obrázek 3.5 pak porovnává reálné hodnoty pro rok 2013 s hodnotami, které poskytl model 3 pro PAP (s daty z roku 2012).



Obr. 3.3: Grafické znázornění regresního modelu pro PAP - Surface plot



Obr. 3.4: Grafické znázornění regresního modelu pro PAP - Contour plot



Obr. 3.5: Srovnání reálných a predikovaných hodnot PAP pro rok 2013

3.2 Analýza modelu

Dosud byly vytvořené modely kvalitativně popisovány „pouze“ z hlediska koeficientu determinace. Při tvorbě modelů je také důležité zkoumat vliv *odlehlých bodů* (outliers) a tzv. *vlivných bodů* (leverage points). Za odlehlý je bod považován, pokud je jeho střední hodnota jiná, než udává model. Vlivné body se pak vyznačují tím, že se u nich hodnoty nezávisle proměnných značně liší od ostatních pozorování [16]. Analyzován bude model pro PAP, u ostatních modelů by tento proces probíhal podobným stylem.

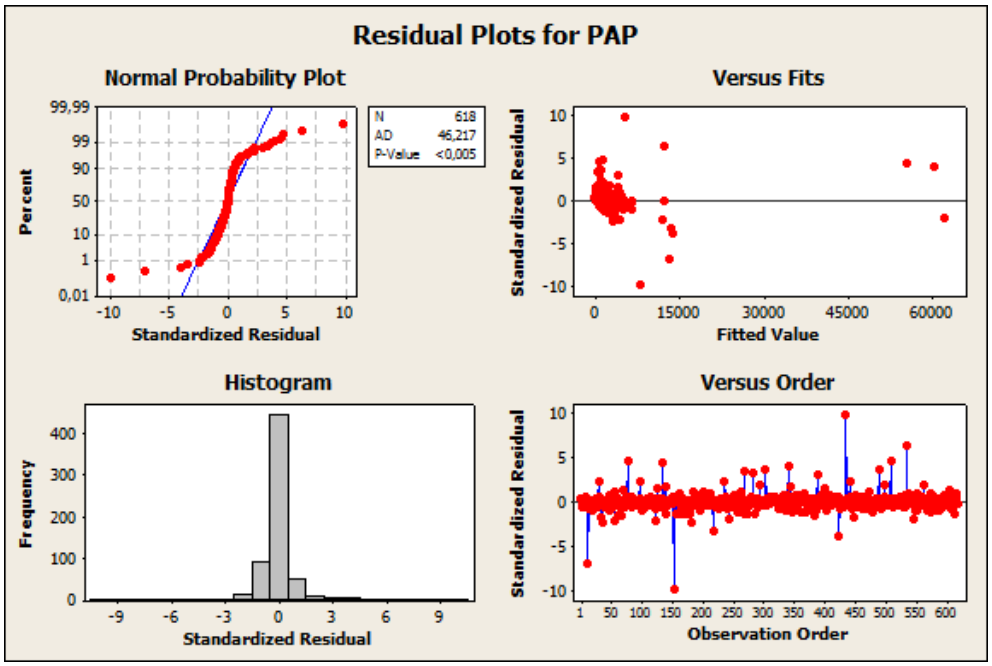
Jako základní model byl použit výsledný model z předchozí části, jehož tvar je uveden na obrázku 3.6 a grafy standardizovaných reziduí na obrázku 3.7. Při pohledu na tyto grafy vidíme „esovitý“ průběh reziduí znamenající (v našem případě) větší špičatost, než u normálního rozdělení. Tuto interpretaci potvrzuje i histogram. Postupným odstraňováním pozorování dokážeme „hrubou silou“ dosáhnout toho, že rezidua mají normální rozdělení. Uvedeným způsobem také samozřejmě zvyšujeme koeficient determinace, ale v tomto případě nejde ani tak o zlepšování modelu, jako spíše přizpůsobování dat modelu. Takovýto model pak nemá velký informační přínos.

```
The regression equation is
PAP = - 533 + 0,0367 OBYV + 0,000432 PAP092


Predictor         Coef      SE Coef        T      P
Constant      -532,53        43,57   -12,22  0,000
OBYV        0,0366791    0,0006513    56,32  0,000
PAP092     0,00043191   0,00002555    16,91  0,000


S = 816,196   R-Sq = 96,7%   R-Sq(adj) = 96,7%
```

Obr. 3.6: Rovnice modelu PAP



Obr. 3.7: Grafy reziduí modelu PAP

Unusual Observations

Obs	OBYV	PAP	Fit	SE Fit	Residual	St Resid
12	371399	7738,2	13226,9	209,1	-5488,8	-6,96RX
30	40196	2793,8	962,3	33,1	1831,5	2,25R
37	97595	1115,1	3051,1	48,0	-1936,0	-2,38R
55	82934	2369,1	4189,1	83,7	-1820,1	-2,24R
77	29305	4377,7	590,0	34,9	3787,7	4,64R
97	42176	2924,3	1046,8	33,0	1877,4	2,30R
123	335425	10201,2	11964,2	183,3	-1763,0	-2,22RX
135	1249026	58602,5	55532,5	421,3	3070,0	4,39RX
153	38476	1104,3	8012,9	424,6	-6908,6	-9,91RX
182	121458	2056,9	3942,2	59,5	-1885,4	-2,32R
218	371371	10711,8	13347,6	203,2	-2635,8	-3,33RX
236	40286	2831,7	978,8	33,1	1852,9	2,27R
269	22552	3049,5	295,1	35,7	2754,4	3,38R
283	29948	3292,4	648,7	35,4	2643,6	3,24R
303	42637	4010,8	1068,3	33,0	2942,5	3,61R
329	333579	12063,2	12152,3	169,8	-89,1	-0,11 X
341	1257158	63117,9	60411,9	468,1	2706,1	4,05RX
388	121699	6386,3	3949,6	59,7	2436,8	2,99R
424	378965	10725,0	13863,1	196,7	-3138,1	-3,96RX
435	154786	13227,3	5261,9	74,1	7965,4	9,80R
442	40212	2848,6	977,0	33,1	1871,5	2,29R
489	30429	3610,8	630,4	34,7	2980,4	3,65R
509	43323	4946,5	1126,0	33,1	3820,5	4,68R
535	329961	17251,9	12198,7	159,3	5053,2	6,31RX
547	1241664	61007,4	62217,4	546,4	-1210,0	-2,00 X

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.

Obr. 3.8: Odlehlá a vlivná pozorování modelu PAP

Pozorování s velkými standardizovanými rezidui pro nás mohou být také přínosná, protože poukazují na případy, kdy se závisle proměnná nechovala ve shodě s modelem, tedy ve shodě s chováním populace. Velké reziduum tak může upozornit na případy, kde je množství shromážděného odpadu odlišné od očekávaného chování. Jednotlivé ORP si pak mohou např. vyměnit informace o tom, jak lépe motivovat občany k třídění daného druhu odpadu. U rozebíraného modelu pro PAP má největší záporné standardizované reziduum Slaný (v roce 2009) a největší kladné pak České Budějovice (2012), interpretaci důvodů však ponecháme odborníkům.

Nenormalita reziduí je samozřejmě nepříjemná v tom, že znamená nesplnění předpokladů lineárního regresního modelu. Tedy testy významnosti jednotlivých koeficientů nemusejí být úplně korektní. Pokud se však podíváme na lineární regresi z pohledu numerického, jako na metodu nejmenších čtverců, mají modely s vysokými koeficienty determinace jistě vypovídající hodnotu. Pro výpočet koeficientu determinace totiž není požadováno normální rozdělení.

4 ZÁVĚR

Obsahem kapitoly prvé, práce kterou laskavý čtenář právě drží ve svých rukou, je shrnutí dosud dosažených a publikovaných výsledků v oblasti přístupu predikování KO. Jsou zde popsána kritéria, podle kterých se používané modely člení a následně jsou uvedeny příklady zkoumaných studií.

V kapitole druhé jsou pak uvedeny nejdůležitější teoretické poznatky o regresní analýze a také krátká exkurze do tajů korelační analýzy.

Dříve uvedené poznatky pak byly využity při tvorbě regresních modelů pro jednotlivé druhy KO a také jejich přepočtů na osobu. Vytvořené modely velmi dobře vyhovují požadavkům na rozličnost zahrnutých regresorů při zachování vysokých koeficientů determinace. Konkrétně u celkových množství se 6 z 8 modelů může honosit koeficientem determinace přesahujícím hodnotu 90 %. Nejhorší model pak vysvětluje kolem 70 % variability, což je stále dobrý výsledek. U odpadů na osobu se pak většina modelů pohybuje nad hranicí 45 %, což už není tak dobré, ale jedná se o očekávaný výsledek.

Na závěr byla zkoumána rezidua vybraného modelu. Přitom došlo k nepříjemnému zjištění v podobě porušení předpokladu na normalitu rozdělení reziduí, některé testy tak nemohou být korektně provedeny. I přesto však budou vytvořené modely použitelné pro nástroj JUSTÝNA, protože lineární regresi lze interpretovat také z numerického pohledu, jako metodu nejmenších čtverců. Tudíž podmínka normality reziduí, která nás limituje v oblasti matematické statistiky není v oblasti numerických metod vyžadována.

Jak si jistě pozorný čtenář povšiml, výsledky této práce jsou spíše jakýmsi prvním krůčkem ve vývoji, než čímkoliv jiným. A to jak z pohledu hloubky zkoumání dané problematiky, tak i z pohledů požadavků kolegů z ÚPI. Bylo by tedy vhodné na dosažené výsledky navázat a pokusit se tak celou snahu posunout o další krok vpřed.

Hlavními přínosy pro autora samotného byly hlubší seznámení se statistickým softwarem, práce s reálnými daty a požadavky, jejich korigování a v neposlední řadě ilustrace pojmů uváděných ve výuce na reálných příkladech.

LITERATURA

[1] BEIGL, P., S. LEBERSORGER a S. SALHOFER. *Modelling municipal solid waste generation: A review*. Waste Management. 2008, 28, 200-214.

[2] INTHARATHIRAT, Rotchana, P. ABDUL SALAM, S. KUMAR a Akarapong UNTONG. *Forecasting of municipal solid waste quantity in a developing country using multivariate grey models*. Waste Management. 2015, 39, 3-14.

[3] BEIGL, P. a S. LEBERSORGER. *Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel*. Waste Management. 2011, 31, 1907-1915.

[4] HOCKETT, Daniel, Douglas LOBER a Keith PILGRIM. *Determinants of Per Capita Municipal Solid Waste Generation in the Southeastern United States*. Journal of Environmental Management. 1995, 45, 205-217.

[5] Organisation for Economic Co-operation and Development (OECD), 2004. *Towards Waste Prevention Performance Indicators.*. ENV/EPOC/WGWPR/SE(2004)1/FINAL. Environment directorate, Paris, France.

[6] SKOVGAARD, M., Moll, S., Andersen, F.M., Larsen, H. *Outlook for waste and material flows: baseline and alternative scenarios.*. Working Paper 1. European Topic Centre on Resource and Waste Management, Copenhagen, Denmark.

[7] KARAVEZYRIS, Vassilios, Klaus-Peter TIMPE a Ruth MARZI. *Application of system dynamics and fuzzy logic to forecasting of municipal solid waste.*. Mathematics and Computers in Simulation. 2002, 60, 149-158.

[8] LILAI, Xu, Gao PEIQING, Cui SHENGHUI a Liu CHUN. *A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China.*. Waste Management. 2013, 33, 1324–1331.

[9] IBÁNEZ, M.V., M. PRADES a A. SIMÓ. *Modelling municipal waste separation rates using generalized linear models and beta regression*. Resources, Conservation and Recycling. 2011, 55, 1129-1138.

[10] HUI-ZHEN, Fu, Li ZHEN-SHAN a Wang RONG-HUA. *Estimating municipal solid waste generation by different activities and various resident groups in five provinces of China*. Waste Management. 2015, 41, 3-11.

[11] KESER, Saniye, Sebnem DUZGUN a Aysegul AKSOY. *Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey*. Waste Management. 2012, 32, 359-371.

[12] AZADI, S., Karimi-Jashni, A. *Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran*. Waste Management. 2015. Dostupné z URL: <`http://dx.doi.org/10.1016/j.wasman.2015.09.034`>.

[13] Hierarchie nakládání s odpady. In: Arnika [online]. [cit. 2016-04-10]. Dostupné z: http://arnika.org/hierarchie-nakladani-s-odpady

[14] KLIMEŠ, Lumír *Slovník cizích slov*. Vyd. 3. Praha: Státní pedagogické nakladatelství, 1986.

[15] HRABEC, P. *Zavedení a aplikace obecného regresního modelu*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2015. 38 s. Vedoucí Ing. Josef Bednář, Ph.D.

[16] ZVÁRA, Karel. *Regrese*. Vyd. 1. Praha: Matfyzpress, 2008. ISBN 978-80-7378-041-8.

[17] ANDĚL, Jiří. *Základy matematické statistiky*. Vyd. 3. Praha: Matfyzpress, 2011. ISBN 978-80-7378-162-0.

[18] *Shlukování podobných v softwaru STATISTICA*. [online]. [cit. 2016-04-10]. Dostupné z: http://www.statsoft.cz/file1/PDF/newsletter/2014_10_08_StatSoft_Shlukovani_podobnych_v_softwaru_statistica.pdf

[19] *Novinky STATISTICA 12: Jednodušší tvorba shluků*. [online]. [cit. 2016-04-10]. Dostupné z: http://www.statsoft.cz/file1/PDF/newsletter/2013_11_05_StatSoft_Novinky_ve_shlukove_analyze.pdf

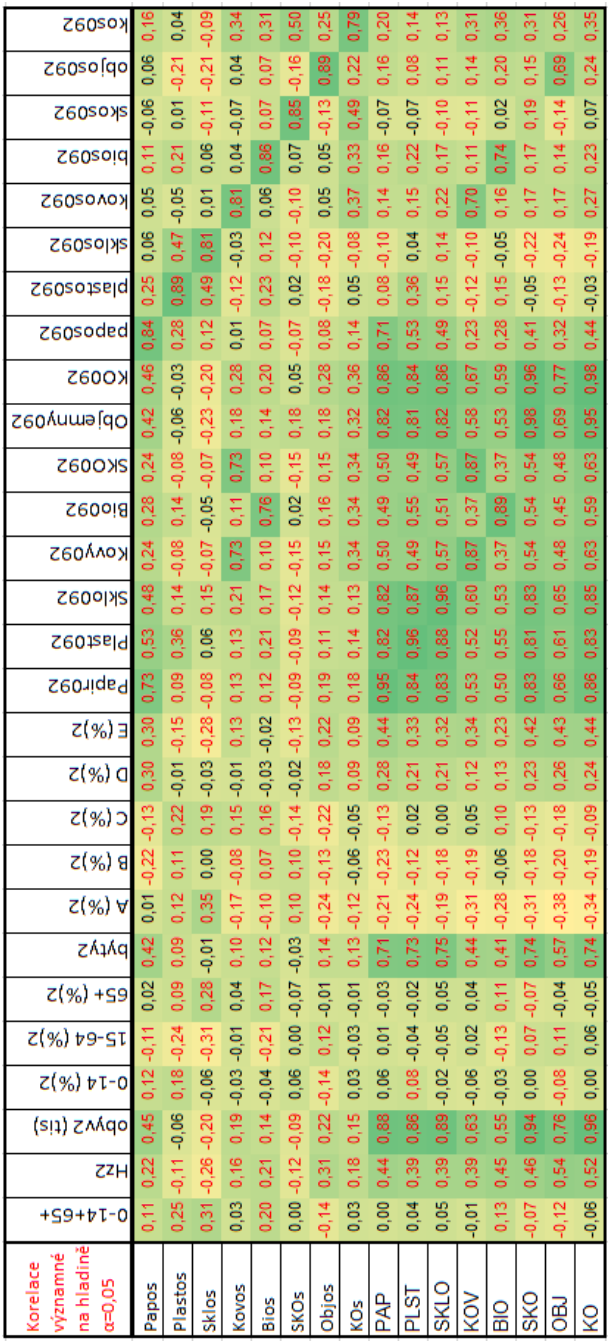
[20] Help and How-To *Minitab Express Support*. [online]. [cit. 2016-04-10]. Dostupné z: $http://support.minitab.com/en-us/minitab-express/1/$

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

$\mathbf{1}$	vektor jedniček
$\mathbf{I}$	jednotková matice
X'	transponovaná matice
$\mathsf{E}\,Y$	střední hodnota náhodné veličiny Y
$\| \mathbf{u} \|$	norma vektoru u
cov (X, Y)	kovariance X a Y
var Y	rozptyl náhodné veličiny Y
ČSÚ	Český statistický úřad
GM	Grey model
GMC	Grey model s konvolučním integrálem
ISOH	Informační systém odpadového hospodářství
KO	komunální odpad
OH	odpadové hospodářství
ORP	obec s rozšířenou působností

SEZNAM PŘÍLOH

A KORELAČNÍ ANALÝZA



Korelace významné na hladině α=0,05	0-14+65+	Hz2	obyv2 (tis)	0-14 (%)2	15-64 (%)2	65+ (%)2	byty2	A (%)2	B (%)2	C (%)2	D (%)2	E (%)2	Papir092	Plast092	Sklo092	kovy092	Bio092	SKO092	Objemny092	KO092	papos092	plast092	sklo092	kovo092	bio092	sko092	obj092	ko092
Papos	0,11	0,22	0,45	0,12	-0,11	0,02	0,42	0,01	-0,22	-0,13	0,30	0,30	0,73	0,53	0,48	0,24	0,28	0,24	0,42	0,46	0,84	0,25	0,06	0,05	0,11	-0,06	0,06	0,16
Plastos	0,25	-0,11	-0,06	0,18	-0,24	0,09	0,09	0,12	0,11	0,22	-0,01	-0,15	0,09	0,36	0,14	-0,08	0,14	-0,08	-0,06	-0,03	0,28	0,89	0,47	-0,05	0,21	0,01	-0,21	0,04
Sklos	0,31	-0,26	-0,20	-0,06	-0,31	0,28	-0,01	0,35	0,00	0,19	-0,03	-0,28	-0,08	0,06	0,15	-0,07	-0,05	-0,07	-0,23	-0,20	0,12	0,49	0,81	0,01	0,06	-0,11	-0,21	-0,09
Kovos	0,03	0,16	0,19	-0,03	-0,01	0,04	0,10	-0,17	-0,08	0,15	-0,01	0,13	0,13	0,13	0,21	0,73	0,11	0,73	0,18	0,28	0,01	-0,12	-0,03	0,81	0,04	-0,07	0,04	0,34
Bios	0,20	0,21	0,14	-0,04	-0,21	0,17	0,12	-0,10	0,07	0,16	-0,03	-0,02	0,12	0,21	0,17	0,10	0,76	0,10	0,14	0,20	0,07	0,23	0,12	0,06	0,86	0,07	0,07	0,31
SKOs	0,00	-0,12	-0,09	0,06	0,00	-0,07	-0,03	0,10	0,10	-0,14	-0,02	-0,13	-0,09	-0,09	-0,12	-0,15	0,02	-0,15	0,18	0,05	-0,07	0,02	-0,10	-0,10	0,07	0,85	-0,16	0,50
Objos	-0,14	0,31	0,22	-0,14	0,12	-0,01	0,14	-0,24	-0,13	-0,22	0,18	0,22	0,19	0,11	0,14	0,15	0,16	0,15	0,18	0,28	0,08	-0,18	-0,20	0,05	0,05	-0,13	0,89	0,25
KOs	0,03	0,18	0,15	0,03	-0,03	-0,01	0,13	-0,12	-0,06	-0,05	0,09	0,09	0,18	0,14	0,13	0,34	0,34	0,34	0,32	0,36	0,14	0,05	-0,08	0,37	0,33	0,49	0,22	0,79
PAP	0,00	0,44	0,88	0,06	0,01	-0,03	0,71	-0,21	-0,23	-0,13	0,28	0,44	0,95	0,82	0,82	0,50	0,49	0,50	0,82	0,86	0,71	0,08	-0,10	0,14	0,16	-0,07	0,16	0,20
PLST	0,04	0,39	0,86	0,08	-0,04	-0,02	0,73	-0,24	-0,12	0,02	0,21	0,33	0,84	0,96	0,87	0,49	0,55	0,49	0,81	0,84	0,53	0,36	0,04	0,15	0,22	-0,07	0,08	0,14
SKLO	0,05	0,39	0,89	-0,02	-0,05	0,05	0,75	-0,19	-0,18	0,00	0,21	0,32	0,83	0,88	0,96	0,57	0,51	0,57	0,82	0,86	0,49	0,15	0,14	0,22	0,17	-0,10	0,11	0,13
KOV	-0,01	0,39	0,63	-0,06	0,02	0,04	0,44	-0,31	-0,19	0,05	0,12	0,34	0,53	0,52	0,60	0,87	0,37	0,87	0,58	0,67	0,23	-0,12	-0,10	0,70	0,11	-0,11	0,14	0,31
BIO	0,13	0,45	0,55	-0,03	-0,13	0,11	0,41	-0,28	-0,06	0,10	0,13	0,23	0,50	0,55	0,53	0,37	0,89	0,37	0,53	0,59	0,28	0,15	-0,05	0,16	0,74	0,02	0,20	0,36
SKO	-0,07	0,46	0,94	0,00	0,07	-0,07	0,74	-0,31	-0,18	-0,13	0,23	0,42	0,83	0,81	0,83	0,54	0,54	0,54	0,98	0,96	0,41	-0,05	-0,22	0,17	0,17	0,19	0,15	0,31
OBJ	-0,12	0,54	0,76	-0,08	0,11	-0,04	0,57	-0,38	-0,20	-0,18	0,26	0,43	0,66	0,61	0,65	0,48	0,45	0,48	0,69	0,77	0,32	-0,13	-0,24	0,17	0,14	-0,14	0,69	0,26
KO	-0,06	0,52	0,96	0,00	0,06	-0,05	0,74	-0,34	-0,19	-0,09	0,24	0,44	0,86	0,83	0,85	0,63	0,59	0,63	0,95	0,98	0,44	-0,03	-0,19	0,27	0,23	0,07	0,24	0,35

B REGRESNÍ ANALÝZA

B.1 Přehled modelů

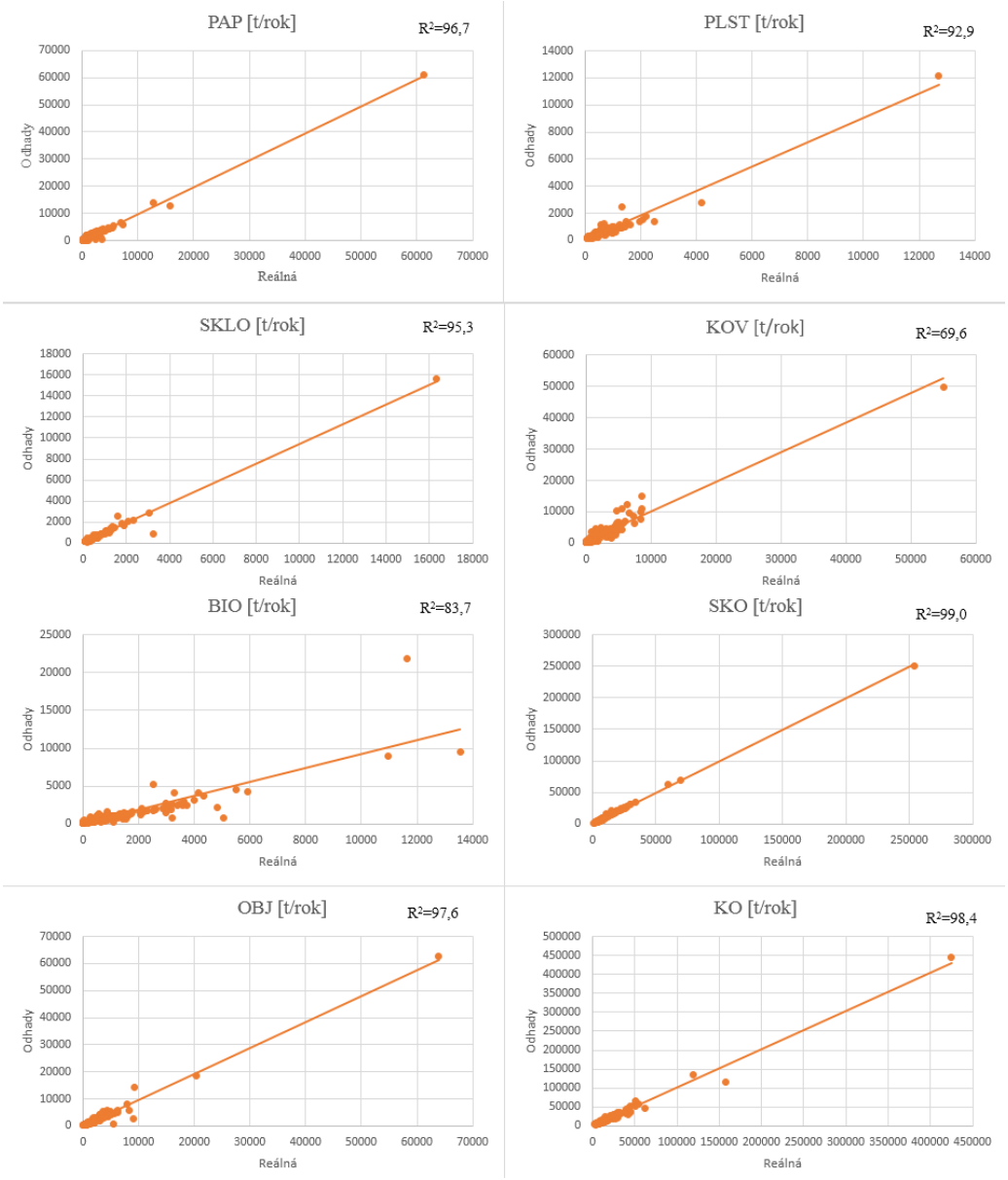
Modely 1	Konstanta	HZ	OBYV	0-14	15-64	65+	BYTY	A	B	C	D	E	R_sq	R_sq_a
PAP	-281,4	-1,43	0,048									-14,7	86,36	86,32
PLST	93,35	-0,92	0,01118										28,7	28,56
SKLO	330,7	-0,555	0,0126				0,217			-4,7	4,7	-11,4	-5,53	95,12	95,08
KOV	19321	2,05	0,0225		-237			-4,52	-55	-35				22	21,54
BIO	-6491		0,0147	297		142		-0,67	-8,4			-23	7,1	54,12	53,79
SKO	26579	-2,14	0,2043		-259	-414					-38	-57	-22,8	98,09	98,07
OBJ	-24503	3,85	0,0496		300	221	1,19				-37,2		-10,1	94,02	93,98
KO	-7239		0,365	488			-3,84							97,64	97,63
Papos	297,9		0,00003		-3,3	-2,5					-0,46			1,68	1,29
Plastos	*														
Sklos	36,02	-0,0036	0,00001		-0,42			0,127			0,221			11,94	11,51
Kovos	93,47	0,0188				3,31	-0,0141	-1,56	-1,46			-1,82	-0,92	10,41	9,8
Bios	-87,39	0,0135	-0,00007	3,31		4,28	0,0087	-0,318	-0,177			-0,58		9,73	9,02
SKOs	568,3				-4,2	-7,1		1,09	0,75			1,72		5,54	5,07
Objos	32,06	0,0372	-0,00007	-1,69					0,44		-0,69	0,66	0,515	14,62	14,03
KOs	110,6	0,073	-0,00013	10,3		5,1								1,69	1,31

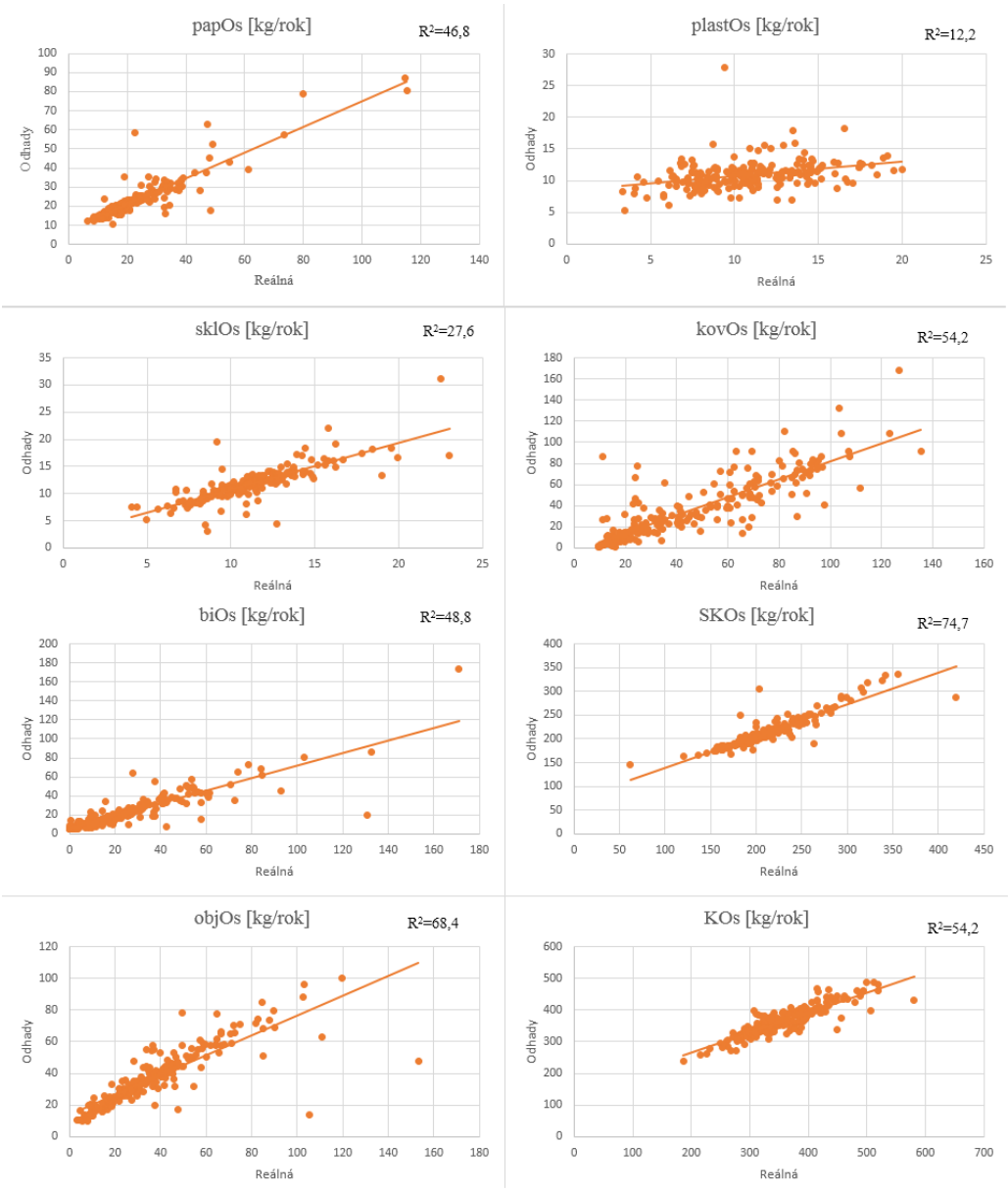
Modely 2	Konstanta	HZ	OBYV	0-14	BYTY	A	C	D	E	PAP09	PLST09	SKLO09	KOV09	BIO09	SKO09	OBJ09	KO09	R sq	R sq adj
PAP	-573,7		0,02406		1,32	7,9				0,487	-0,421							97,54	97,52
PLST	-51,602	-0,596	0,00781		0,385		5,6	5,8		0,0415	-0,0233	0,155	0,0089	0,0861	-0,0071	-0,0256		96,71	96,63
SKLO	21,83	-0,345	0,0027		0,083							0,7651		0,0284				97,38	97,36
KOV	-459,5		0,0627		-1,65					-0,454	0,415	-1,38	0,806			-0,257		74,7	74,39
BIO	17,69		0,00683		0,59							-0,3	0,0238	0,734				84,39	84,25
SKO	211,02		0,0813		0,88								-0,068	0,227	0,5967	-0,091		99,38	99,37
OBJ	2692	1,29		-213	2,97				10,8			-0,34		-0,175	0,0372	0,672		97,87	97,84
KO	-683,2		0,191		4,44					-0,44	-0,68	-2,05	0,21	0,34		-0,244	0,594	99,29	99,28

Modely 2	Konstanta	HZ	0-14	15-64	BYTY	A	B	C	Papos09	Plastos09	Sklos09	Kovos09	Bios09	SKOs09	Objos09	R sq	R sq adj
Papos	17,91				0,0035		-0,08		0,529	-0,464						40,26	39,87
Plastos	-3,653	-0,003	0,64		0,00191	0,06		0,163			0,107		0,025			11,3	10,28
Sklos	4,923					0,114		0,165			0,294					16,13	15,72
Kovos	-194,11			2,9				0,46				0,704				47,09	46,83
Bios	95,049			-1,29									0,828			49,38	49,21
SKOs	54,05													0,762	-0,075	74,58	74,5
Objos	15,178					-0,157		-0,39							0,749	69,13	68,98
KOs	128,6											0,712	0,82	0,748	0,723	55,44	55,14

Modely 3	Konstanta	OBYV09	BYTY09	C (%)09	PAP09	PLST09	SKLO09	KOV09	BIO09	SKO09	OBJ09	KO09	PAP092	sklos092	kovos092	skos092	kos092	Rsq	Rsq_adj
PAP	-532,5	0,03668											0,00043					96,67	96,66
PLST	17,3974	0,00575					0,31											92,89	92,87
SKLO	29,987						0,9699											95,26	95,25
KOV	250	0,00485						0,946										69,63	69,53
BIO	32,837	0,0059							0,736									83,69	83,63
SKO	44,75									0,9966								98,99	98,99
OBJ	52,87		2,53								0,768							97,56	97,55
KO	148,2	0,3569																98,4	98,39
	Konstanta	OBYV09	BYTY09	C (%)09	papOs09	plastOs09	sklOs09	kovOs09	biOs09	SKOs09	objOs09	KOs09	papOs092	sklOs092	kovOs092	SKOs092	KOs092	Rsq	Rsq_adj
Papos	7,439	0,00001			0,69								-0,00064					46,81	46,55
Plastos	3,634			0,091			0,508							0,00341				12,21	11,79
Sklos	0,596						1,08							-0,0068				27,63	27,4
Kovos	8,31							1,163							-0,00257			54,17	54,02
Bios	5,061								0,839									48,76	48,67
SKOs	136,56															0,00159		74,69	74,65
Objos	7,999										0,774							68,38	68,33
KOs	67,59											0,952					-0,00033	54,22	54,07

B.2 Srovnání odhadů s reálnými daty pro rok 2013




papOs [kg/rok]
R²=46,8
plastOs [kg/rok]
R²=12,2
sklOs [kg/rok]
R²=27,6
kovOs [kg/rok]
R²=54,2
biOs [kg/rok]
R²=48,8
SKOs [kg/rok]
R²=74,7
objOs [kg/rok]
R²=68,4
KOs [kg/rok]
R²=54,2
Odhady
Reálná