



Ekonomická
fakulta
Faculty
of Economics

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Jihočeská univerzita v Českých Budějovicích

Fakulta ekonomická

Katedra aplikované matematiky a informatiky

Bakalářská práce

Statistické zpracování nezaměstnanosti

Vypracoval: Patrik Lojda

Vedoucí práce: doc. RNDr. Tomáš Mrkvička Ph.D.

České Budějovice 2017

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Patrik LOJDA**

Osobní číslo: **E14139**

Studijní program: **B6208 Ekonomika a management**

Studijní obor: **Účetnictví a finanční řízení podniku**

Název tématu: **Statistické zpracování nezaměstnanosti**

Zadávací katedra: **Katedra aplikované matematiky a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Cílem je vytvoření modelu nezaměstnanosti v závislosti na různých charakteristikách okresu.

Metodický postup:

1. Studium literatury - 1. semestr.
2. Sběr dat - 2. semestr.
3. Vyhodnocení informací - 2. semestr.
4. Zpracování výsledků - 3. semestr.

Rozsah grafických prací: dle potřeby
Rozsah pracovní zprávy: 40 - 50 stran
Forma zpracování bakalářské práce: tištěná
Seznam odborné literatury:

1. Anděl, J. (1998). *Statistické metody*. Praha: Matfyzpress.
2. Howell, D. R. (2004). *Fighting Unemployment: The Limits of Free Market Orthodoxy*. Oxford: Oxford University Press.
3. https://www.czso.cz/csu/czso/zamestnanost_nezamestnanost_prace

Vedoucí bakalářské práce: doc. RNDr. Tomáš Mrkvička, Ph.D.
Katedra aplikované matematiky a informatiky

Datum zadání bakalářské práce: 15. ledna 2016
Termín odevzdání bakalářské práce: 15. dubna 2017

02 
doc. Ing. Ladislav Rolínek, Ph.D.
děkan

JIHOČESKÁ UNIVERZITA
V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
Studentská 13 (1)
370 05 České Budějovice


prof. RNDr. Pavel Tlustý, CSc.
vedoucí katedry

V Českých Budějovicích dne 8. února 2016

Prohlášení

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47 zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to - v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích 13. dubna 2017

.....

Patrik Lojda

Poděkování

Mé poděkování právem patří doc. RNDr. Tomášovi Mrkvičkovi, Ph.D. Jeho ochota, cenné rady, laskavost a podpora pro mne tvořily základní kameny při vypracování této práce a bez nich byste tohle nečetli.

Obsah

1. Úvod.....	8
2. Teorie.....	9
2.1. Trh práce, nabídka a poptávka po práci	9
2.1.1. Trh práce.....	9
2.1.2. Nabídka práce.....	9
2.1.3. Substituční efekt zvýšení mzdy.....	10
2.1.4. Důchodový efekt zvýšení mzdy.....	11
2.1.5. Vliv důchodového a substitučního efektu na nabídku práce	11
2.1.6. Poptávka po práci.....	12
2.2. Nezaměstnanost.....	12
2.2.1. Míra nezaměstnanosti	12
2.2.2. Frikční nezaměstnanost	13
2.2.3. Strukturální nezaměstnanost.....	13
2.2.4. Cyklická nezaměstnanost	14
2.2.5. Dobrovolná a nedobrovolná nezaměstnanost.....	14
2.2.6. Příčiny nedobrovolné nezaměstnanosti.....	15
2.2.6.1. Odbory.....	16
2.2.6.2. Stát.....	16
2.2.7. Substituční vztah mezi nezaměstnaností a inflací	16
3. Metodika	18
3.1. Korelační a regresní analýza	18
3.1.1. Korelační analýza	18
3.1.2. Regresní analýza.....	20
3.1.3. Typy regresních funkcí	21
3.2. Test normality.....	21
3.2.1. Porovnání mediánu a aritmetického průměru	22
3.2.2. Histogram	22
3.2.3. Q-Q graf	23
3.3. N-P graf reziduí.....	25
4. Proměnné do modelu	26
4.1. Dálnice, silnice, železnice, vzdálenosti mezi okresními městy, krajskými městy a hlavním městem	26

4.2.	Státní hranice	26
4.3.	Vybavenost domácností informačními a komunikačními technologiemi	27
4.4.	Cizinci	27
4.5.	Průměrný věk	28
4.6.	Průměrná mzda	28
5.	Zkoumání závislostí mezi mírou nezaměstnanosti a proměnnými	29
5.1.	Dálnice	29
5.2.	Silnice 1. a 2. tříd	30
5.3.	Vzdálenost okresního města od krajského města a hlavního města Prahy	31
5.4.	Státní hranice	32
5.5.	Cizinci	33
5.6.	Informační technologie	33
5.7.	Průměrný věk	34
5.8.	Průměrná hrubá mzda	35
5.9.	Železnice	36
5.10.	Korelační koeficienty	37
6.	Vytváření modelu	38
6.1.	Model z dat roku 2015	40
6.2.	Model z dat roku 2005	46
6.3.	Model z dat roku 2010	48
6.4.	Model z dat roku 2015 s prostorovou korelací	49
6.5.	Použití modelů pro predikci hodnot roku 2016	51
7.	Závěr	54
8.	Summary	55
9.	Seznam literatury a použitých zdrojů	56
	Seznam grafů, tabulek a kartogramu	57

1. Úvod

Ekonomický růst, nezaměstnanost, stabilita cenové hladiny a obchodní bilance jsou čtyři základní pilíře, podle kterých se hodnotí hospodářská politika a výkonnost ekonomiky. Tato bakalářská práce se zabývá nezaměstnaností. Cílem je vytvoření modelu nezaměstnanosti v závislosti na různých charakteristikách okresu. Pro statistické výpočty a grafy je použit program Excel 2007 a Statistika 13.2.

V začátku se nachází popis trhu, nabídky a poptávky po práci. Na příkladech je vysvětleno, jak ovlivňuje substituční a důchodový efekt křivku nabídky práce. Pokračuje se nezaměstnaností, kde se vysvětluje daná problematika. Rozebírá se výpočet míry nezaměstnanosti, dělení podle příčiny a dobrovolnosti, příčiny nedobrovolné nezaměstnanosti a substituční vztah mezi nezaměstnaností a inflací. Hlavním pramenem z kterého je čerpáno je 4. vydání učebnice s názvem *Ekonomie* napsané Robertem Holmanem.

Následně je vysvětleno na příkladech kouzlo korelační a regresní analýzy, bez kterých by v této práci nebylo možno pokračovat. Ve stručných kapitolách je popsán korelační koeficient a typy regresních funkcí.

Náplní další kapitoly se stane výběr proměnných, z kterých se budou modely vytvářet. Volí se takové, které by mohly ovlivňovat míru nezaměstnanosti, byly kvantifikovatelné, různé v závislosti na okrese a hlavně spolehlivé. Většinu z nich je nalezena na webových stránkách Českého statistického úřadu. Korelační analýzou a grafy se ověří, zda zvolené proměnné se opravdu chovají, jak je očekáváno. Následně se korelační analýzou zkoumá, jestli mezi proměnnými nejsou data dvakrát, jen s jiným názvem.

Primární cíl práce je rozvinut. Vytváří se tři regresní modely z různých let a zkouší se v praxi. Odhaduje se míra nezaměstnanosti pro rok 2016 a porovnávají se výsledky. Sekundární cíl tvoří zkoumání, zda existuje závislost mezi mírami nezaměstnanosti okresů se společnou hranicí. Pokud ano, vytvoří se další model, který bude brát v úvahu sousedící okresy.

V závěru si řekneme, jestli lze nějaký z modelů použít v praxi. Popřípadě přidáme další poznatky, ke kterým v práci dospějeme.

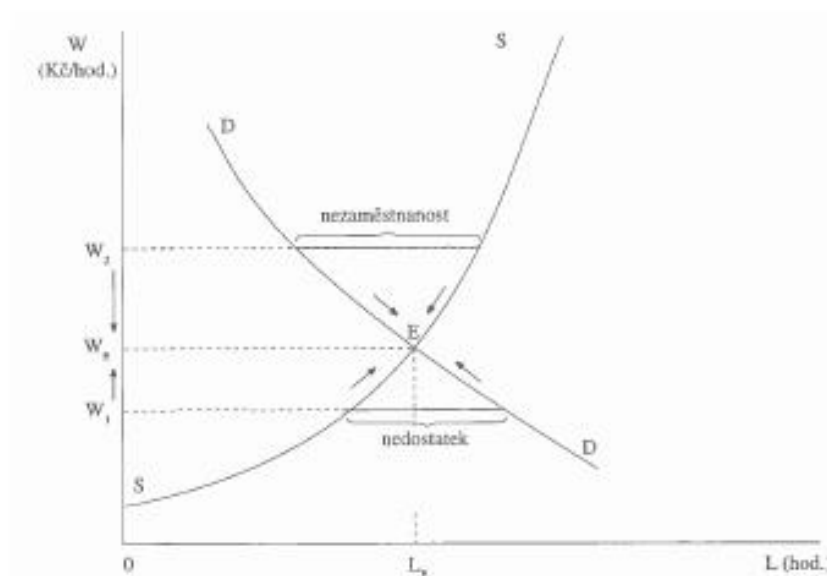
2. Teorie

2.1. Trh práce, nabídka a poptávka po práci

2.1.1. Trh práce

Trh jako každý jiný, kde se střetává nabídka s poptávkou, v tomto případě nabídka práce a poptávka po práci. Firmy hledají pracovní sílu a nabízejí za ni mzdu, naopak lidé tuto mzdu poptávají a nabízejí svůj volný čas. Zkráceně řečeno, že lidé prodávají svůj čas. Jako na každém trhu i zde existuje konkurence na straně poptávky a na straně nabídky (Holman, 2013).

Graf 1 - Trh práce



zdroj: Holman, 2005

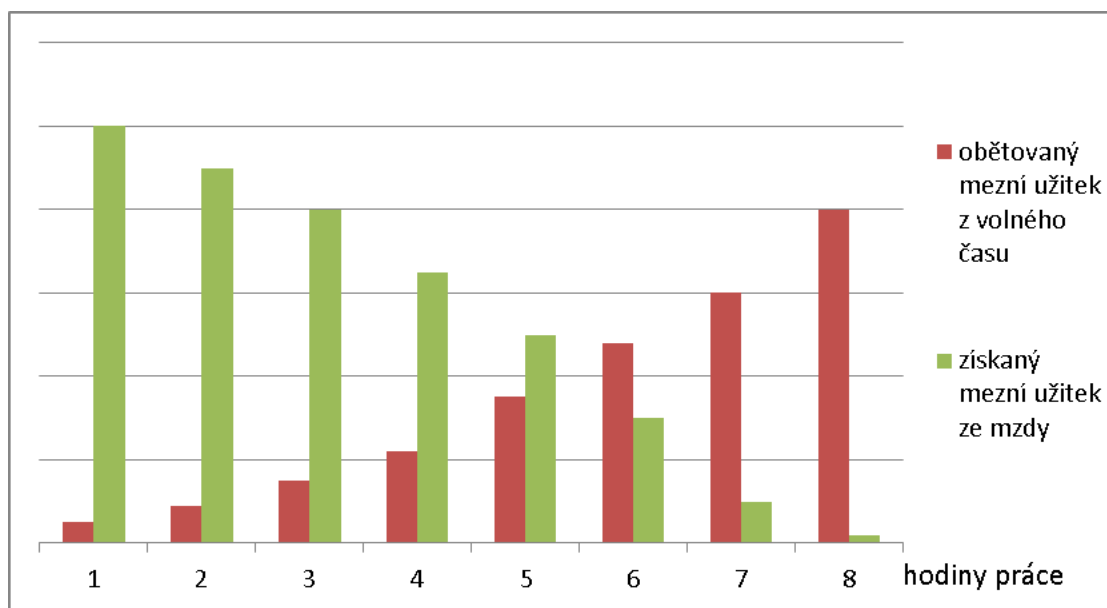
Na grafu 1 je vidět trh práce, kde se střetává nabídka a poptávka po práci. Vodorovnou osu tvoří množství práce v hodinách a na svislou je vynášena hodinová mzda. Trh se drží v rovnováze E , jako každý jiný trh a pokud je nějakým způsobem vychýlen, má tendenci se vrátit do rovnovážného stavu (Holman, 2013).

2.1.2. Nabídka práce

Představme si, že máme nezaměstnaného člověka, řekněme mu Adam, který má za celý den 24 hodin volného času. Pokud bychom Adamovi nabídli práci za nějakou mzdu, pravděpodobně by ji přijal, poněvadž mezní užitek z peněz bude větší, než mezní užitek,

který ztratí obětováním jedné hodiny svého volného času. Takto to půjde až do doby, dokud odpracování další hodiny nebude představovat obětování většího mezního užitku než získání mezního užitku, který dostane v podobě větší mzdy. Tohle je bod, ve kterém nám Adam řekne, že už další hodinu pro nás pracovat nebude a dohoda se ustálí na konkrétní době, kterou bude ochotný odpracovat (Holman, 2013).

Graf 2 - Mezní užítiky



zdroj: vlastní tvorba

V tomto případě se Adamova nabídka ustálí na pěti hodinách práce, protože šestá hodina práce už pro Adama znamená ztrátu celkového užitku (Holman, 2013).

2.1.3. Substituční efekt zvýšení mzdy

Adam nám řekl, že za tu mzdu, co jsme mu nabídli, nám nabízí určitý počet hodin. Ze mzdy si platí nájem, stravu, oblečení a touží si koupit automobil. Při stávající mzdě se rozhodl, že pro něj 2 hodiny volného času jsou více než automobil, který by si pořídil, kdyby pracoval o 2 hodiny denně navíc. Co by se stalo, kdybychom mu zvýšili mzdu? Adam by zaplatil všechno, co platil dříve a ještě by mu zbylo. Jelikož už by vlastnil nějaké finanční prostředky navíc, na vysněný automobil by mu stačilo pracovat už jen o půl hodiny více. V tomto případě se Adam rozhodne pro nový automobil a obětuje práci další čas. Řečeno našimi mezními užítiky, Adam by byl ochotný pracovat více hodin denně, protože by mu mzda přinášela větší mezní užitek než u mzdy předchozí. Ve výsledku by byl ochotný odpracovat větší počet hodin než předtím. Zkráceně řečeno,

pokud budeme zvyšovat mzdu, lidé budou ochotni pracovat delší dobu. Jenže proti substitučnímu efektu nám zde působí i důchodový efekt (Holman, 2013).

2.1.4. Důchodový efekt zvýšení mzdy

Řekli jsme si, že při zvyšování mzdy bude Adam nabízet více hodin práce. Důchodový efekt nám pojednává o tom, že pokud budeme zvyšovat mzdu, Adam nám bude nabízet méně hodin. Teď vás určitě zarazilo, co jste se to vlastně dočetli. Proč by někdo pracoval méně, pokud bychom mu nabídli větší mzdu? Představte si, že Adam v práci vydělává tolik, že všechno pro něj důležité si může koupit a další peníze už mu nepřinášejí výrazný užitek. Už takhle stráví práci 10 hodin denně a nemá skoro žádný volný čas. Nyní, aby obdržel stejnou výši mzdy, nemusí pracovat 10 hodin, ale pouze 9 hodin. Peníze mu postačí ve stejné výši, jako měl předtím a raději vybere variantu, kde má o hodinu více volného času, aby si svůj majetek mohl užívat. Tomuto jevu říkáme důchodový efekt (Holman, 2013).

2.1.5. Vliv důchodového a substitučního efektu na nabídku práce

Na nabídku práce působí oba jevy současně. Při nižších mzdách převládá substituční jev, lidé chtějí více pracovat, aby obdrželi vyšší mzdy. A v jistém bodě se křivka začne otáčet zpět, začíná převládat důchodový efekt. Hranice od kdy lidé začínají chodit do práce méně, aby měli volný čas si své peníze užít (Holman, 2013).

2.1.6. Poptávka po práci

2.2. Nezaměstnanost

Nezaměstnanost se jako absolutní hodnota v praxi příliš nepoužívá. V praxi se používá míra nezaměstnanosti, abychom mohli lépe porovnávat nezaměstnanost v různě velkých okresech, krajích, státech, atd. Vypočteme ji poměrem mezi počtem nezaměstnaných proti součtu nezaměstnaných a zaměstnaných viz vzorec 1. Kde za L dosazujeme počet zaměstnaných a za U počet nezaměstnaných (Holman, 2013).

$$u = \frac{U}{U+L} \quad (1)$$

Koho považovat za nezaměstnaného? Označujeme tak pouze lidi, kteří jsou bez zaměstnání a hledají práci. Například studenty mezi nezaměstnané nezařazujeme. V praxi se nezaměstnanost zjišťuje pomocí evidence na úřadech práce, kam se nezaměstnaní hlásí z důvodu hledání nového zaměstnání či pro získání podpory z nezaměstnanosti. Takto zjištěnou nezaměstnanost nazýváme registrovanou, avšak skutečná nezaměstnanost bývá o nějaké procento vyšší než registrovaná. Ne všichni nezaměstnaní se hlásí na úřad práce, důvodů může být několik. Například pracovní poměr mi končí v půlce tohoto měsíce a začátkem dalšího měsíce už mi začíná jiný pracovní poměr (Holman, 2013).

2.2.2. Frikční nezaměstnanost

Nezaměstnanost rozdělujeme podle příčin na frikční, strukturální a cyklickou. O frikční nezaměstnanosti mluvíme, pokud nezaměstnaný hledá novou práci, avšak nemění svůj obor. Ze svého původního zaměstnání odešel nebo byl propuštěn. Děje se tak z důvodu stěhování, špatných vztahů na pracovišti a spoustu dalších. Doba, než naleznou své nové zaměstnání, může být delší z důvodu, že hledají konkrétní zaměstnání a “nesáhnou“ po první nabídce, která jim přijde pod ruku jako v následujícím příkladu (Holman, 2013).

Paní Božena pracuje v kadeřnictví, jejímu manželovi nabídli vysněnou práci v jiném městě. Dojíždění nepřipadá v úvahu, tak se celá rodina rozhodla přestěhovat. Božena tímto ztratila práci, kterou měla ráda. Nic jiného za svůj život nedělala a rozhodla se hledat novou práci ve stejném odvětví. V novém městě Božena první práci odmítla, protože mzda nebyla dostačující. Hledala dál a dostala se k jiné nabídce, jenže tohle kadeřnictví mělo špatnou pověst, proto ji také odmítla. Následující nabídku už přijala, ale byla po celou dobu od přestěhování až doteď nezaměstnaná (Holman, 2013).

2.2.3. Strukturální nezaměstnanost

Za strukturální považujeme takovou nezaměstnanost, pokud jsme propuštěni ze současné práce z důvodu redukce pracovních míst v daném odvětví a jsme nuceni si hledat zaměstnání v jiném oboru jako pan Cyril v našem příkladu (Holman, 2013).

Pan Cyril byl velice zručný a pracoval v textilním průmyslu, bohužel došlo k redukci zaměstnanců a byl propuštěn. Zprvu hledal v okolí podobné podniky, ale nikde nenabízeli volné pracovní pozice. Cyril se smířil s tím, že už práci podobnou té poslední nenajde a koukal se po volných pozicích v ostatních odvětvích. Zaujala ho nabídka v dopravě, práce

řidiče nákladního automobilu ho lákala, jenže neměl profesní a řidičský průkaz skupiny C. Zaměstnání řidiče si našel, až když získal oba průkazy, jenže to trvalo delší dobu, než paní Boženě. Strukturální nezaměstnanost je zpravidla delší než frikční nezaměstnanost z důvodu rekvalifikace (Holman, 2013).

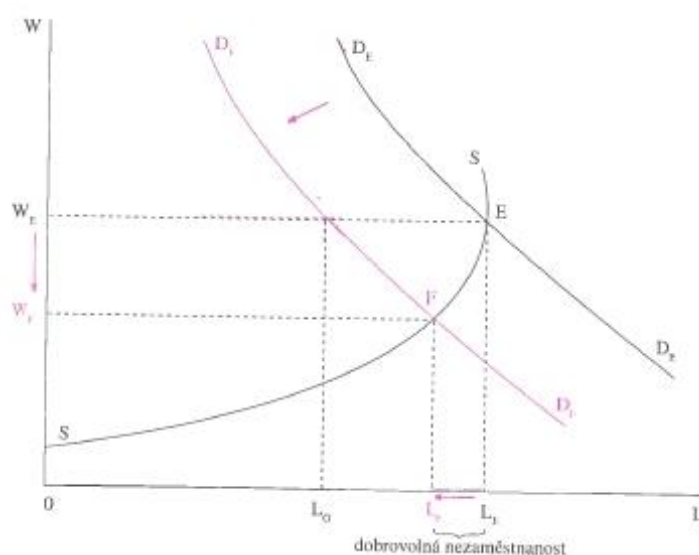
2.2.4. Cyklická nezaměstnanost

O cyklické nezaměstnanosti mluvíme v souvislosti s hospodářským cyklem. Slábnutí jednoho odvětví má vliv na celou ekonomiku. Nejdříve ochabuje jen dané odvětví, následně na to ochabují i odvětví, která jsou s ním spojená. Postupem času jak roste počet nezaměstnaných, klesají příjmy domácností a tím klesá i spotřeba ostatních produktů. Ostatní firmy z důvodu menší poptávky po jejich produktech jsou také nuceny propouštět (Holman, 2013).

2.2.5. Dobrovolná a nedobrovolná nezaměstnanost

Hlavním rozdílem mezi dobrovolnou a nedobrovolnou nezaměstnaností je to, že v případě dobrovolné můžeme nastoupit na pracovní pozici ihned, ale za nižší mzdu. U nedobrovolné nezaměstnanosti na nás pracovní příležitost nezbývá a musíme vyčkávat, zdali se nějaká na trhu práce objeví (Holman, 2013).

Graf 4 - Dobrovolná nezaměstnanost

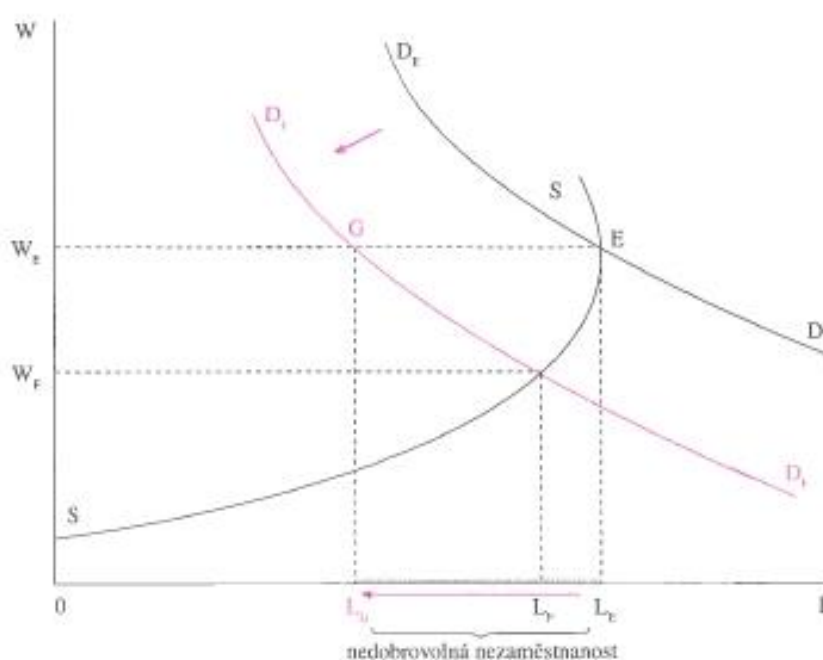


zdroj: Holman, 2005

Na obrázku 4 máme příklad dobrovolné nezaměstnanosti. Z nějaké příčiny poklesne nabídka práce z D_E na D_F . To vede k propouštění, dostáváme se z původní zaměstnanosti

L_E v rovnováze E na zaměstnanost L_G . Zvýšením počtu nezaměstnaných si firmy mohou dovolit snížit mzdy, protože fronta nezaměstnaných se rozrostla a jsou ochotni přijmout práci i za nižší mzdu, než pobírali v práci předchozí. Mzda klesá na výši W_F . Za tuto mzdu jsou ochotny firmy doplnit pracovníky, avšak na nové pracovní pozice se nehlásí všichni, kteří o práci přišli. Trh práce se dostává do nového rovnovážného stavu F a rozdíl mezi původním stavem zaměstnaných L_E a novým stavem zaměstnaných L_F nám tvoří tzv. dobrovolně nezaměstnaní. Lidé, kteří by práci našli, ale nechtějí za ní dostávat menší mzdu, tak radši nepracují vůbec.

Graf 5 - Nedobrovolná nezaměstnanost



zdroj: Holman, 2005

Zde je začátek stejný, jako u obrázku 4. Z nějakého důvodu poklesne poptávka po práci a firmy začínají propouštět. Zaměstnanost se dostává z bodu L_E na bod L_G . I když je nabídka práce podstatně vyšší než její poptávka, trh je omezen a z nějakého důvodu mzdy neklesají, jak bychom očekávali. To zapříčiní to, že na trhu práce se objeví spousta pracovníků, kteří by rádi pracovali i za nižší mzdu, než pobírali původně, ale takové místa neexistují. Těmto lidem říkáme nedobrovolně nezaměstnaní (Holman, 2013).

2.2.6. Příčiny nedobrovolné nezaměstnanosti

Důvodu nesnižování mezd při poklesu poptávky po práci je několik. Například odbory můžou zamezit poklesu mezd a vytvořit tak bariéru, přes kterou se mzda nedostane. Tento

jev nazýváme nepružnost mezd a má za důsledek nedobrovolnou nezaměstnanost (Holman, 2013).

2.2.6.1. Odbory

Ačkoli hlavním smyslem existence odborů je zlepšování podmínek zaměstnanců, jsou zde i výjimky, u kterých to neplatí. Jedním takový příkladem je, pokud si odbory s podniky dohodnou výši minimální mzdy, pod kterou nesmí mzda zaměstnance v podniku klesnout. Zatím to zní jako pohádka pro zaměstnance, ale zeptám se vás, co se stane, pokud poklesne poptávka po práci. Ano, podniky propustí i ty zaměstnance, kteří by přistoupili na nižší mzdu, jen aby si udrželi současné zaměstnání (Holman, 2013).

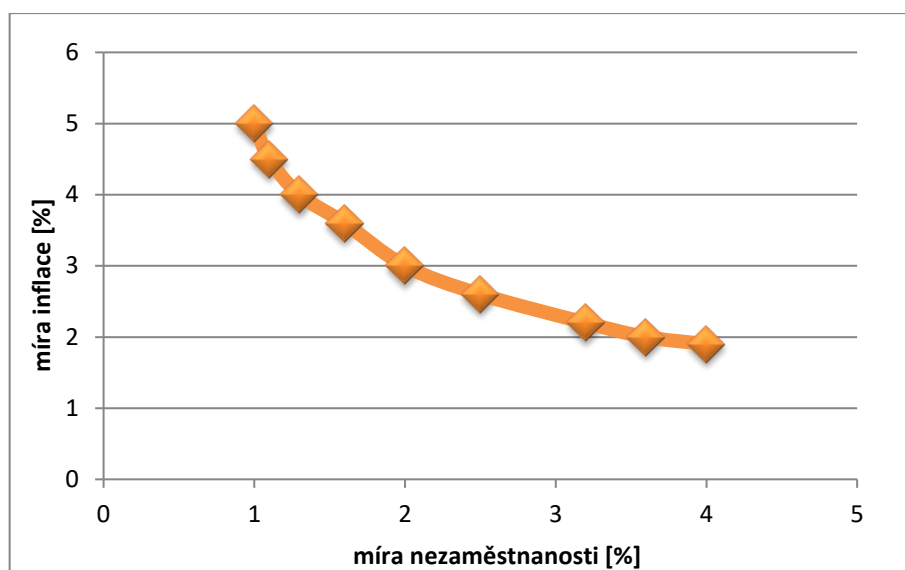
2.2.6.2. Stát

Obdobně jako odbory i stát přispívá ke vzniku nedobrovolně nezaměstnaných, uzákoněním minimální mzdy. Tím vytváří jakousi pomyslnou hranici, pod kterou nesmí mzda klesnout (Holman, 2013).

2.2.7. Substituční vztah mezi nezaměstnaností a inflací

Jako v životě, i v ekonomii nemůžeme mít všechno najednou. Vždy je něco na úkor něčeho jiného. Jistý pan Phillips přišel s tím, že nezaměstnanost a inflace mají mezi sebou vztah. Konkrétně pokud roste jedna hodnota, klesá druhá a naopak. Závislost nám popisuje Phillipsova křivka na grafu 6. V krátkém období se osvědčila, ale snažili se použít křivku i pro dlouhé období, ale hodnoty neodpovídaly. Zjistilo se, že v dlouhém období jsou nezaměstnanost a inflace na sobě nezávislé (Dornbusch & Fischer, 1994).

Graf 6 - Phillipsova křivka



zdroj: vlastní tvorba

3. Metodika

V této kapitole si rozebereme všechny statistické metody, které budeme dále používat k výpočtům.

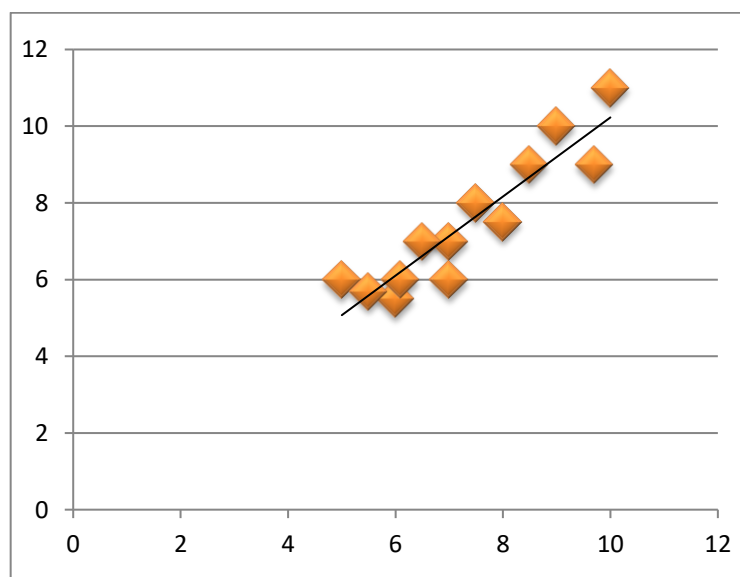
3.1. Korelační a regresní analýza

Doteď jsme se zajímali pouze o jednu proměnnou a tou byla míra nezaměstnanosti. Postupem přibývají další proměnné a my se potřebujeme zabývat vztahy mezi nimi. K tomuto problému nám slouží regresní a korelační analýza. Ve zkratce, v korelační analýze se zkoumá, zda mají proměnné mezi sebou závislost a do jaké míry. V regresní analýze, jde o odhad jedné proměnné pomocí druhé (Hindls & spol., 2002).

3.1.1. Korelační analýza

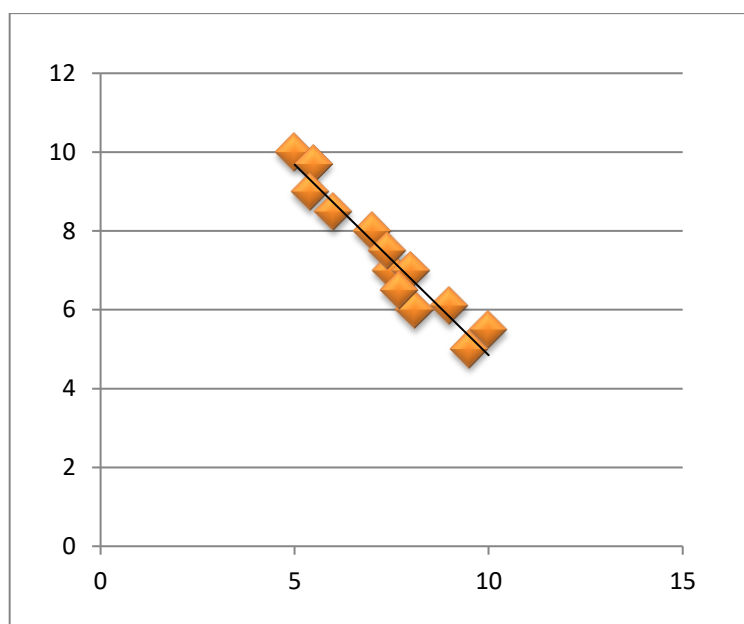
Korelace je vztah mezi jednou a druhou proměnnou, abychom mohli tento vztah kvantifikovat, musíme se bavit o nejznámějším koeficientu v této oblasti. Jedná se Pearsonův korelační koeficient. Nabývá hodnot od -1 do 1. Při hodnotě -1 mluvíme o naprosté negativní korelaci a při 1 naprosté kladné korelaci. Kladné hodnoty znamenají, že při zvyšování jedné roste i druhá hodnota, nehledě na to, která z proměnných je závislá na té druhé. U záporného koeficientu je to naopak. Znamená to, že pokud roste jedna, druhá klesá viz následující grafy 7 a 8 (Hindls & spol., 2002).

Graf 7 - Korelační koeficient blížící se 1



zdroj: vlastní tvorba

Graf 8 - Korelační koeficient blíží se -1



zdroj: vlastní tvorba

Vysvětlíme si to na příkladu. Představme si, že je chladný den a prodáváme čaj ve stánku u sjezdovky. Pro statistické účely si zaznamenáváme, kolik lidí projede okolo a kolik jsme prodali kelímků s čajem. Po několika dnech si z těchto dat vypočítáme korelační koeficient a vyjde nám 0,8. Co to znamená? Znamená to, že ve dnech, kdy projelo kolem stánku více lyžařů, jsme prodali více kelímků. Statistické údaje nás zaujaly natolik, že jsme si začali všimnout i ostatních proměnných, jako je třeba teplota vzduchu a pečlivě si po několik dalších dnů zapisovali údaje. Po týdnu prodávání jsme si jeden večer sedli k záznamům a kalkulačce a vypočítali si korelační koeficient mezi teplotou a počtem prodaných kelímků. Na kalkulačce jsme se dostali k výsledku -0,7 a my už víme, co nám výsledek sděluje. Že ve chladnějších dnech se prodává více kelímků s čajem než ve dnech teplejších.

Už nám jen zbývá si říct, jak korelační koeficient $\text{kor}(x,y)$ mezi proměnnými x a y vypočítáme. Pro výpočet používáme vzorec

$$\text{kor}(x, y) = \frac{\overline{xy} - \bar{x} \bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)}} \quad (2)$$

kde $\bar{x}$ se počítá jako aritmetický průměr viz rovnice 3.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Obdobně pak následuje výpočet $\bar{y}$, $\overline{xy}$, $\overline{x^2}$ a $\overline{y^2}$ (Hindls & spol., 1997).

3.1.2. Regresní analýza

Regresní analýza slouží pro odhady vysvětlované (závislé) proměnné pomocí jedné či více vysvětlujících (nezávislých) proměnných. Pokud si sestojíme regresní funkci, jsme schopni po dosažení ostatních hodnot odhadnout velikost odhadované hodnoty. Při jedné nezávislé proměnné mluvíme o jednoduché regresi, když jsou dvě a více, tak potom o vícenásobné regresi (Hindls & spol., 2002).

Budeme pro lepší názornost pokračovat v příkladu z korelační analýzy. Jelikož každý den vyrážíme na svah do našeho stánku a potřebujeme si sebou vzít dostatek pitné vody pro vaření čaje, protože ve stánku máme pouze rychlovarnou konvici, v které vaříme vodu na zalévání čaje. K případnému vaření čaje ze sněhu se neodvažujeme, kvůli sankcím z Krajské hygienické stanice. Abychom zbytečně nenosili příliš mnoho lahví nebo naopak, aby nám voda v půlce dne nedošla, musíme mít také odhad, kolik kelímků v jednotlivých dnech prodáme. Pro tento odhad si sestojíme jednoduchou regresní funkci z naměřených dat. Z dat jsme zjistili, že o víkendu a ve svátcích je na svahu daleko více lidí, než v pracovních dnech. Máme první nezávislou proměnnou, jako druhou volíme odhadovanou teplotu v poledne. Pomocí statistického programu pak vypočítáme regresní funkci, která může vypadat následovně

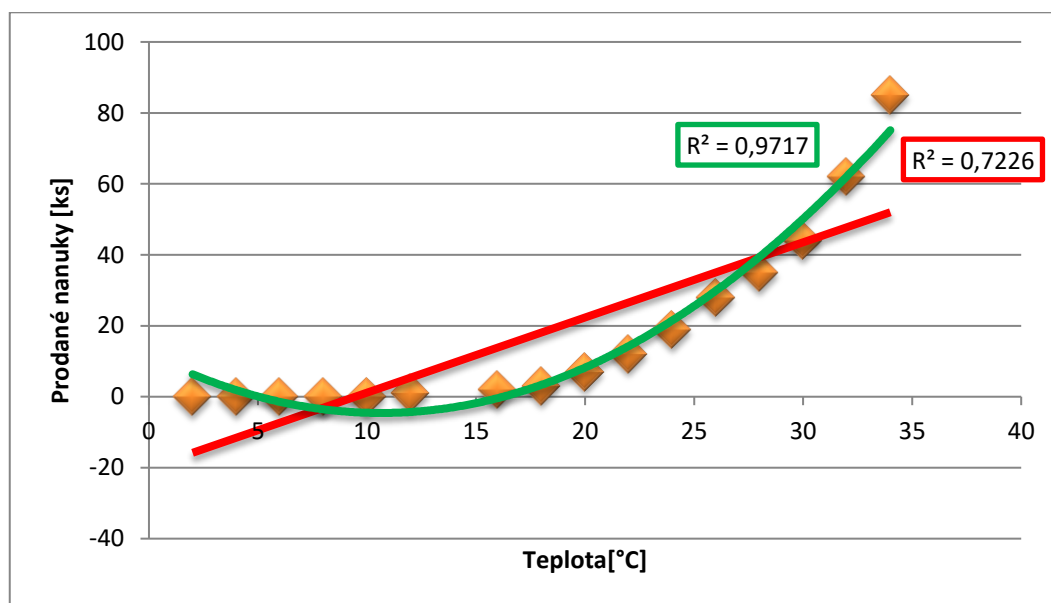
$$\text{prodaných kelímků} = 0,4 * \text{den} - 5,4 * \text{teplota}, \quad (4)$$

kde dosazujeme za den hodnotu 100 nebo 200, podle toho, zda je všední den nebo víkend či svátek. Odhadovanou teplotu v poledne dosazujeme ve stupních Celsia. Pro sobotu s předpovědí -5°C pak odhadujeme 107 kelímků.

3.1.3. Typy regresních funkcí

Pro svou jednoduchost se používá lineární regresní funkce, ale existuje mnohem více typů. Řadíme mezi ně parabolickou, polynomickou, hyperbolickou, logaritmickou, exponenciální regresní funkci. Každá z těchto funkcí má své uplatnění, vždy volíme takovou, která nejlépe odhadne závislou proměnnou. To znamená s nejmenším rozdílem (reziduem) mezi odhadovanou hodnotou a naměřenou hodnotou při stejných vstupních hodnotách (Hindls & spol., 1997).

Graf 9 - Typy regresních funkcí



zdroj: vlastní tvorba

Ve výše uvedeném grafu 9 zobrazujeme kusy prodaných nanuků při různých venkovních teplotách. Z grafu je patrné, že mají parabolický růst a ne lineární. Pro názornější ukázkou, jsou do grafu vloženy dva trendy, červený lineární a zelený parabolický. K trendům jsou přidány hodnoty spolehlivosti R^2 , které nabývají hodnot mezi 0 a 1. Čím blíže je hodnota k 1, tím trend lépe popisuje hodnoty v grafu. Při výběru typu pak volíme parabolickou regresi, protože předpokládáme, že bude mít méně reziduí.

3.2. Test normality

Většina statistických metod předpokládá rozdělení dat podle Gaussova normálního rozdělení. Máme hned několik možností, jak toto testovat (Netolická, 2008).

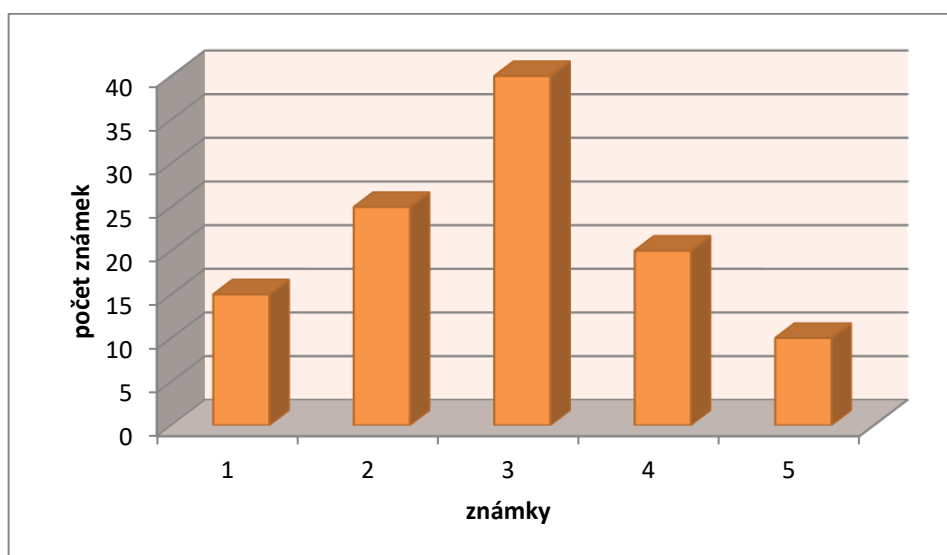
3.2.1. Porovnání mediánu a aritmetického průměru

Spíše orientační metoda, která nám rozliší jen to, pokud více hodnot na jedné straně nachází dál od středu (mediánu) normálního rozdělení. Všeobecně se u této metody připouští rozdíl mezi aritmetickým průměrem a mediánem do výše 10% (Netolická, 2008).

3.2.2. Histogram

Jednoduchá konstrukce je zde výhodou. Hlavní myšlenka tkví v tom, že vytváříme graf četností. Pokud máme jen několik možných výsledků, jako například udělené známky ve škole, rozumíme od jedničky po pětku, je konstrukce jednoduchá a výsledný graf by měl připomínat Gaussovu křivku viz graf 10.

Graf 10 - Udělené známky studentům



zdroj: vlastní tvorba

Problémem u histogramu je to, pokud množina řešení je spojitá, nebo existuje příliš mnoho různých výsledků. V takovém případě se používá rozdělení hodnot do intervalů a počet těchto intervalů se volí podle Sturgesova pravidla. Například u histogramu výšky studentů. Výpočet intervalů se počítá vzorcem

$$k = 1 + 3,3 * \log n, \quad (5)$$

kde k je počet intervalů a n je počet pozorování. Výsledek se doporučuje zaokrouhlit nahoru a na liché číslo, následuje pak jen výpočet šířky intervalu. Jednoduše se rozdíl

mezi nejmenší a největší naměřenou hodnotou vydělí počtem intervalů (Kasal a Hladíková, 1995).

3.2.3. Q-Q graf

Další možností, jak ověřit, jestli hodnoty měření připomínají normální rozdělení je použití Q-Q grafu. Je poněkud složitější na konstrukci, než histogram. Nejprve musíme seřadit hodnoty od nejmenších po největší a následně je přepočítat na kvantily pomocí vzorce

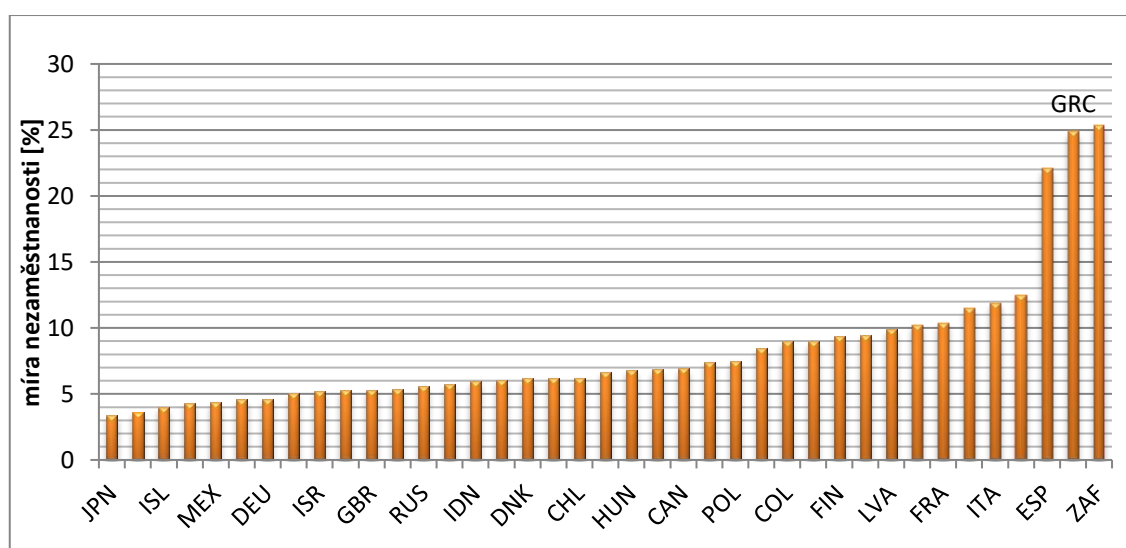
$$k_i = (x_i - \text{průměr hodnot } x) / \text{směrodatná odchylka hodnot } x \quad (6)$$

Na druhou osu nanášíme kvantily normálního normovaného rozdělení počítané inverzní funkcí k distribuční funkci. Do funkce dosazujeme

$$y_i = \frac{i - 0,5}{\text{počet hodnot}} \quad (7)$$

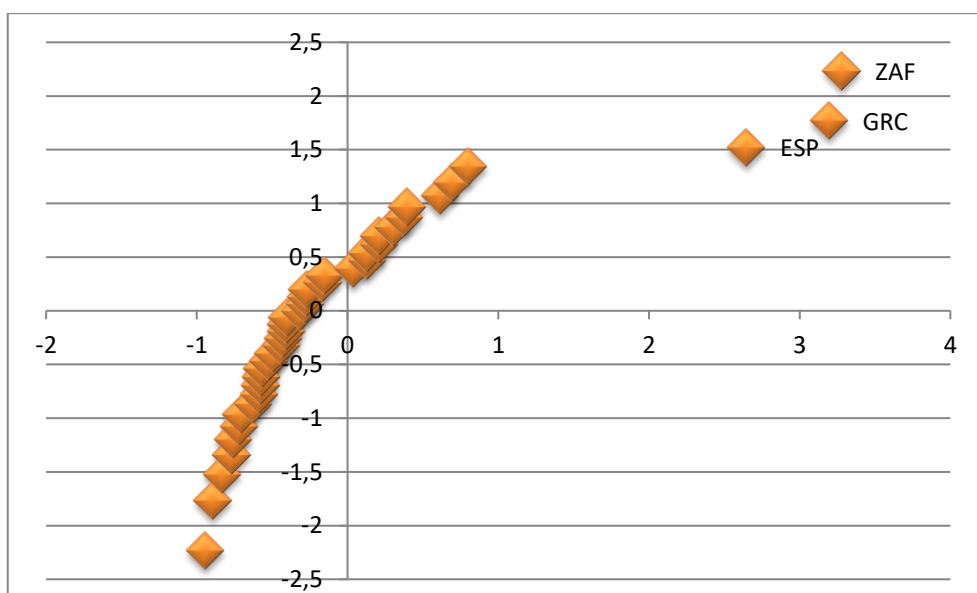
Pro názornou ukázkou si zvolíme míru nezaměstnanosti ve státech OECD v roce 2015. Už z grafu hodnot je patrné, že hodnoty Španělska, Řecka a Jihoafrické republiky budou problémem. Nyní sestojíme dva Q-Q grafy, v prvním budeme brát všechny data a druhý očistíme o data za tyto tři anomálie (Holčík a Komenda, 2015).

Graf 11- Míra nezaměstnanosti ve státech OECD v roce 2015



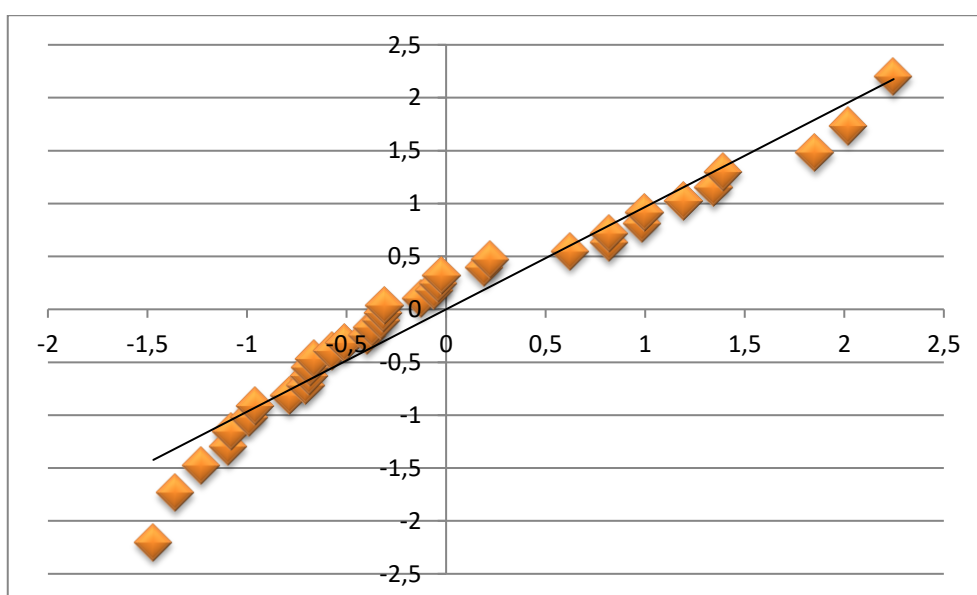
zdroj: Míry nezaměstnanosti ve státech OECD 2015, vlastní tvorba

Graf 12 - Q-Q graf měr nezaměstnaností států OECD



zdroj: Míry nezaměstnanosti ve státech OECD 2015, vlastní tvorba

Graf 13 - Q-Q graf měr nezaměstnaností států OECD bez ESP, GRC a ZAF



zdroj: Míry nezaměstnanosti ve státech OECD 2015, vlastní tvorba

Pro splnění normality by měly hodnoty ležet na jedné přímce, jak je tomu v grafu 13. V grafu 12 si nelze nevšimnout právě výše zmíněných tří států, které se výrazně liší od ostatních a jejich příčinou je takto zdeformovaný graf (Holčík a Komenda, 2015).

3.3. N-P graf reziduí

Nejprve si řekneme, co je to reziduum. Je to rozdíl, mezi skutečnou hodnotou a vypočtenou hodnotou vytvořeným modelem. N-P graf nám tedy zobrazí, do jaké míry mají rezidua vytvořeného modelu normální rozdělení. Vyhledáváme modely, které mají tyto rezidua nejbližší k spojnici trendu a zároveň spojnice se blíží počátku

Sestrojení je obdobné jako u Q-Q grafu. Na osu x vynášíme seřazené rezidua od nejmenší po největší hodnotu a na osu y nanášíme kvantily normálního normovaného rozdělení počítané inverzní funkcí k distribuční funkci přesně jako o kapitulu výše (Holčík a Komenda, 2015).

4. Proměnné do modelu

Nejprve je důležité zvolit si proměnné. Volíme takové, které nabývají různých hodnot v jednotlivých okresech. Například minimální mzda, ačkoli má jistě vliv na nezaměstnanost, se do našeho modelu příliš nehodí, protože pro všechny okresy má stejnou hodnotu. Špatná volba by také byla zahrnout do modelu proměnné, které vychází z nezaměstnanosti. Pokud bychom zvažovali podpory v nezaměstnanosti přepočtené na osoby v okrese, které souvisí a porostou s nezaměstnaností, to také není to, co hledáme. Snažíme se volit takové, které nejsou ovlivnitelné nezaměstnaností, ale onu nezaměstnanost můžou nějak ovlivňovat. Pokud jsme správně odhadli jejich vliv není jisté, to teprve budeme zkoumat. Zvolili jsme si několik proměnných, jako jsou:

4.1. Dálnice, silnice, železnice, vzdálenosti mezi okresními městy, krajskými městy a hlavním městem

Do modelu jsou zahrnuty vzdálenosti dálnic a silnic v 1. a 2. třídy jednotlivých okresech dostupné z webových stránek Českého statistického úřadu. Také vzdálenosti okresních měst od krajských měst a hlavního města, kde nebereme vzdálenost vzdušnou čarou, nýbrž cestou po pozemních komunikacích. Tyto vzdálenosti nepočítáme jako nejkratší, ale nejrychlejší varianty. Předpokládáme, že pro zásobování nepoužíváme trasy po málo frekventovaných silnicích třetích tříd, přes které by nás to vedlo, kdybychom hledali nejkratší vzdálenosti.

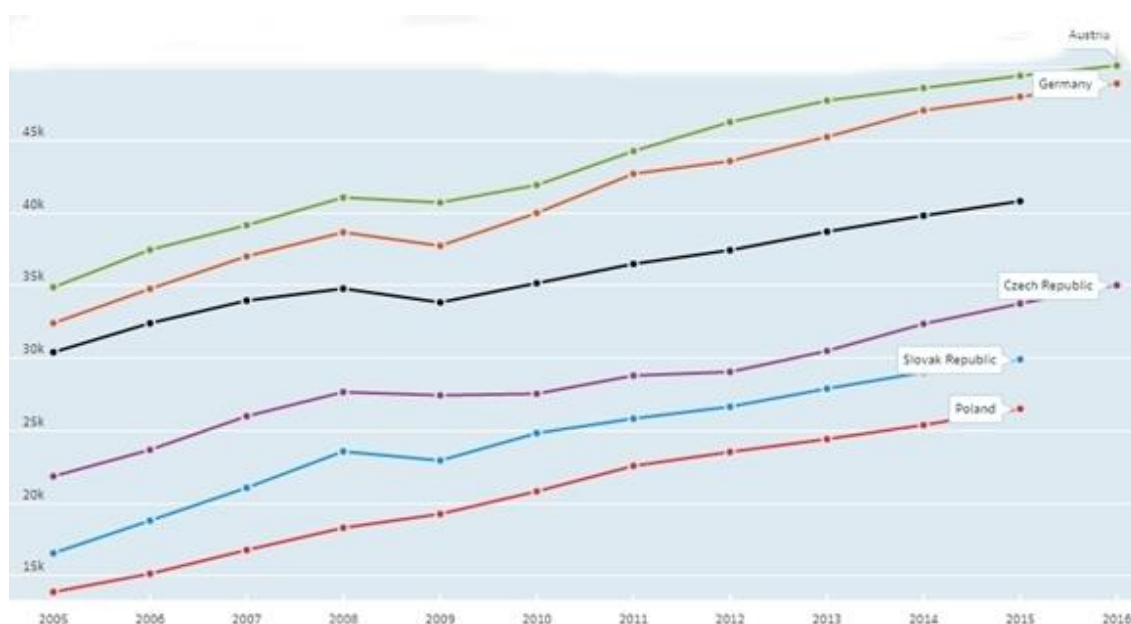
Pro snižování nezaměstnanosti je zásadní vytvářet pracovní pozice. Tyto pozice vytváří nové společnosti, které do dané lokality přijdou podnikat. Výběr lokality pro vytvoření společnosti do jisté míry závisí i na snadné možnosti zásobování, která je ovlivněna hustotou dopravní infrastruktury. Proto předpokládáme, že menší míra nezaměstnanosti bude v lokalitách s vyšším počtem komunikací a zároveň menší vzdáleností od krajských měst a hlavního města Prahy.

4.2. Státní hranice

V našem modelu zohledňujeme možnost dojíždění do zaměstnání do jiného členského státu. Proto bereme vzdálenost státní hranice jako další proměnnou. Jenže pozor, i naši sousedé k nám můžou dojíždět za prací a pro české občany nezbude tolik pracovních pozic. Dobré je se nad tímto zamyslet a říct si, jestli okresy v sousedství jiného státu jsou pro nás spíše přínosem nebo ztrátou. U hranic s Německem a Rakouskem spíše jezdí Češi

do zahraničí. Naproti tomu u hranic se Slovenskem a Polskem je to naopak, sousedé pracují více u nás, než my u nich. Česká ekonomika je na tom lépe, než slovenská či polská, ale naopak zaostáváme za německou nebo rakouskou ekonomikou. Toto můžeme vypočítat z následujícího grafu týkajícího se hrubého domácího produktu, přepočteného na jednoho občana. Hranice s Polskem a Slovenskem jsou brány jako záporné hodnoty, abychom vnesli do modelu jejich spíše negativní vliv na nezaměstnanost oproti hranicím s Německem a Rakouskem.

Graf 14 - HDP na obyvatele [\$/rok]



zdroj: HDP na obyvatele v ČR a sousedních státech

4.3. Vybavenost domácností informačními a komunikačními technologiemi

Dříve se zaměstnání hledalo hůře, dnes díky výpočetní technice a přístupu k internetu můžeme snáze najít zaměstnání, proto by v modelu neměla chybět vybavenost domácností informačními a komunikačními technologiemi.

4.4. Cizinci

Pro ekonomicky slabší státy je nepochybně Česká republika příležitostí, kterou si někteří nenechají ujít. V našem modelu uvažujeme jako další proměnou poměr mezi cizinci a celkovým obyvatelstvem v okrese. Cizinci jsou často ochotni pracovat za nižší mzdu, než

čeští zaměstnanci, kteří místo toho radši vyčkávají na lépe ohodnocené zaměstnání a tím přispívají k vyšším hodnotám nezaměstnanosti.

4.5. Průměrný věk

Nezaměstnanost mladých lidí, konkrétně mezi patnácti a dvaceti čtyřmi lety je vyšší než u ostatních věkových skupin, jak můžeme vypožorovat z tabulky 1. Z tohoto důvodu do modelu vkládáme průměrný věk obyvatelstva v okrese, i když může být dosti zkreslený vlivem příliš velkého nebo malého počtu lidí ve starobním důchodu.

Tabulka 1 - Míry nezaměstnanosti podle věku v ČR 2015

	Celkem	Muži	Ženy
15 až 19 let	31,8	31,1	32,7
20 až 24 let	10,7	9,5	12,6
25 až 29 let	6,7	6,2	7,4
30 až 34 let	5,9	4,8	7,5
35 až 39 let	4,6	3,1	6,5
40 až 44 let	3,6	2,3	5,1
45 až 49 let	3,5	2,9	4,2
50 až 54 let	4,2	3,6	4,9
55 až 59 let	5,1	4,5	5,8
60 až 64 let	2,8	3,3	2,0
65 a více let	1,1	1,1	1,2
15 až 64 let	5,1	4,3	6,2

zdroj: Soubor dat a tabulek

4.6. Průměrná mzda

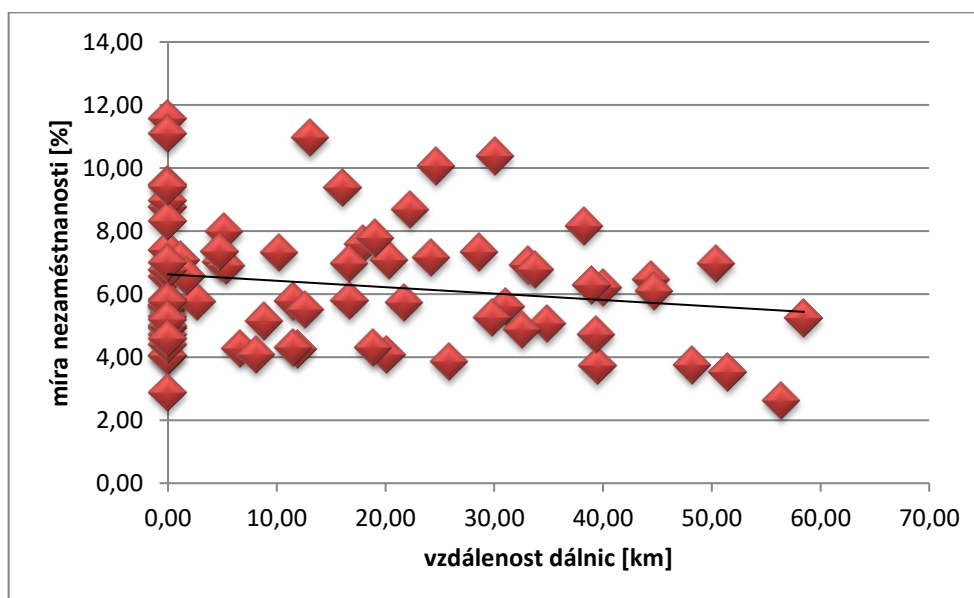
Je hned několik kritérií, podle kterých si lidé volí své budoucí zaměstnání. Odvětví, dojezdová vzdálenost, možnost kariérního růstu, dobrý kolektiv, ale na peníze slyší téměř každý. Proto použijeme pro náš model průměrné mzdy v okrese. Někteří lidé jsou schopni za vidinou vyššího ohodnocení jezdit za prací do jiných okresů, ačkoli stráví delší dobu na cestě do zaměstnání.

5. Zkoumání závislosti mezi mírou nezaměstnanosti a proměnnými

V kapitole 4 jsme pouze odhadovali, jak by mohly proměnné ovlivňovat míru nezaměstnanosti. V této kapitole si ukážeme, jestli opravdu se s měnícími hodnotami míra nezaměstnanosti mění tak, jak jsme odhadovali či nikoliv. Posuzovat to budeme tak, že na svislou osu budeme nanášet míru nezaměstnanosti a na vodorovnou volené proměnné. Následně do grafu přidáme lineární spojnici trendu. Pokud bude tento trend klesat, znamená to, že s většími hodnotami bude míra nezaměstnanosti klesat. Pro rostoucí trend je tomu přímo naopak.

5.1. Dálnice

Graf 15 - Míra nezaměstnanosti na vzdálenosti dálnic 2015

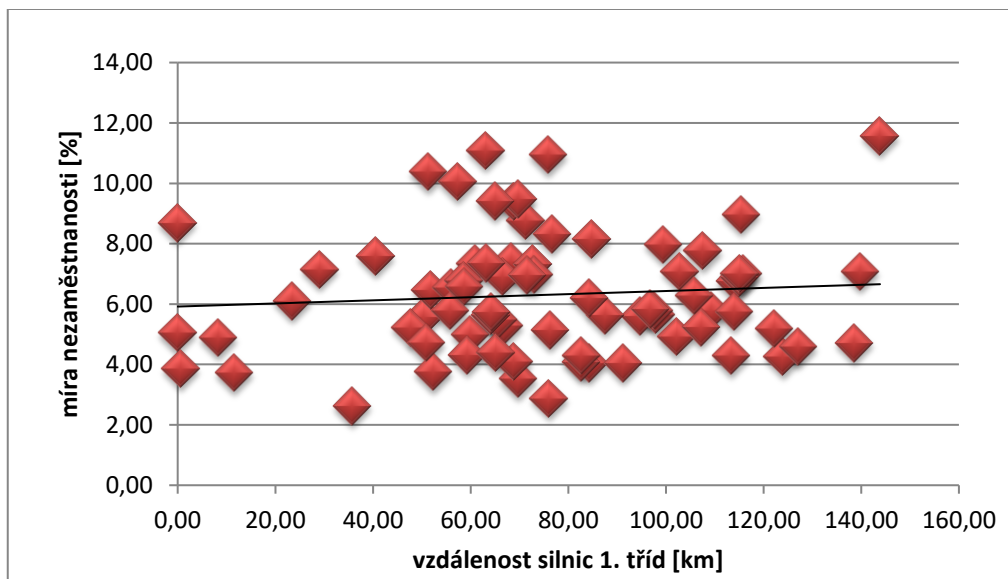


zdroj: Soubor dat a tabulek, vlastní tvorba

Trend v grafu 15 je klesající, to znamená, že jsme odhadovali správně, že větší míra nezaměstnanosti bude v okresech s menším počtem dálnic.

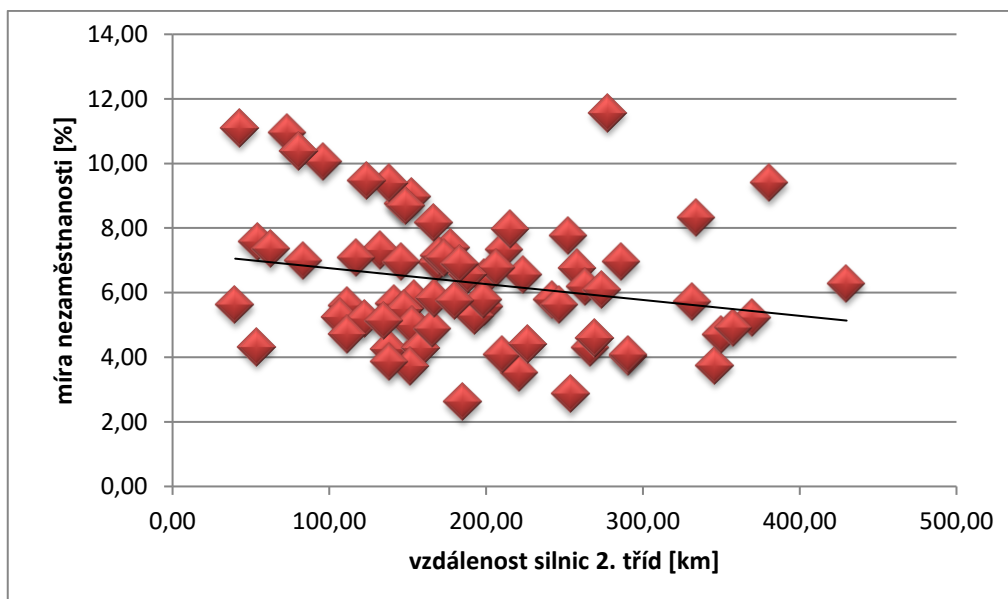
5.2. Silnice 1. a 2. tříd

Graf 16 - Míra nezaměstnanosti na vzdálenosti silnic 1. tříd v roce 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

Graf 17 - Míra nezaměstnanosti na vzdálenosti silnic 2. tříd v roce 2015



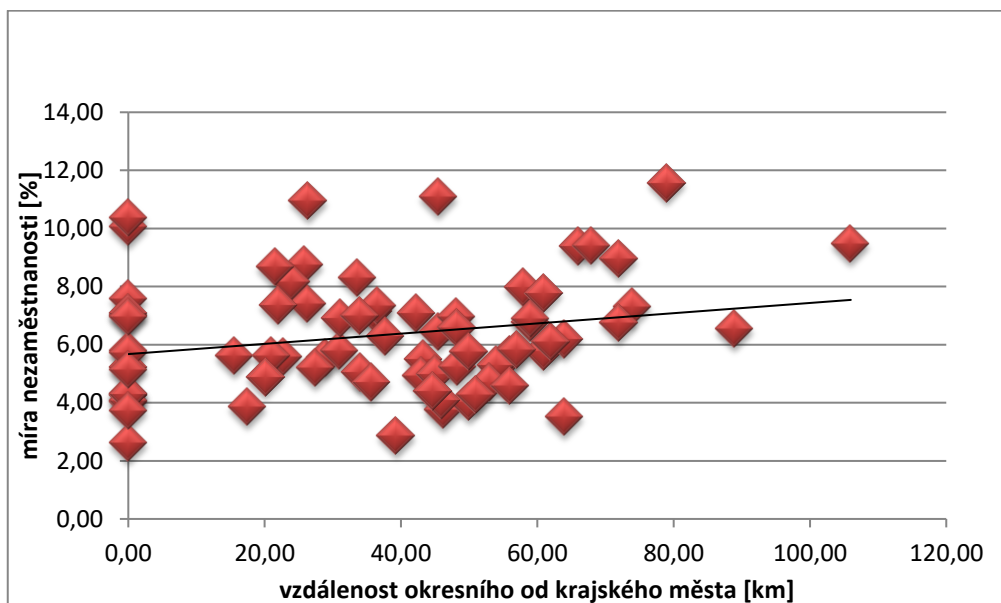
zdroj: Soubor dat a tabulek, vlastní tvorba

Z výše vyobrazeného grafu 16 lze vypožorovat, že u silnic prvních tříd je tomu jinak, než jsme odhadovali. Trend je opačný a to znamená, že se vyšší míra nezaměstnanosti vyskytuje v okresech s větším počtem kilometrů silnic 1. třídy. Ale trend u silnic druhé

třídy dopadl podle očekávání. Okresy, které mají více kilometrů silnic, jsou na tom se zaměstnaností výrazně lépe.

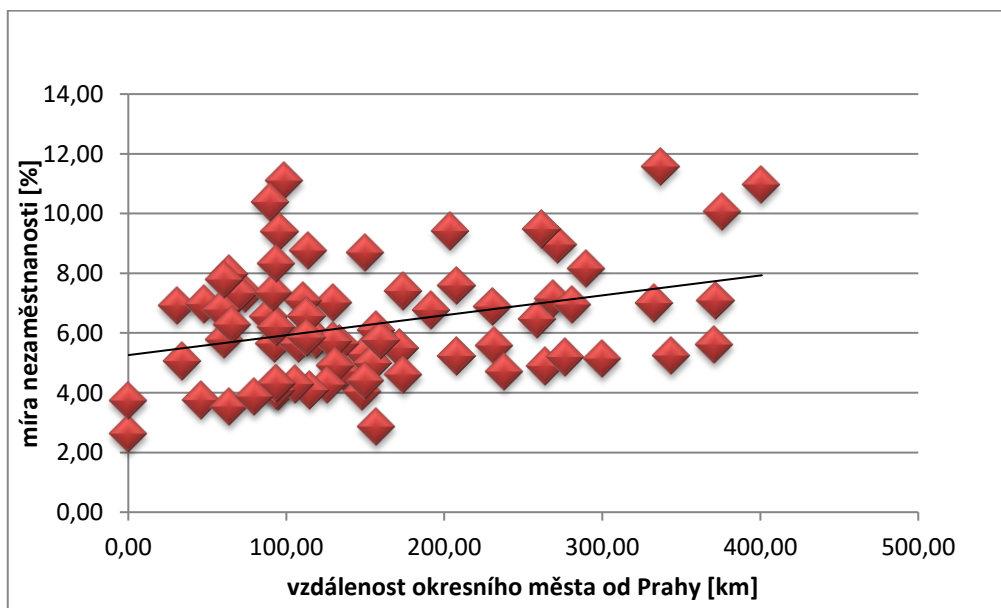
5.3. Vzdálenost okresního města od krajského města a hlavního města Prahy

Graf 18 - Míra nezaměstnanosti na vzdálenosti okresního od krajského města



zdroj: Soubor dat a tabulek, vlastní tvorba

Graf 19 - Míra nezaměstnanosti na vzdálenosti okresního města od Prahy

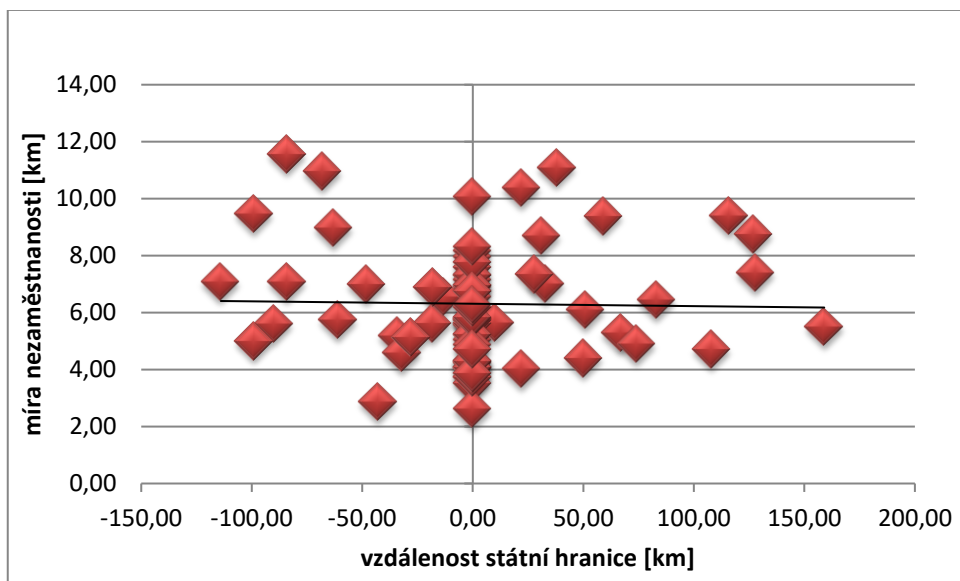


zdroj: Soubor dat a tabulek, vlastní tvorba

Podle výsledků grafů 18 a 19 můžeme tvrdit, že odhad byl správný. Čím blíže je okresní město k Praze a krajskému městu, klesá v okrese míra nezaměstnanosti.

5.4. Státní hranice

Graf 20 - Míra nezaměstnanosti na vzdálenosti státní hranice

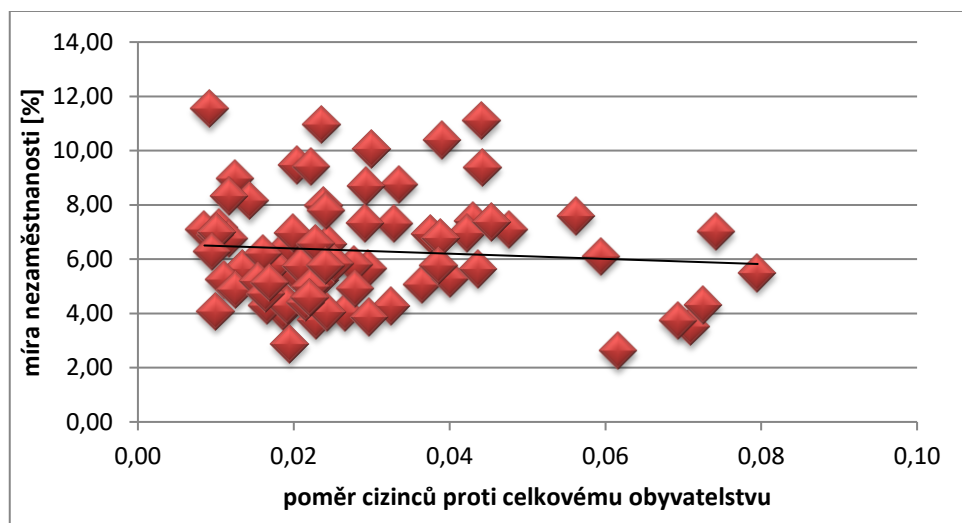


zdroj: Soubor dat a tabulek, vlastní tvorba

Názor, že státní hranice mají vliv na míru nezaměstnanosti, může znít zajímavě. Při této problematice na grafu 20 jsme zjistili něco jiného a konstantní trend je toho důkazem. Na vyšší míry nezaměstnanosti nemá vliv délka státní hranice.

5.5. Cizinci

Graf 21 - Míra nezaměstnanosti na poměru cizinců v roce 2015

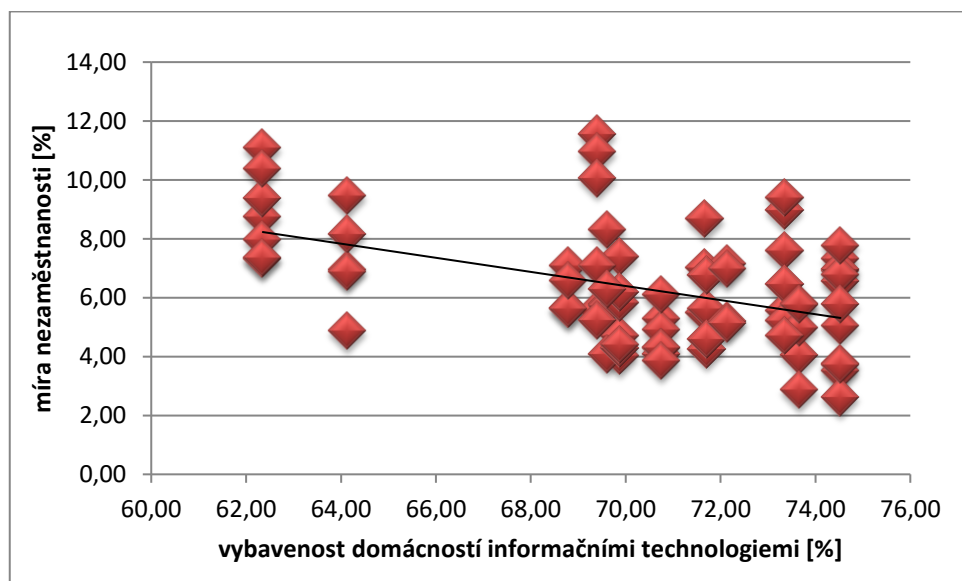


zdroj: Soubor dat a tabulek, vlastní tvorba

Ani v grafu 21, u poměru cizinců proti všem obyvatelům jsme nezaznamenali změnu na míře nezaměstnanosti. Výsledkem je, že na míru nezaměstnanosti v okrese nemá vliv poměr cizinců, kteří zde žijí.

5.6. Informační technologie

Graf 22 - Míra nezaměstnanosti na vybavenosti domácností IT v roce 2015

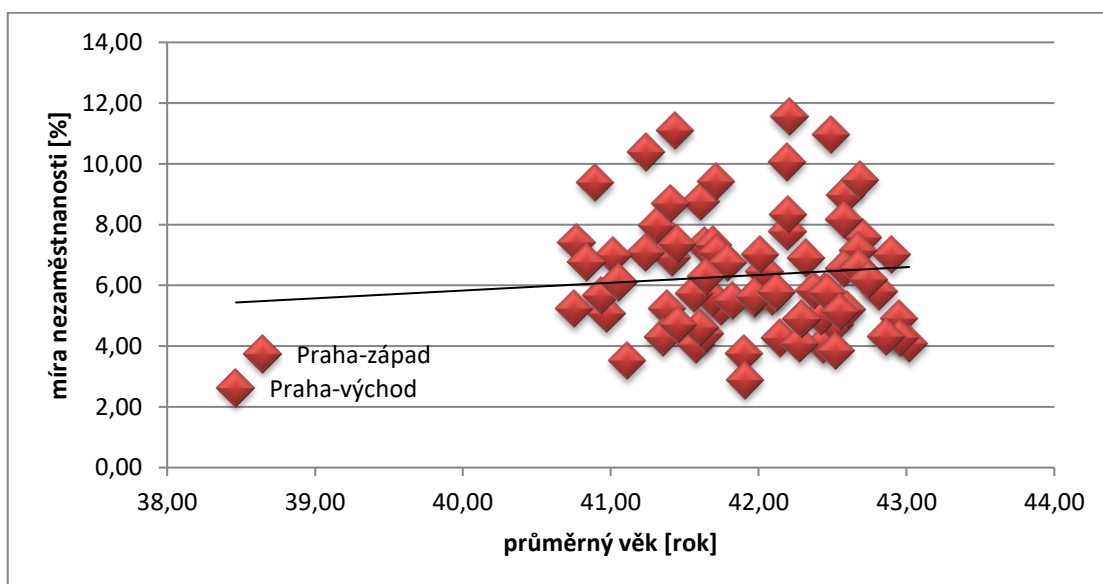


zdroj: Soubor dat a tabulek, vlastní tvorba

Značně klesající trend v grafu 22 nám říká, že v místech s lepší vybaveností informačními technologiemi je míra nezaměstnanosti podstatně nižší, než je tomu v okresech, kde jsou informační technologie v domácnostech v menším měřítku. Bohužel se nám nepodařilo shromáždit hodnoty pro jednotlivé okresy zvlášť, ale pouze hodnoty pro celé kraje, takže vždy mají okresy ze stejného kraje stejné hodnoty vybavenosti. V grafu je snadné tento jev vypožorovat. Jednotlivé okresy ze stejného kraje jsou vždy v grafu nad sebou, protože mají stejné hodnoty vybavenosti, akorát míra nezaměstnanosti se u těchto okresů mění.

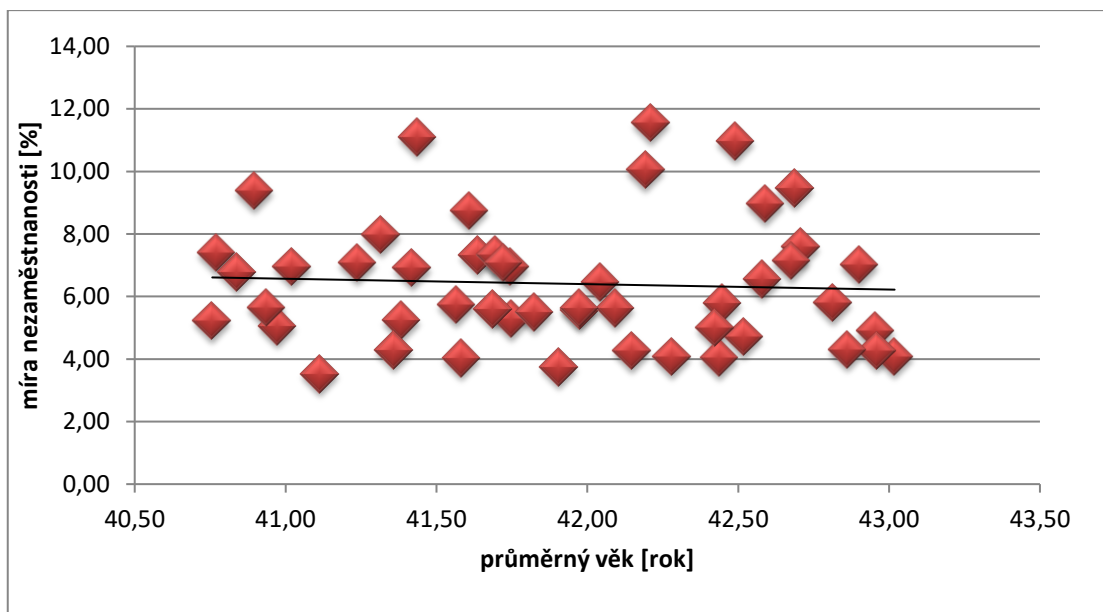
5.7. Průměrný věk

Graf 23 - Míra nezaměstnanosti na průměrném věku v roce 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

Graf 24 - Míra nezaměstnanosti na průměrném věku v roce 2015 bez okresů Praha-západ a Praha-východ

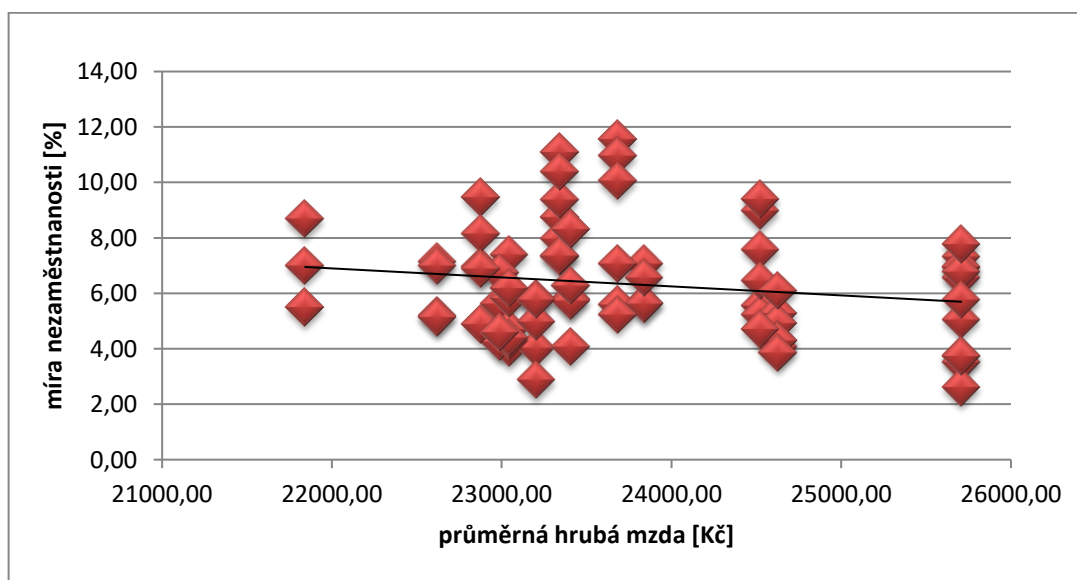


zdroj: Soubor dat a tabulek, vlastní tvorba

Na grafu 23 leží všechny okresy, ale na první pohled je vidět, že se na něm objevují dva, které se poměrně dost vzdalují od ostatních. Jde o okresy Praha-východ a Praha-západ. Pokud bychom uvažovali, že tyto hodnoty nebudeme brát v úvahu, pak výsledný graf 24 bude mít lehký klesající trend proti původnímu rostoucímu trendu v grafu 23. Na tomto lze vypožorovat, jak je trend lehce ovlivnitelný několika vychýlenými hodnotami od ostatních hodnot.

5.8. Průměrná hrubá mzda

Graf 25 - Míra nezaměstnanosti na průměrné hrubé mzdě v roce 2015

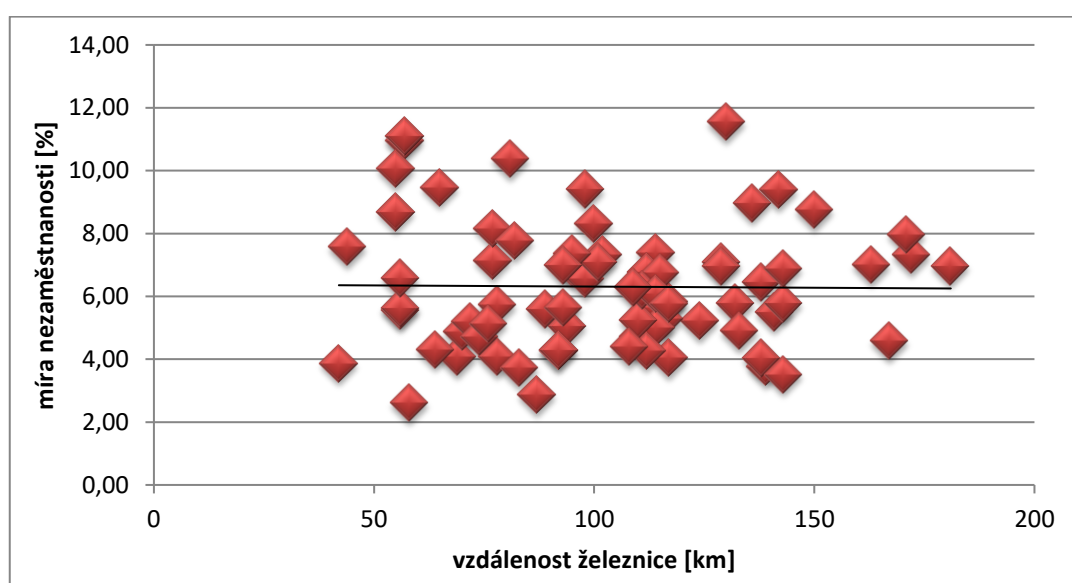


zdroj: Soubor dat a tabulek, vlastní tvorba

Podobně vypadající graf už jsme tu měli. Na grafu 25 vidíme několik okresů při stejné výši průměrné hrubé mzdy, protože jako u vybavenosti IT technologiemi, i zde byly hodnoty dostupné pouze pro celé kraje. To však nemění nic na tom, že jsme trend odhadli správně, míra nezaměstnanosti bývá nižší tam, kde dosahuje průměrná mzda vyšších hodnot.

5.9. Železnice

Graf 26 - Míra nezaměstnanosti na vzdálenosti železnice v roce 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

Z grafu 26 je patrné, že množství železničních kolejí nijak neovlivňuje míru nezaměstnanosti a vůbec by nevadilo, pokud bychom tuto proměnnou do našeho modelu neuvedli.

5.10. Korelační koeficienty

Tabulka 2 - Korelační koeficienty mezi mírou nezaměstnanosti a proměnnými

Variable	Correlations (sheet2015) Marked correlations are significant at $p < ,05000$ N=76 (Casewise deletion of missing data)					
	Dálnice	Silnice 1. třídy	Silnice 2. třídy	Státní hranice	Krajské město	Praha
Míra nezaměstnanosti	-0,176104	0,085126	-0,212142	-0,021363	0,226980	0,326044
	IT	Cizinci	Průměrný věk	Hrubá mzda	Železnice	
Míra nezaměstnanosti	-0,446344	-0,082923	0,105906	-0,174048	-0,012424	

zdroj: Soubor dat a tabulek, vlastní tvorba

Druhou možností, jak lze popsat závislosti mezi mírou nezaměstnanosti a vysvětlujícími proměnnými, je vytvoření si jednoduché tabulky s korelačními koeficienty. Rozdíl mezi kladným a záporným koeficientem je již rozepsán v metodice, takže teď už jen velmi stručně. Kladný koeficient vyjde, když rostou obě hodnoty současně, záporný koeficient vyjde, když roste jedna a klesá druhá. Pozorný čtenář si může všimnout, že u klesajícího trendu jsou hodnoty korelačního koeficientu záporné a u rostoucího trendu jsou kladné. Dokonce i výše koeficientu je zde patrná. U větších hodnot je zkosení větší. Proto také u posledního grafu s železnicí, je trend skoro rovnoběžný s osou x a korelační koeficient se nejvíce z uvedených blíží k nule.

6. Vytváření modelu

Nejprve budeme vytvářet model z dat roku 2015 ze všech proměnných, které jsme zmínili výše a postupným odebíráním vždy nejnevhodnější proměnné se dobereme ke konečnému modelu. Následně to samé uděláme i pro data z roku 2005 a 2010, abychom měli s čím model porovnat.

Pro začátek by ještě bylo vhodné si ověřit, jestli nejsou některá data v proměnných dvakrát. Sice by se jinak jmenovaly, hodnoty by byly mírně odlišné, ale chovaly by se podobně.

Pro představu, pokud bychom konstruovali model pro stánek s občerstvením, který má v nabídce párky v rohlíku, avšak dále z těchto dvou surovin nic jiného nevyrábí. V tomto příkladu se dá uvažovat, že spotřeba párků bude stejná jako spotřeba rohlíků, i když se může stát, že se budou v nějakém malém měřítku lišit.

Z uvedeného příkladu je doufám jasné, jak je tento problém myšlený. Ještě je dobré si připomenout, že i doplňkový jev mezi proměnnými nechceme.

Například tvoříme model pro dopravní společnost a mezi proměnnými máme celkovou spotřebu pohonných hmot, podíl dieselu na spotřebě a podíl benzínu na spotřebě. Pro vysvětlení neuvažujeme možnost, že by jsme spotřebovávali ještě jiný druh paliva. Potom se podíl dieselu rovná podílu benzínu odečteného od jedničky

$$\text{podíl dieselu} = 1 - \text{podíl benzínu} \quad (8)$$

Ted' jen zůstává otázkou, jak tyto proměnné odhalit. Nejjednodušší a zároveň spolehlivou metodu jsme již v této práci použili. Ano, máme na mysli Pearsonův korelační koeficient. Sestrojíme si tedy matici korelačních koeficientů a budeme hledat podezřelé hodnoty, tím je myšleno koeficienty blížíící se k 1 nebo -1.

Tabulka 3 - Korelační koeficienty mezi jednotlivými proměnnými 1. část

Variable	Correlations (sheet2015) Marked correlations are significant at $p < ,05000$ N=76 (Casewise deletion of missing data)				
	Dálnice	Silnice 1. třídy	Silnice 2. třídy	Státní hranice	Krajské město
Dálnice	1,000000	-0,461962	0,136784	0,012856	-0,274178
Silnice 1. třídy	-0,461962	1,000000	0,159974	-0,226327	0,219652
Silnice 2. třídy	0,136784	0,159974	1,000000	0,209496	0,172202
Státní hranice	0,012856	-0,226327	0,209496	1,000000	-0,017391

Krajské město	-0,274178	0,219652	0,172202	-0,017391	1,000000
Praha	-0,091273	0,247423	-0,180234	-0,279518	-0,008158
IT	0,095803	-0,075143	0,171609	-0,088514	0,033593
Cizinci	0,191457	-0,351889	-0,319381	0,332553	-0,209569
Průměrný věk	-0,424335	0,331563	-0,002393	-0,144532	0,162364
Hrubá mzda	0,346877	-0,322892	0,086395	-0,031407	0,139495
Železnice	-0,008327	0,391338	0,368784	0,179293	0,267088

zdroj: Soubor dat a tabulek, vlastní tvorba

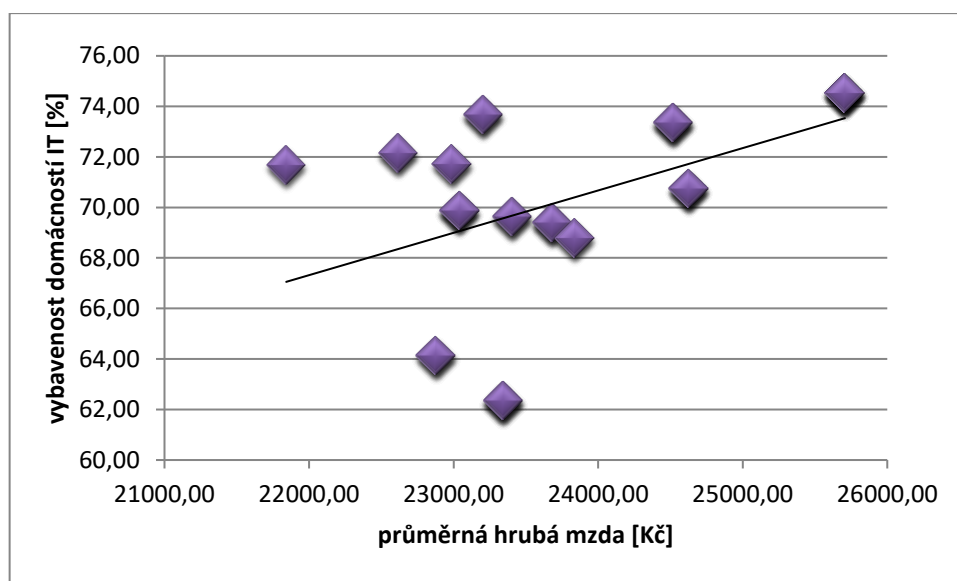
Tabulka 4 - Korelační koeficienty mezi jednotlivými proměnnými 2. část

Variable	Correlations (sheet2015) Marked correlations are significant at $p < ,05000$ N=76 (Casewise deletion of missing data)					
	Praha	IT	Cizinci	Průměrný věk	Hrubá mzda	Železnice
Dálnice	-0,091273	0,095803	0,191457	-0,424335	0,346877	-0,008327
Silnice 1. třídy	0,247423	-0,075143	-0,351889	0,331563	-0,322892	0,391338
Silnice 2. třídy	-0,180234	0,171609	-0,319381	-0,002393	0,086395	0,368784
Státní hranice	-0,279518	-0,088514	0,332553	-0,144532	-0,031407	0,179293
Krajské město	-0,008158	0,033593	-0,209569	0,162364	0,139495	0,267088
Praha	1,000000	-0,135527	-0,442558	0,322919	-0,408371	-0,137493
IT	-0,135527	1,000000	0,106077	-0,054895	0,482666	-0,120508
Cizinci	-0,442558	0,106077	1,000000	-0,399379	0,254472	0,045267
Průměrný věk	0,322919	-0,054895	-0,399379	1,000000	-0,378889	-0,062316
Hrubá mzda	-0,408371	0,482666	0,254472	-0,378889	1,000000	-0,086333
Železnice	-0,137493	-0,120508	0,045267	-0,062316	-0,086333	1,000000

zdroj: Soubor dat a tabulek, vlastní tvorba

Zjistili jsme, že nejvyšší hodnotu z tabulky tvoří korelace mezi informačními technologiemi a průměrnou hrubou mzdou. Ačkoli koeficient dosahuje čísla 0,482666, rozhodli jsme se obě tyto proměnné v modelu ponechat. Ještě pro kontrolu jsme si sestrojili graf, abychom mohli porovnat variabilitu, tím je myšleno jak jsou jednotlivé hodnoty od sebe vzdálené. Hodnoty v grafu 27 jsou poměrně dost daleko od sebe vzdálené a pomyslnou přímkou ani zdaleka netvoří.

Graf 27 - Variabilita IT a průměrné hrubé mzdy v roce 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

6.1. Model z dat roku 2015

V předchozích kapitolách jsme se rozhodli ponechat všechny proměnné a teď už následuje jen jakési prohnání dat mlýnkem. Vytvoříme si model a odebereme nejméně statisticky nevýznamnou možnost. Vytvoříme nový model ze zbylých proměnných a budeme takto pokračovat až do stavu, kdy nebude co "semlít". Pro výběr, jestli je proměnná statisticky významná, nebo ne se zjišťuje tak, že p-hodnota (p-value) nepřekročí hranici 0,05.

Tabulka 5 - Všechny proměnné pro regresní analýzu dat z roku 2015

Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,62152213 R2= ,38628976 Adjusted R2= ,28080831 F(11,64)=3,6622 p<,00047 Std.Error of estimate: 1,6748						
N=76	b*	Std.Err. of b*	b	Std.Err. of b	t(64)	p-value
Intercept			15,76829	14,62280	1,07834	0,284931
Dálnice	-0,123388	0,135428	-0,01425	0,01564	-0,91109	0,365665
Silnice 1. třídy	-0,008376	0,141935	-0,00051	0,00858	-0,05901	0,953128
Silnice 2. třídy	-0,027999	0,140233	-0,00065	0,00325	-0,19966	0,842381
Státní hranice	0,019428	0,121035	0,00077	0,00477	0,16051	0,872982
Krajské město	0,244150	0,116355	0,01894	0,00902	2,09833	0,039828

Praha	0,387956	0,129811	0,00793	0,00265	2,98863	0,003971
IT	-0,499385	0,120214	-0,26810	0,06454	-4,15414	0,000099
Cizinci	0,150987	0,148943	17,48283	17,24626	1,01372	0,314535
Průměrný věk	-0,005819	0,121703	-0,01409	0,29471	-0,04781	0,962015
Hrubá mzda	0,188132	0,143797	0,00035	0,00027	1,30832	0,195445
Železnice	-0,066334	0,131769	-0,00397	0,00789	-0,50341	0,616403

zdroj: Soubor dat a tabulek, vlastní tvorba

Není divu, že jako první vypadla proměnná pro průměrný věk. Celý trend je zde zkreslený dvěma hodnotami viz kapitola 5.7. Jako další předpokládáme, že vypadne vzdálenost silnic 1. třídy. Druhá nejhorší p-hodnota, hned v těsném závěsu za průměrným věkem.

Tabulka 6 - Proměnné pro regresní analýzu - odebrání první proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,62150450 R2= ,38626784 Adjusted R2= ,29184751 F(10,65)=4,0909 p<,00022 Std.Error of estimate: 1,6619					
	b*	Std.Err. of b*	b	Std.Err. of b	t(65)	p-value
Intercept			15,12983	5,91274	2,55885	0,012839
Dálnice	-0,122024	0,131371	-0,01409	0,01517	-0,92885	0,356405
Silnice 1. třídy	-0,008764	0,140611	-0,00053	0,00850	-0,06232	0,950495
Silnice 2. třídy	-0,027923	0,139144	-0,00065	0,00323	-0,20067	0,841579
Státní hranice	0,019474	0,120099	0,00077	0,00474	0,16215	0,871693
Krajské město	0,243564	0,114815	0,01889	0,00890	2,12136	0,037710
Praha	0,387501	0,128464	0,00792	0,00263	3,01642	0,003648
IT	-0,500099	0,118364	-0,26848	0,06354	-4,22508	0,000076
Cizinci	0,152248	0,145460	17,62885	16,84291	1,04666	0,299132
Hrubá mzda	0,189703	0,138915	0,00035	0,00026	1,36560	0,176771
Železnice	-0,065758	0,130206	-0,00394	0,00779	-0,50503	0,615245

zdroj: Soubor dat a tabulek, vlastní tvorba

Znovu se odstraní proměnná s nejhorší p-hodnotou a sestrojíme regresní funkci. Silnice 1. třídy k dalšímu modelování nepoužijeme.

Tabulka 7 - Proměnné pro regresní analýzu - odebrání druhé proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,62147499 R2= ,38623116 Adjusted R2= ,30253541 F(9,66)=4,6147 p<,00010 Std.Error of estimate: 1,6493					
	b*	Std.Err. of b*	b	Std.Err. of b	t(66)	p-value
Intercept			15,09136	5,83590	2,58595	0,011925
Dálnice	-0,118313	0,116218	-0,01366	0,01342	-1,01803	0,312382
Silnice 2. třídy	-0,028926	0,137163	-0,00067	0,00318	-0,21089	0,833624
Státní hranice	0,021638	0,114099	0,00085	0,00450	0,18964	0,850172
Krajské město	0,243931	0,113795	0,01892	0,00883	2,14359	0,035755
Praha	0,386238	0,125895	0,00790	0,00257	3,06795	0,003125
IT	-0,500543	0,117255	-0,26872	0,06295	-4,26884	0,000064
Cizinci	0,153109	0,143705	17,72859	16,63970	1,06544	0,290559
Hrubá mzda	0,190508	0,137266	0,00036	0,00026	1,38787	0,169844
Železnice	-0,069469	0,114914	-0,00416	0,00688	-0,60453	0,547563

zdroj: Soubor dat a tabulek, vlastní tvorba

Pro další krok se odstraní délka státních hranic, protože jiná proměnná nemá vyšší p-hodnoty než 0,850172.

Tabulka 8 - Proměnné pro regresní analýzu - odebrání třetí proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,62120585 R2= ,38589671 Adjusted R2= ,31257095 F(8,67)=5,2628 p<,00004 Std.Error of estimate: 1,6374					
	b*	Std.Err. of b*	b	Std.Err. of b	t(67)	p-value
Intercept			15,28209	5,70707	2,67775	0,009313
Dálnice	-0,120096	0,115001	-0,01387	0,01328	-1,04431	0,300097
Silnice 2. třídy	-0,019725	0,127369	-0,00046	0,00295	-0,15487	0,877393
Krajské město	0,244581	0,112922	0,01897	0,00876	2,16592	0,033880
Praha	0,384880	0,124783	0,00787	0,00255	3,08439	0,002962
IT	-0,504065	0,114939	-0,27061	0,06171	-4,38549	0,000042
Cizinci	0,164234	0,130238	19,01679	15,08037	1,26103	0,211672
Hrubá mzda	0,187784	0,135527	0,00035	0,00025	1,38558	0,170469

Železnice	-0,070521	0,113952	-0,00422	0,00682	-0,61887	0,538100
-----------	-----------	----------	----------	---------	----------	----------

zdroj: Soubor dat a tabulek, vlastní tvorba

V tomto kroku se vyřadí proměnná Silnice 2. třídy s hodnotou 0,877393.

Tabulka 9 - Proměnné pro regresní analýzu - odebrání čtvrté proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,62102890 R2= ,38567689 Adjusted R2= ,32243774 F(7,68)=6,0987 p<,00002 Std.Error of estimate: 1,6256					
	b*	Std.Err. of b*	b	Std.Err. of b	t(68)	p-value
Intercept			15,29539	5,66533	2,69982	0,008746
Dálnice	-0,124529	0,110579	-0,01438	0,01277	-1,12616	0,264057
Krajské město	0,244034	0,112054	0,01893	0,00869	2,17782	0,032892
Praha	0,391741	0,115814	0,00801	0,00237	3,38250	0,001195
IT	-0,508898	0,109824	-0,27320	0,05896	-4,63378	0,000017
Cizinci	0,174690	0,110571	20,22743	12,80308	1,57989	0,118773
Hrubá mzda	0,189553	0,134072	0,00035	0,00025	1,41381	0,161979
Železnice	-0,077646	0,103500	-0,00465	0,00619	-0,75020	0,455720

zdroj: Soubor dat a tabulek, vlastní tvorba

Ze 7 zbývajících proměnných se volí jako nejslabší délka železnic v okrese, která se nebude podílet na dalším modelování. Asi vás napadá, proč rovnou nevytvoříme model jen ze tří statisticky významných proměnných? Je zde možnost, že postupným odebráním se nám může navýšit počet statisticky významných vysvětlujících (proměnných). S každým dalším krokem se mění p-hodnoty a stává se, že se ze statisticky nevýznamné stane statisticky významná.

Tabulka 10 - Proměnné pro regresní analýzu - odebrání páté proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,61692172 R2= ,38059241 Adjusted R2= ,32673089 F(6,69)=7,0661 p<,00001 Std.Error of estimate: 1,6205					
	b*	Std.Err. of b*	b	Std.Err. of b	t(69)	p-value
Intercept			13,80205	5,28727	2,61043	0,011083
Dálnice	-0,136653	0,109044	-0,01578	0,01259	-1,25318	0,214369
Krajské město	0,215843	0,105230	0,01674	0,00816	2,05115	0,044051
Praha	0,407540	0,113522	0,00833	0,00232	3,58998	0,000614

IT	-0,504685	0,109332	-0,27094	0,05869	-4,61609	0,000018
Cizinci	0,168854	0,109947	19,55173	12,73084	1,53578	0,129166
Hrubá mzda	0,210298	0,130773	0,00039	0,00024	1,60811	0,112378

zdroj: Soubor dat a tabulek, vlastní tvorba

Konec modelování se pomalu, ale jistě blíží a proměnné ubývají. Mezi množinu vyřazených se nyní zařadí i vzdálenost dálnic.

Tabulka 11 - Proměnné pro regresní analýzu - odebrání šesté proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,60538786 R2= ,36649446 Adjusted R2= ,32124407 F(5,70)=8,0993 p<,00000 Std.Error of estimate: 1,6271					
	b*	Std.Err. of b*	b	Std.Err. of b	t(70)	p-value
Intercept			15,81611	5,05758	3,12721	0,002571
Krajské město	0,259890	0,099590	0,02016	0,00772	2,60959	0,011077
Praha	0,391911	0,113293	0,00801	0,00232	3,45926	0,000926
IT	-0,488928	0,109048	-0,26248	0,05854	-4,48360	0,000028
Cizinci	0,159942	0,110163	18,51979	12,75584	1,45187	0,151006
Hrubá mzda	0,145032	0,120443	0,00027	0,00022	1,20415	0,232588

zdroj: Soubor dat a tabulek, vlastní tvorba

Další jako nevýznamnou vysvětlující vyřazujeme průměrnou hrubou mzdu a hned v následujícím kroku se rozhodne, jak na tom bude poměr cizinců.

Tabulka 12 - Proměnné pro regresní analýzu - odebrání sedmé proměnné

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,59445106 R2= ,35337206 Adjusted R2= ,31694232 F(4,71)=9,7001 p<,00000 Std.Error of estimate: 1,6322					
	b*	Std.Err. of b*	b	Std.Err. of b	t(71)	p-value
Intercept			19,95893	3,71880	5,36704	0,000001
Krajské město	0,280902	0,098359	0,02179	0,00763	2,85588	0,005624
Praha	0,348006	0,107605	0,00711	0,00220	3,23410	0,001853
IT	-0,427208	0,096556	-0,22935	0,05184	-4,42446	0,000034
Cizinci	0,175275	0,109771	20,29518	12,71043	1,59673	0,114767

zdroj: Soubor dat a tabulek, vlastní tvorba

Očekávání se naplnilo a nyní se vytvoří model už jen ze tří statisticky významných závislých.

Tabulka 13 - Výsledek regresní analýzy z roku 2015

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015) R= ,57458864 R2= ,33015211 Adjusted R2= ,30224178 F(3,72)=11,829 p<,00000 Std.Error of estimate: 1,6497					
	b*	Std.Err. of b*	b	Std.Err. of b	t(72)	p-value
Intercept			20,53685	3,740758	5,49002	0,000001
Krajské město	0,243227	0,096509	0,01886	0,007485	2,52024	0,013945
Praha	0,271414	0,097353	0,00555	0,001990	2,78793	0,006780
IT	-0,417731	0,097405	-0,22426	0,052292	-4,28860	0,000055

zdroj: Soubor dat a tabulek, vlastní tvorba

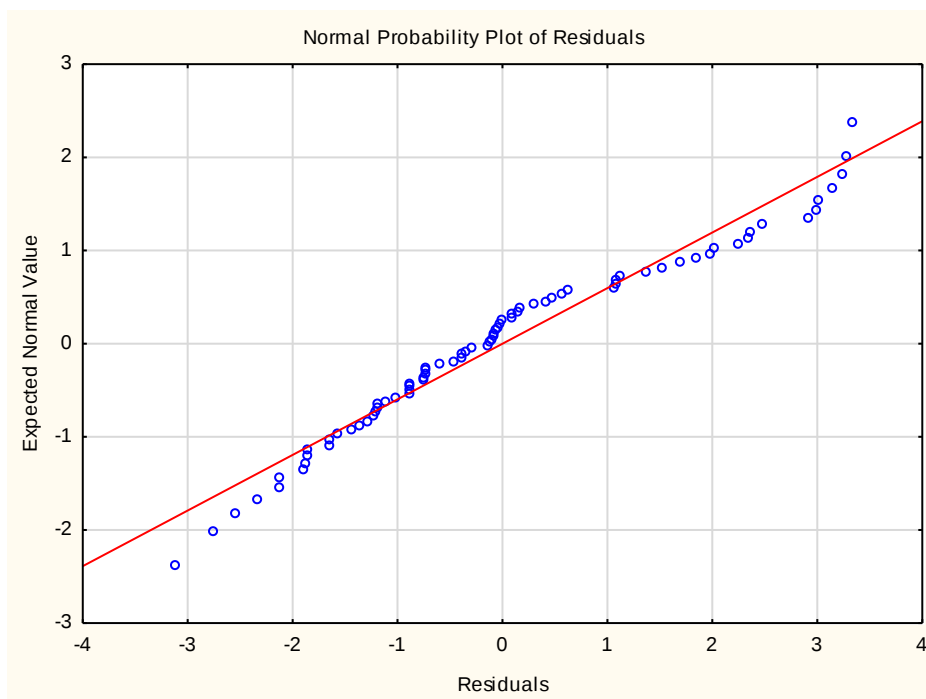
Výsledná regresní funkce pak má tvar

$$MN_x = 20,53685 + 0,01886 * KM_x + 0,00555 * PR_x - 0,22426 * IT_x, \quad (9)$$

ve které MN_x je odhadovaná míra nezaměstnanosti okresu x, KM_x je vzdálenost okresu x od krajského města, PR_x je vzdálenost okresního města okresu x od Prahy a IT_x je průměrná vybavenost domácností informačními technologiemi v okrese x.

Každý odhad není stoprocentní, proto se dále budeme bavit o již zmíněném reziduu. Pro opakování, reziduum je rozdíl mezi původní naměřenou hodnotou a odhadovanou hodnotou. Přirozeně chceme mít tento rozdíl co nejmenší, protože to znamená, že náš odhad je přesnější. Pro porovnání slouží tzv. N-P graf reziduí. V ideálním případě, by rezidua měla ležet na přímce. I když rezidua u sestrojeného N-P grafu neleží na přímce, také se od přímky příliš nevzdalují, což jsme posoudili jako uspokojivé.

Graf 28 - N-P graf reziduí modelu z roku 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

6.2. Model z dat roku 2005

Totéž, co v kapitole 6.1., vytvoříme i zde. Postup bude naprosto stejný, jen budeme používat data z roku 2005. Pro zkrácení zde uvedeme pouze tabulku finálního modelu a nikoliv jednotlivé kroky.

Tabulka 14 - Výsledek regresní analýzy z roku 2005

Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2005) R= ,67724346 R2= ,45865870 Adjusted R2= ,42816060 F(4,71)=15,039 p<,00000 Std.Error of estimate: 2,0166						
N=76	b*	Std.Err. of b*	b	Std.Err. of b	t(71)	p-value
Intercept			51,74432	12,87893	4,01775	0,000144
Dálnice	-0,260174	0,090325	-0,04248	0,01475	-2,88044	0,005247
Praha	0,337669	0,089399	0,00932	0,00247	3,77708	0,000327
IT	-0,327115	0,090435	-0,23811	0,06583	-3,61715	0,000554
Průměrný věk	-0,264469	0,089859	-0,96607	0,32825	-2,94315	0,004387

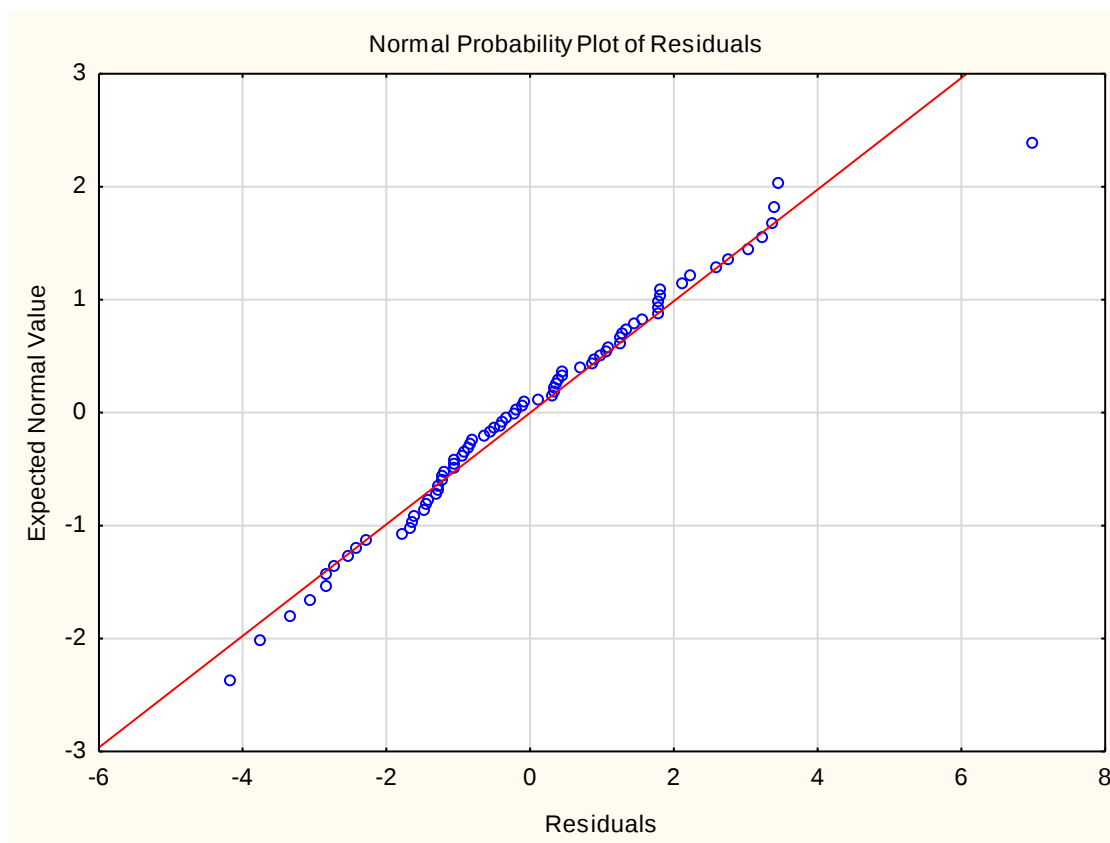
zdroj: Soubor dat a tabulek, vlastní tvorba

V tomto roce jsme měli také zprvu tři statistické významné proměnné, jen v průběhu jednotlivých kroků mezi ně přibyla i čtvrtá s názvem Dálnice. Z výsledné tabulky 14 si sestojíme regresní funkci

$$MN_x = 51,74432 - 0,04248 * D_x + 0,00932 * PR_x - 0,23811 * IT_x - 0,96607 V_x, \quad (10)$$

ve které MN_x je odhadovaná míra nezaměstnanosti okresu x , D_x je vzdálenost dálnic v okrese x , PR_x je vzdálenost okresního města okresu x od Prahy, IT_x je průměrná vybavenost domácností informačními technologiemi v okrese x a V_x je průměrný věk v okrese x .

Graf 29 - N-P graf reziduí modelu z roku 2005



zdroj: Soubor dat a tabulek, vlastní tvorba

N-P graf je přesvědčivější, než u modelu z roku 2015, protože rezidua jsou mnohem blíže k přímce. To znamená, že odhadované hodnoty se blíží více ke skutečným hodnotám a tato regresní funkce je přesnější.

6.3. Model z dat roku 2010

Postup nám zůstává stejný jako u obou modelů v předchozích kapitolách, jen použijeme údaje z roku 2010 a dostaneme se postupným odebíráním proměnných k následující tabulce.

Tabulka 15 - Výsledek regresní analýzy z roku 2010

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2010) R= ,67251741 R2= ,45227967 Adjusted R2= ,42142219 F(4,71)=14,657 p<,00000 Std.Error of estimate: 1,5399					
	b*	Std.Err. of b*	b	Std.Err. of b	t(71)	p-value
Intercept			17,33058	3,204666	5,40792	0,000001
Hranice	0,186632	0,091681	0,00754	0,003705	2,03567	0,045517
Krajské město	0,404563	0,087854	0,03217	0,006985	4,60494	0,000018
Praha	0,472552	0,091865	0,00990	0,001925	5,14397	0,000002
IT	-0,335540	0,088266	-0,20980	0,055190	-3,80146	0,000301

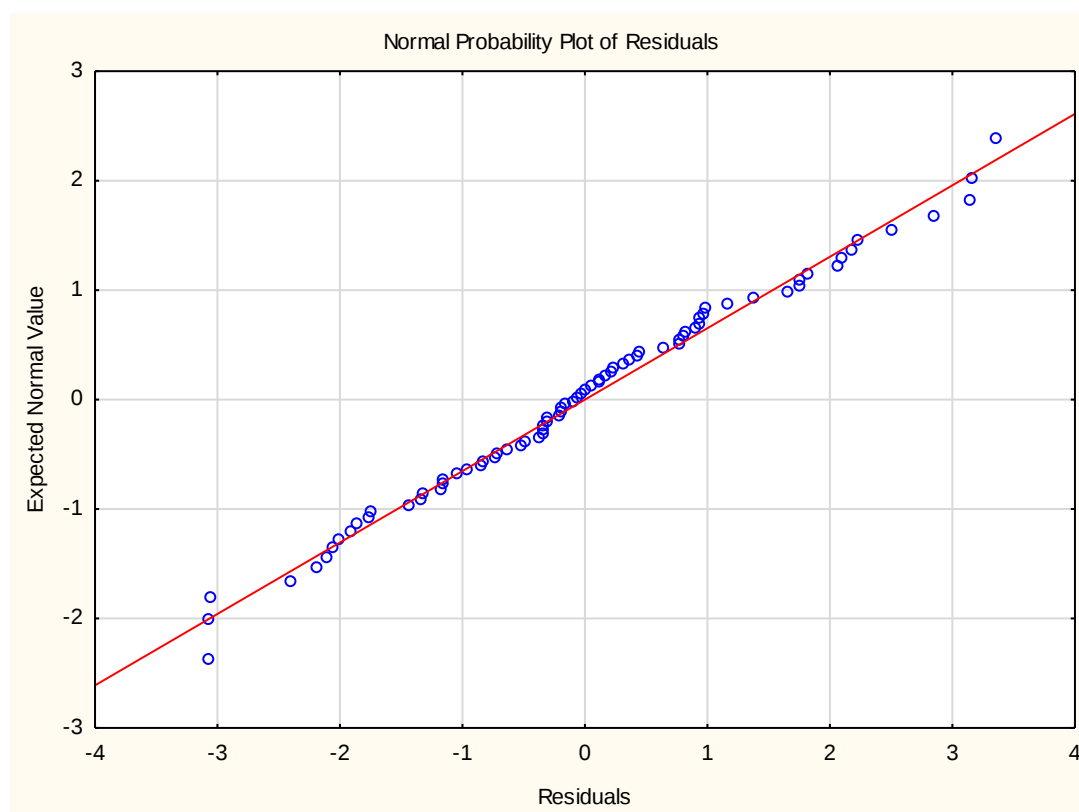
zdroj: Soubor dat a tabulek, vlastní tvorba

Výsledná regresní funkce vypadá

$$MN_x = 17,33058 + 0,00754 * H_x + 0,03217 * KM_x + 0,00990 * PR_x - 0,20980 IT_x, \quad (11)$$

kde MN_x je odhadovaná míra nezaměstnanosti okresu x, H_x je státní hranice okresu x, KM_x je vzdálenost okresu x od krajského města, PR_x je vzdálenost okresního města okresu x od Prahy a IT_x je průměrná vybavenost domácností informačními technologiemi v okrese x. I zde je N-P graf uspokojivější, než graf 28.

Graf 30 - N-P graf reziduí modelu z roku 2010



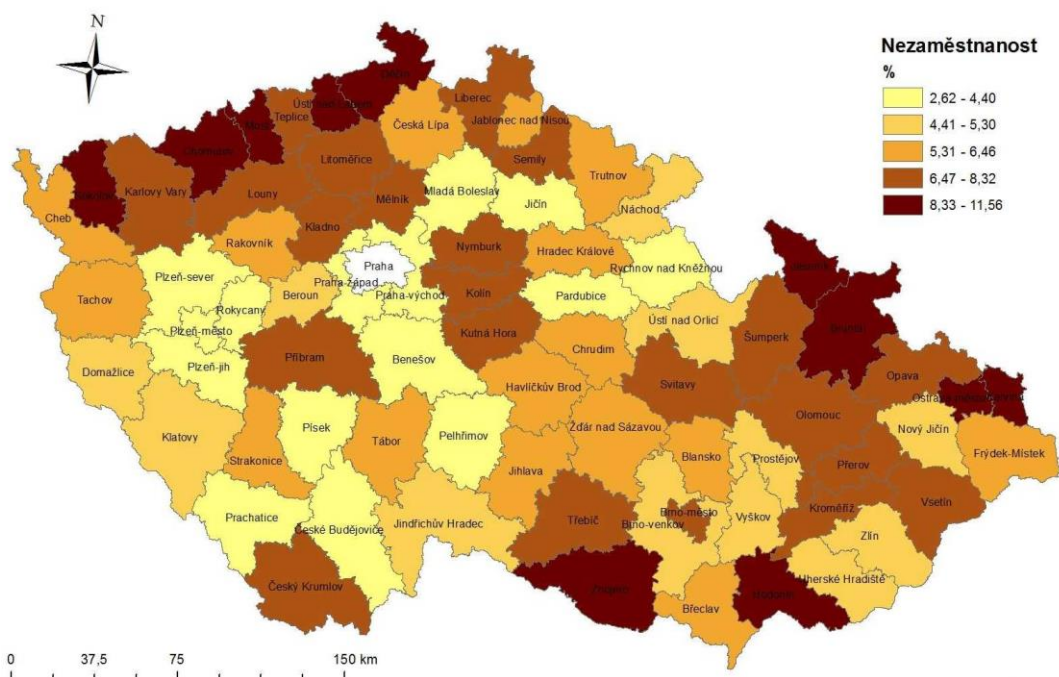
zdroj: Soubor dat a tabulek, vlastní tvorba

6.4. Model z dat roku 2015 s prostorovou korelací

Nejprve je dobré se zamyslet, co by se stalo, kdyby v okrese byla vysoká míra nezaměstnanosti a v sousedním okrese naopak nízká. Pravděpodobně by hodně lidí cestovalo za prací a rozdíl by se postupem času snižoval.

Proto přicházíme s myšlenkou prostorové korelace, protože se domníváme, že míra nezaměstnanosti v našem okrese bude vyšší, pokud našimi sousedy budou okresy s vyšší nezaměstnaností. Jednou z možností, jak toto zkoumat je, že si sestojíme kartogram.

MÍRA NEZAMĚŠTNANOSTI V OKRESECH ČR V ROCE 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

V kartogramu 1 je vidět, jak některé okresy tvoří pomyslné celky s podobnou mírou nezaměstnanosti. Například oblast kolem Plzně s hodnotami kolem 4%, oblast kolem Mostu a Ústí nad Labem je naopak charakteristická vyšší hodnotou stejně jako oblast kolem Bruntálu.

Druhou možností je vytvořit si všechny dvojice sousedních okresů s jejich hodnotami. Výsledné dvojice pro okres Český Krumlov vypadají následně.

Tabulka 16 - Skupina dvojic okresů pro ilustraci sestavení tabulky pro výpočet prostorového korelačního koeficientu

Český Krumlov	7,403904	České Budějovice	4,042228
Český Krumlov	7,403904	Prachatice	4,408312

zdroj: Soubor dat a tabulek, vlastní tvorba

Stejným způsobem pak konstruujeme další okresy. Výsledná tabulka je poté tvořena 378 dvojicemi a my si z nich vypočítáme korelační koeficient. Ten vychází 0,274284.

Zjistili jsme, že v roce 2015 na míru nezaměstnanosti v okrese má značný vliv i okolní okresy. Z tohoto důvodu rozšiřujeme proměnné průměrnou mírou nezaměstnanosti sousedních okresů a dále vytváříme regresní funkci.

Tabulka 17 - Výsledek regresní analýzy pro rok 2015 s prostorovou korelací

N=76	Regression Summary for Dependent Variable: Míra nezaměstnanosti (sheet2015+MN) R=,54742448 R2=,29967357 Adjusted R2=,28048654 F(2,73)=15,619 p<,00000 Std.Error of estimate: 1,6752					
	b*	Std.Err. of b*	b	Std.Err. of b	t(73)	p-value
Intercept			0,618849	1,048766	0,590073	0,556964
Krajské město	0,206446	0,098030	0,016012	0,007603	2,105956	0,038647
Průměr soused	0,498573	0,098030	0,828042	0,162810	5,085942	0,000003

zdroj: Soubor dat a tabulek, vlastní tvorba

Výsledná regresní funkce má tvar

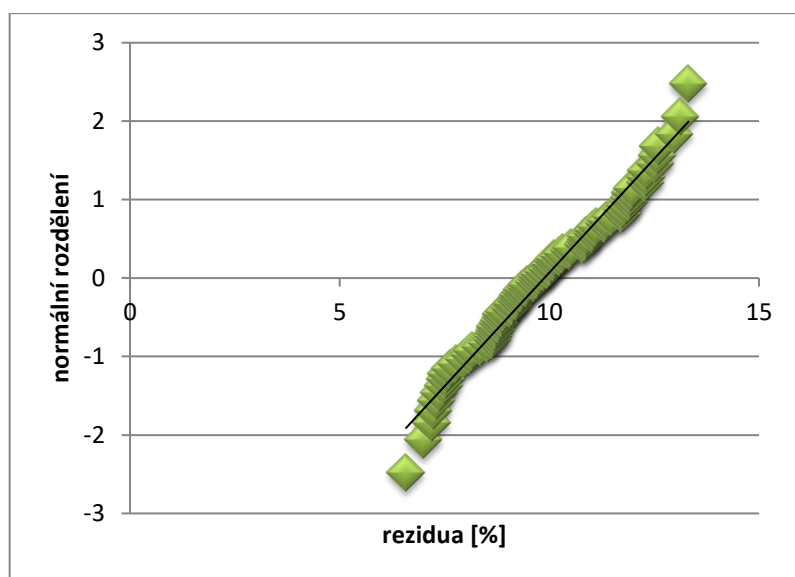
$$MN_x = 0,618849 - 0,016012 * KM_x + 0,828042 * S_x, \quad (12)$$

ve které MN_x je odhadovaná míra nezaměstnanosti okresu x, KM_x je vzdálenost okresu x od krajského města a S_x je průměrná míra nezaměstnanosti okolních okresů.

6.5. Použití modelů pro predikci hodnot roku 2016

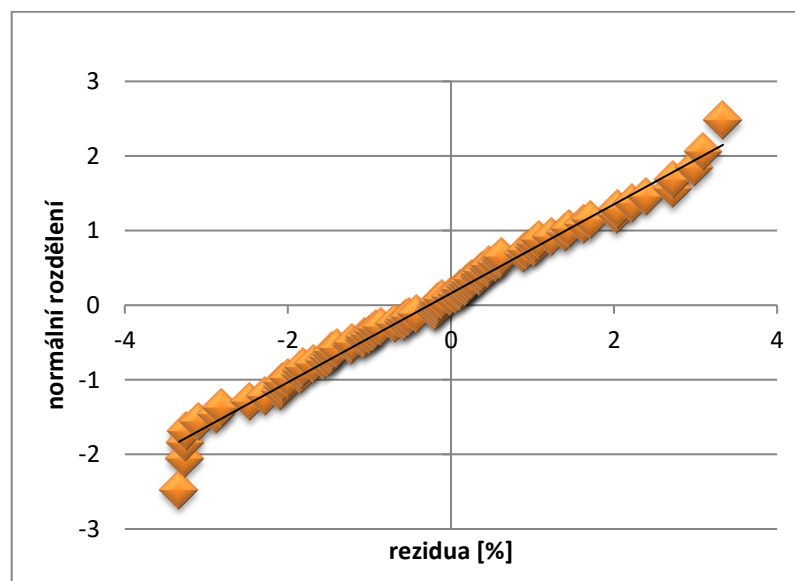
Z kapitol výše jsme si vypočítali tři regresní funkce z různých let. Jde o funkce 9, 10 a 11, u kterých si nyní vyzkoušíme, jak přesně nám dokážou odhadnout míru nezaměstnanosti. Na datech z roku 2016 s nimi vypočteme hodnoty, které porovnáme se skutečností tak, že si sestojíme N-P grafy reziduí. Regresní funkci z kapitoly 6.4 nebereme v úvahu, protože potřebujeme pro výpočet samotnou míru nezaměstnanosti. Takže tento model je pro predikci budoucích dat nepoužitelný.

Graf 31 - N-P graf reziduí 2016 pro model z roku 2005



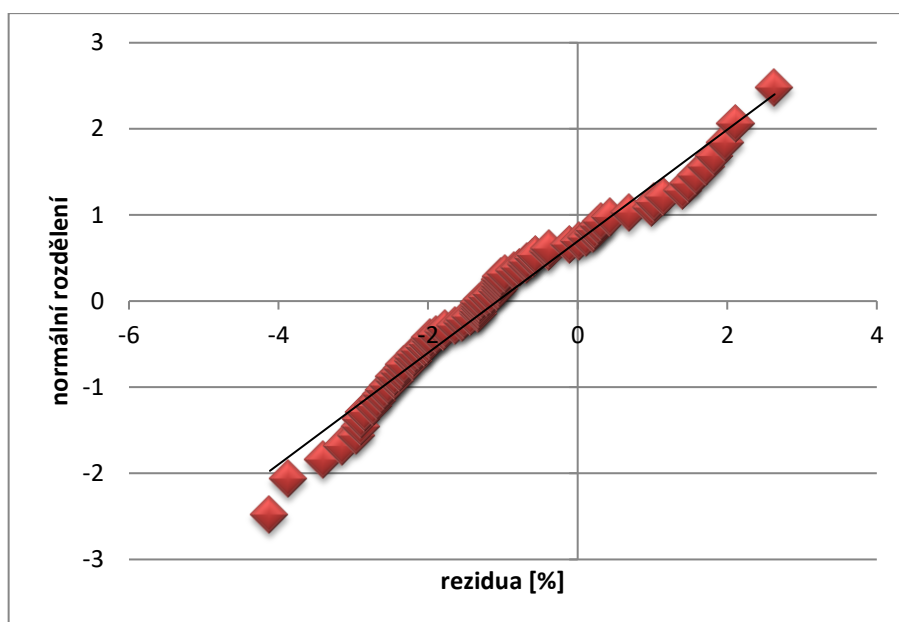
zdroj: Soubor dat a tabulek, vlastní tvorba

Graf 32 - N-P graf reziduí 2016 pro model z roku 2010



zdroj: Soubor dat a tabulek, vlastní tvorba

Graf 33 - N-P graf reziduí 2016 pro model z roku 2015



zdroj: Soubor dat a tabulek, vlastní tvorba

Odhadovali jsme hodnoty pro každý model zvlášť a následně z nich zkonstruovali N-P grafy reziduí. Model z roku 2005 je úplně nepoužitelný, střední hodnota reziduí se pohybuje kolem 10%. U zbylých dvou modelů je rozpětí reziduí podobné. U modelu 2010 je střední hodnota někde kolem nuly, naopak u modelu z roku 2015 je střední hodnota záporná. Z tohoto můžeme vyvodit závěr, že tento model má tendenci odhadovat hodnoty převážně vyšší, než jsou skutečně naměřené hodnoty.

7. Závěr

Vytvořili jsme tři modely a zjistili jsme, že ve všech třech modelech figurují dvě proměnné. Vzdálenost okresních měst od hlavního města Prahy a vybavenosti domácností informačními technologiemi. Kladný koeficient u Prahy nasvědčuje, že míra nezaměstnanosti roste se vzdáleností od Prahy. A záporný u proměnné IT říká, že míra nezaměstnanosti klesá při větší vybavenosti domácností výpočetní technikou.

V roce 2005 se navíc objevily další dvě proměnné. Vzdálenost dálnic a průměrný věk. Při zkoumání koeficientů z regresní funkce jsme došli k závěru, že nižší míra nezaměstnanosti se objevuje v okresech s větším počtem kilometrů dálnic a starším obyvatelstvem.

Pro rok 2010 tyto dvě proměnné zmizely a nahradily je vzdálenost státních hranic a vzdálenost okresních měst od měst krajských. Závěr zní, že okresy vzdálenější od svých krajských měst trpí větší mírou nezaměstnanosti. Zároveň se větší nezaměstnanost nacházela u hranic s ekonomicky silnějšími státy, než se státy slabšími.

V roce 2015 už záleželo pouze, jak jsou okresy vzdálené od Prahy, od krajských měst a na vybavenosti domácností informačními technologiemi. Menší míra se nacházela blíže ku Praze, ke krajským městům a v lépe vybavených okresech.

Při zkoumání prostorové korelace a vytváření následného regresního modelu jsme zjistili, že v roce 2015 míra nezaměstnanosti v okrese do jisté míry závisí na vzdálenosti od krajských měst a na průměru měr nezaměstnaností sousedních okresů.

V praxi pro predikci míry nezaměstnanosti pro rok 2016 obstál pouze model z roku 2010, ale pro přesné predikce také není vhodný, protože tvoří poměrně velké rezidua.

Hlavní poznatky z celé práce jsou tři. Míra nezaměstnanosti je zpravidla menší směrem ke krajským městům a k hlavnímu městu. Lépe jsou na tom okresy s lepší vybaveností domácností informační technikou, ale u tohoto zjištění se dá polemizovat, jestli tomu není právě naopak. Že domácnosti v lépe zaměstnaných okresech mají více příjmů a mohou si proto toto zboží dovolit. A třetí zjištění je takové, že míru nezaměstnanosti daného okresu možná ovlivňují okresy sousední. Abychom toto tvrzení mohli potvrdit, museli bychom se danou problematikou zabývat více.

8. Summary

My bachelor thesis is focusing on creation of an unemployment rate model based on various characteristics regarding all districts in the Czech Republic.

Beginning of this thesis is dedicated to labour market, income and substitution effect which both affect the job offer. Afterwards we uncover various types of unemployment, including detailed description of statistical methods used inside and description of correlation and regress analysis which are important parts for next steps.

Main point of the during model creation is definitely choosing variables. We try to find suitable ones. Correlation analysis and graphs are used for checking the validity of all variables. This way 3 models are created while using linear regress method in Statistica 13.2. Models are made of data considering specific years. We use them for 2016 unemployment rate prediction and compare final results. Second point of this thesis is focused on area correlation.

We discovered that the 2005 year model is totally useless for predication. The 2015 model seems to be fine, but results are on higher average than true values. The 2010 year model is the best out of those three, but isn't suitable for accurate predictions. After analysis of all models several findings have been found. Variables „distance from Prague“ and „household IT equipment“ are present in all models. We can say that unemployment rate is generally lower in districts closer to Prague and in households with lower values of IT equipment. We found a correlation of unemployment rate in neighborhood, but for it to be taken as a fact we should do deeper analysis.

9. Seznam literatury a použitých zdrojů

Seznam literatury

- **HOLMAN**, Robert. *Ekonomie*. 4., aktualiz. vyd. Praha: C.H. Beck, 2005. Beckovy ekonomické učebnice. ISBN 80-7179-891-6.
- **DORNBUSCH**, Rudiger. *Makroekonomie*. Praha: SPN, 1994. ISBN 80-04-25556-6.
- **SAMUELSON**, Paul Anthony a William D. **NORDHAUS**. *Ekonomie: 19. vydání*. Praha: NS Svoboda, 2013. ISBN 978-80-205-0629-0.
- **HOWELL**, David R., ed. *Fighting unemployment: the limits of free market orthodoxy*. New York: Oxford University Press, 2005.
- **HINDLS**, Richard, Stanislava **HRONOVÁ** a Jan **SEGER**. *Statistika pro ekonomy*. 2. vyd. Praha: Professional Publishing, 2002. ISBN 80-86419-30-4.
- **HINDLS**, Richard, Ilja **NOVÁK** a Jara **KAŇOKOVÁ**. *Metody statistické analýzy pro ekonomy*. Praha: Management Press, 1997. ISBN 80-85943-44-1.
- **ANDĚL**, J. *Statistické metody*. Praha: Matfyzpress, 1998.
- **NETOLICKÁ**, Veronika. *Testy normality*. Přírodovědecká fakulta Univerzity Palackého v Olomouci : Katedra matematické analýzy a aplikací matematiky, 2008.

Seznam internetových zdrojů

- **KASAL**, Pavel a Marie **HLADÍKOVÁ**. Koutek pro statistiky amatéry. *Pelikán : akademický bulletin 2. LF UK* [online]. 1995, roč. 3, no. 3, s. -, Dostupné z <http://stary.lf2.cuni.cz/projekty/pelikan/peli0395/statis1.htm>.
- **HOLČÍK**, Jiří, **KOMENDA**, Martin (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vydání. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9.
- *Míry nezaměstnanosti ve státech OECD 2015*. Dostupné z <https://data.oecd.org/unemp/unemployment-rate.htm>
- *HDP na obyvatele v ČR a sousedních státech*. Dostupné z <https://data.oecd.org/gdp/gross-domestic-product-gdp.htm>
- *Soubor dat a tabulek*. Dostupné z <https://www.czso.cz>

Seznam grafů, tabulek a kartogramu

Seznam grafů

Graf 1 - Trh práce.....	9
Graf 2 - Mezní užítky.....	10
Graf 3 - Nabídka práce.....	12
Graf 4 - Dobrovolná nezaměstnanost.....	14
Graf 5 - Nedobrovolná nezaměstnanost.....	15
Graf 6 - Phillipsova křivka.....	17
Graf 7 - Korelační koeficient blížící se 1.....	18
Graf 8 - Korelační koeficient blížící se -1.....	19
Graf 9 - Typy regresních funkcí.....	21
Graf 10 - Udělené známky studentům.....	22
Graf 11- Míra nezaměstnanosti ve státech OECD v roce 2015.....	23
Graf 12 - Q-Q graf měr nezaměstnaností států OECD.....	24
Graf 13 - Q-Q graf měr nezaměstnaností států OECD bez ESP,GRC a ZAF.....	24
Graf 14 - HDP na obyvatele [\$/rok].....	27
Graf 15 - Míra nezaměstnanosti na vzdálenosti dálnic 2015.....	29
Graf 16 - Míra nezaměstnanosti na vzdálenosti silnic 1. tříd v roce 2015.....	30
Graf 17 - Míra nezaměstnanosti na vzdálenosti silnic 2. tříd v roce 2015.....	30
Graf 18 - Míra nezaměstnanosti na vzdálenosti okresního od krajského města.....	31
Graf 19 - Míra nezaměstnanosti na vzdálenosti okresního města od Prahy.....	31
Graf 20 - Míra nezaměstnanosti na vzdálenosti státní hranice.....	32
Graf 21 - Míra nezaměstnanosti na poměru cizinců v roce 2015.....	33
Graf 22 - Míra nezaměstnanosti na vybavenosti domácností IT v roce 2015.....	33
Graf 23 - Míra nezaměstnanosti na průměrném věku v roce 2015.....	34
Graf 24 - Míra nezaměstnanosti na průměrném věku v roce 2015 bez okresů Praha-západ a Praha-východ.....	35
Graf 25 - Míra nezaměstnanosti na průměrné hrubé mzdě v roce 2015.....	35
Graf 26 - Míra nezaměstnanosti na vzdálenosti železnice v roce 2015.....	36
Graf 27 - Variabilita IT a průměrné hrubé mzdy v roce 2015.....	40
Graf 28 - N-P graf reziduí modelu z roku 2015.....	46
Graf 29 - N-P graf reziduí modelu z roku 2005.....	47
Graf 30 - N-P graf reziduí modelu z roku 2010.....	49
Graf 31 - N-P graf reziduí 2016 pro model z roku 2005.....	52
Graf 32 - N-P graf reziduí 2016 pro model z roku 2010.....	52
Graf 33 - N-P graf reziduí 2016 pro model z roku 2015.....	53

Seznam tabulek

Tabulka 1 - Míry nezaměstnanosti podle věku v ČR 2015	28
Tabulka 2 - Korelační koeficienty mezi mírou nezaměstnanosti a proměnnými	37
Tabulka 3 - Korelační koeficienty mezi jednotlivými proměnnými 1. část	38
Tabulka 4 - Korelační koeficienty mezi jednotlivými proměnnými 2. část	39
Tabulka 5 - Všechny proměnné pro regresní analýzu dat z roku 2015	40
Tabulka 6 - Proměnné pro regresní analýzu - odebrání první proměnné	41
Tabulka 7 - Proměnné pro regresní analýzu - odebrání druhé proměnné	42
Tabulka 8 - Proměnné pro regresní analýzu - odebrání třetí proměnné	42
Tabulka 9 - Proměnné pro regresní analýzu - odebrání čtvrté proměnné	43
Tabulka 10 - Proměnné pro regresní analýzu - odebrání páté proměnné	43
Tabulka 11 - Proměnné pro regresní analýzu - odebrání šesté proměnné	44
Tabulka 12 - Proměnné pro regresní analýzu - odebrání sedmé proměnné	44
Tabulka 13 - Výsledek regresní analýzy z roku 2015	45
Tabulka 14 - Výsledek regresní analýzy z roku 2005	46
Tabulka 15 - Výsledek regresní analýzy z roku 2010	48
Tabulka 16 - Skupina dvojic okresů pro ilustraci sestavení tabulky pro výpočet prostorového korelačního koeficientu	50
Tabulka 17 - Výsledek regresní analýzy pro rok 2015 s prostorovou korelací	51

Seznam kartogramů

Kartogram 1 - Míra nezaměstnanosti v okresech ČR v roce 2015	50
--	----