

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA STROJNÍ



Diplomová PRÁCE

Závislost HDP na ekonomických faktorech

Praha, 2015

Autor: Jakub Šangala

Prohlášení

Prohlašuji, že jsem svou diplomovou prací vypracoval samostatně a použil jsem pouze podklady, uvedené v příloženém seznamu.

V Praze

Podpis

Poděkování

Chtěl bych tímto poděkovat všem, kteří mi při vytváření této práce pomáhali. Především děkuji svému vedoucímu práce doc. Ing. Janu Kožíškovi CSc.

Abstrakt

Diplomová práce se zabývá zkoumáním závislostí mezi velkým počtem veličin. Čtenářům napomáhá vytvořit postup výběru a následné eliminace veličin, za účelem sestavení optimálního regresního modelu. K tomu poslouží aplikace, která byla v rámci této diplomové práce vytvořena. Funkčnost této aplikace byla ověřena skrze analýzu závislosti HDP na 7 mnou zvolených regresorech.

Abstract

This master thesis is examining the dependency of a numerous figures. It helps the users of this thesis to develop a procedure for elimination those variables, which are not relevant to the development of an optimal regression model. For this procedure an application was developed. The functionality of this application was tested via analyse of the dependency between GDP and 7 independent variables.

Obsah

1	Úvod a cíle práce.....	9
1.1	Závislost jevů.....	9
1.2	Pojmy regrese a korelace	11
1.3	Regresní a korelační analýza	11
1.4	Jednoduchá lineární regrese.....	15
1.5	Vícenásobná lineární regresní a korelační analýza.....	16
2	Identifikace a charakteristika veličin.....	24
2.1	Způsoby výpočtu HDP.....	24
2.1.1	Výdajová metoda.....	24
2.1.2	Důchodová metoda.....	25
2.1.3	Produkční metoda.....	25
2.2	Výběr a charakteristika regresorů	26
2.2.1	Počet obyvatel	26
2.2.2	Počet nezaměstnaných a počet zaměstnaných.....	27
2.2.3	Roční míra inflace	27
2.2.4	Roční spotřeba elektrické energie.....	27
2.2.5	Počet malých a středních podniků.....	28
2.2.6	Produktivita práce.....	28
3	Volba regresního modelu.....	30
3.1	Postup volby regresního modelu.....	31
3.1.1	Metoda linearity.....	31
3.1.2	Multikolinearita	32
3.1.3	Grubbsův test.....	35
3.1.4	Analýza reziduí.....	37
4	Analýza závislost HDP na zvolených regresorech	38
4.1	Příručka použití aplikace	38

4.1.1	Zápis dat	38
4.1.2	Test linearity	40
4.1.3	Test multikolinearity	41
4.1.4	Grubbsův test	42
4.1.5	Analýza reziduí	43
4.1.6	Regresní a korelační analýza	43
4.2	Analýza HDP	44
4.2.1	Zápis dat	44
4.2.2	Metoda linearity	47
4.2.3	Metoda multikolinearity	48
4.2.4	Grubbsův test	52
4.2.5	Analýza reziduí	55
4.2.6	Regresní a korelační úkon	56
5	Vyhodnocení RKA	59
6	Závěr	60
	Publikace	61
	Přílohy	61
	Seznam obrázků	61-62
	Seznam tabulek	63

Seznam použitých zkratk

EU	European union – Evropská unie
RKA	Regresní a korelační analýza
HDP	Hrubý domácí produkt
C	Household consumption - spotřeba domácností
I	Business investment spending, purchases of final products - investiční výdaje firem, nákup finální produkce
G	State spending on the purchase of final products - výdaje státu na nákup finální produkce
NX	Net exports - čistý export
PB	Personal expenses Companies - osobní náklady firem
Tx	Indirect taxes - nepřímé daně
Pr	Corporate profits - zisky firem
Am	Amortisation - odpisy
In	Interest Companies - úroky firem
CO	Net mixed income, bonuses owners - čistý smíšený důchod, odměny vlastníkům
CG	Cash grants- peněžní dotace
PH	Performances - výkony

MSP	Malé a střední podniky
OECD	Organisation for Economic Co-operation and Development - Organizace pro hospodářskou spolupráci a rozvoj
HDP	Hrubý domácí produkt
PZ	Počet zaměstnaných
ΦΡΟΗΤ	Průměrný počet odpracovaných hodin týdně
PS	Počet státních svátků, dní bez práce
TSS	Total sum of squares - celkový součet čtverců
MSS	Model sum of squares – modelová suma čtverců
RSS	Residual sum of squares - reziduální součet čtverců
PO	Počet obyvatel
PN	Počet zaměstnaných
MI	Míra inflace
SEE	Spotřeba elektrické energie
PMSP	Počet malých a středních podniků
PP	Produktivita práce

1 Úvod a cíle práce

Hlavním cílem mé diplomové práce je vytvoření excelovské aplikace, která usnadní zkoumání závislosti mezi velkým počtem veličin. Aplikace pomůže jejím uživatelům najít ideální počet veličin pro jejich dané empirické šetření a následně, pro tyto vybrané veličiny provede regresní a korelační úkon. Funkčnost aplikace a její použití ukáží na příkladu zkoumání závislosti hrubého domácího produktu zemí EU na mých identifikovaných ekonomických faktorech. Výsledky této analýzy shrnu v 5. kapitole.

1.1 Závislost jevů

S pojmy jako jsou závislost, souvislost, vztah příčina a následek se nesetkáváme pouze při řešení teoretických otázek některého z vědního oboru. Přicházíme s nimi do styku běžně i v reálném životě. Poznání obecných principů mezi sledovanými veličinami, slouží k analýze a následnému rozvoji většiny disciplín. Samotné výsledky analýzy nesmí vycházet pouze z teoretických poznatků, je zde důležitá schopnost zobecňovat empirické informace o vlivy působení prostředí.

Pojem funkční závislost veličiny Y na veličině X představuje funkční předpis $y = f(x)$, který jednoznačně přiřazuje hodnotám závisle proměnné Y hodnoty nezávisle proměnné X . Jde zde o jednoznačný zápis. Pro každou hodnotu X nám vyjde hodnota Y . Nepřipouštíme možnost, že by při určitých podmínkách tento zápis platil, a při jiných neplatil. Nebo dokonce, že by pro určité podmínky vyhovoval lépe nebo hůře. Je zde plně vyloučen vliv vedlejších činitelů. Proto se s tímto postupem v praxi příliš často nesetkáme.

V pravděpodobnostním pojetí se dva jevy považují za závislé, jestliže výskyt jednoho jevu podmiňuje pravděpodobnostní výskyt jevu druhého. Zcela stejně si v pravděpodobnosti vykládáme závislost dvou náhodných proměnných: Dvě náhodné veličiny jsou na sobě závislé v případě, že zákon rozdělení jedné veličiny závisí na tom, jakých hodnot bude nabývat druhá veličina. Prostě a jednoduše řečeno, každá hodnota jedné proměnné má nestejnou pravděpodobnost výskytu hodnoty druhé veličiny.

Pravděpodobnostní pojetí se velmi podobá pojetí statistickému. Zde je závislost jedné veličiny na druhé podmíněna určitým systematickým směřováním. Například růst hodnot jedné závisle proměnné Y , povede k poklesu hodnot druhé nezávislé X . Zde mně napadá jeden příklad statistické závislosti. Růst cen elektrické energie $Kč/kWh$ bude mít jistě za následek průměrný pokles odběrů elektrické energie domácností v kWh . Zvýšení cen povede u většiny domácností ke spoření s energií. Poskytovatelé elektrické energie, se jistě nespokojí pouze s teoretickým vyjádřením této závislosti. Bude důležité zobecnit empirické údaje a podle nich vyjádřit ekonomický dopad, který zdražení energie přineslo. Pokud zvýšení cen i při snížené spotřebě povede k větším ziskům, můžeme mluvit o správném kroku. Na příkladu jsem chtěl demonstrovat, že ve většině příkladu nevystačíme pouze s teoretickými znalostmi. Každý případ je sám o sobě unikátní a záleží na našem vyhodnocení výsledků. Tohle obecné systematické směřování hodnot jedné závislé proměnné při změnách hodnot druhé proměnné, však zpravidla nemusí znamenat příčinný vztah mezi těmito veličinami.

V různých oblastech zkoumání používáme statistické metody k vyjádření závislosti jednotlivých veličin pomocí matematických modelů. Matematický model sestavujeme vždy pomocí empirických údajů, které jsme získali buď z provedených experimentů, nebo z pozorování a zapisování určitých jevů reálného světa. Informace jsou v dnešní době ceněnou komoditou, důležitou povinností každého statistika je zvolit vhodnou metodu jejich sběru. Informace musí mít co největší vypovídací schopnosti, respektovat zvláštnosti daného problému a zároveň musí být časově a nákladově přijatelné. Není tedy pro statistika vůbec jednoduché vybrat správnou metodu a následně se rozhodnout v jakém rozsahu ji praktikovat.

V přírodních vědách se hlavně setkáváme s organizovanými systémy, s veličinami jejichž příčinné vazby dobře známe. Víme, jaké množství daného prvku vyvolá v roztoku danou reakci. Kolika % uhlíková ocel má danou tvrdost. Tyto jevy jsou řízeny přírodou. Proto proměnné jsou zatíženy výhradně chybami pozorování a měření. Matematické modely vychází především z odvozených křivek a grafů, které jsme odvodili pomocí experimentů. Jde hlavně o vyrovnání experimentálních údajů vhodnou křivkou (přímkou, parabolou exponenciálou apod.).

Pro technické a společenské vědy je typický velký počet vzájemně závislých veličin. Je tedy složité odhalit a vybrat vhodné veličiny. Snažíme se vždy vybrat veličiny, které mají co největší vliv na sledovaný případ, jsou časově i nákladově přijatelné a naplňují požadavek multikolinearity. To bude úkolem 3 kapitoly naší práce.

1.2 Pojmy regrese a korelace

Pojem regrese pochází už z konce 19. Století, kdy ho poprvé použil anglický vědec Francis Galton. Galton byl renesančním člověkem, činným v mnoha vědních disciplínách: psychologii, antropologii, geografii, statistice a dalších. Zabýval se také otázkami dědičnosti. Pozoroval vztah mezi výškou otců a jejich dětí, své výsledky publikoval v časopise „Antropolog“, zde také prvně použil terminologii „regrese“. Galton svým rozsáhlým pozorováním, došel k rovnici, ze které vyplývalo, že vysocí otcové mají vysoké syny, tedy vyšší než je průměr. Ale zároveň v průměru nižší než jsou sami otcové. A obdobně, malí otcové mají malé syny, ale v průměru vyšší než jsou oni sami. Tento trend následující generace směrem k průměru nazval Galton regresí. Proč to všechno uvádím? Abychom si uvědomili, že současné pojetí regresní problematiky má jen málo společného s tím Galtonovým, ale myšlenka použití empirických údajů k znázornění nějaké tendence, neboli trendu, zůstává stejná.

V současnosti se pod pojmem regrese, potažmo regresní analýza, rozumí určité uspořádané změny jedné veličiny, při změnách jiných veličin a následné zobrazení průběhů těchto změn pomocí matematických funkcí. Regresní analýza je jedna z nejdůležitějších metod matematické statistiky a samotně nebo ve spojení s jinými metodami patří k nejužívanějším metodám prakticky v každé oblasti vědy. Často ji také používají podniky, nehledě na odvětví a velikosti. Je to nejlepší metoda prognózování hodnot vystupující veličiny (závisle proměnné), při znalosti nebo odhadu vstupujících veličin (nezávisle proměnných). V naší diplomové práci budeme označovat jako závisle proměnnou y , a nezávisle proměnné $x_1, x_2 \dots x_k$.

S pojmem regrese úzce souvisí pojem korelace. Pochází už z 16. století, českými synonymy jsou slova souvztažnost, vzájemný vztah, souvislost atd. Obecně lze říct, že korelace znamená vzájemnou souvztažnost veličin, nebo procesů. V případě, že se jedna veličina mění, mění se i ta druhá a naopak.

1.3 Regresní a korelační analýza

Jak již jsem psal v úvodní kapitole, zkoumání závislostí mezi jednotlivými jevy a veličinami není výhra pouze vědních disciplín. Ale i vy si určitě vybavíte situace v běžném

životě, kdy jste uvažovali nad tím, jak jedna proměnná ovlivní tu druhou a naopak. Jak se projeví počet účastníků naší domácnosti na spotřebě energie, počet pracovníků naší firmy na celkovém hospodářském výsledku atd... Nejjednodušším postupem je podívat se na hodnoty veličin z let minulých a snažit se pochopit vztahy mezi nimi, a následně odvodit určitý vztah. Vezmeme si příklad se spotřebou elektrické energie. Když nám přibude člověk v domácnosti, určitě se spotřeba elektrické energie zvýší. Ale jak mnoho? Snažíme se tedy podle určitých vzorců kvantitativně popsat závislost mezi sledovanými jevy a veličinami. Regresní a korelační analýza nám pomáhá tuto závislost kvantitativně popsat podle určitých rovnic (přímky, paraboly, exponenciály, mocninné funkce ...) a následně určit těsnost regresního odhadu.

Regresním úkolem se tedy rozumí kvantitativní popsání průběhu sledovaného vztahu mezi proměnnými. Prvním krokem regresního úkolu je odhad typu křivky (přímky, paraboly, exponenciály, mocninné funkce ...) dle průběhu sledovaného vztahu. Pokud si nejsme jistí správnosti volby dané křivky, provedme výpočet pro více křivek a dle těsnosti korelační závislosti vyberme tu optimální, s nejvyšší hodnotou korelačního koeficientu. Dalším krokem je výpočet korelačních ukazatelů (neboli koeficientů). To provádíme podle metody nejmenších čtverců.

Metoda nejmenších čtverců

Vychází z požadavku minimalizace reziduálního součtu čtverců. Zjednodušeně řečeno, hledáme, nulový průměrný součet všech čtverců odchylek empirických hodnot od hodnot vyrovnaných. Empirické hodnoty y_i jsou skutečné hodnoty, dříve naměřené a zaznamenané. Vyrovnané hodnoty \hat{y}_i jsou vypočtené hodnoty na základě odvozených regresních funkcí. Tento průměrný součet čtverců odchylek nese ve statistice označení reziduální rozptyl.

$$s_{y,x}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min$$

Posledním krokem regresního úkolu je stanovení výsledného tvaru funkcionální regrese. Kde $b_1 \dots b_k$ jsou vypočtené regresní koeficienty a $F_i(x_i)$ jsou funkce nezávisle proměnné x , které již neobsahují žádný další parametr.

$$\hat{y}_i = b_0 + b_1 F_1(x_1) + b_2 F_2(x_2) + b_3 F_3(x_3) \dots b_k F_k(x_k)$$

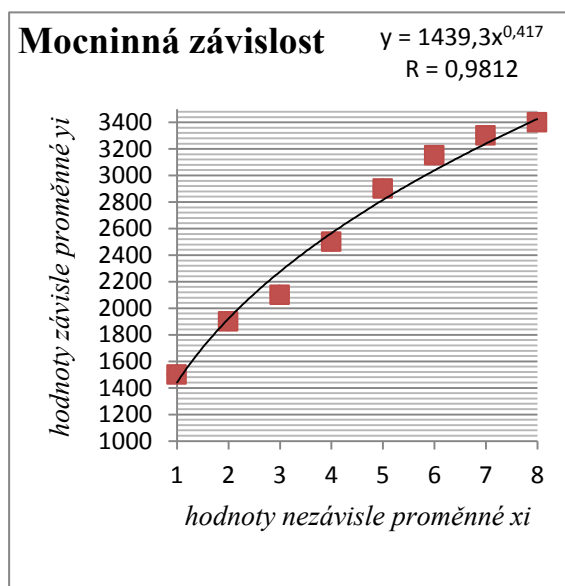
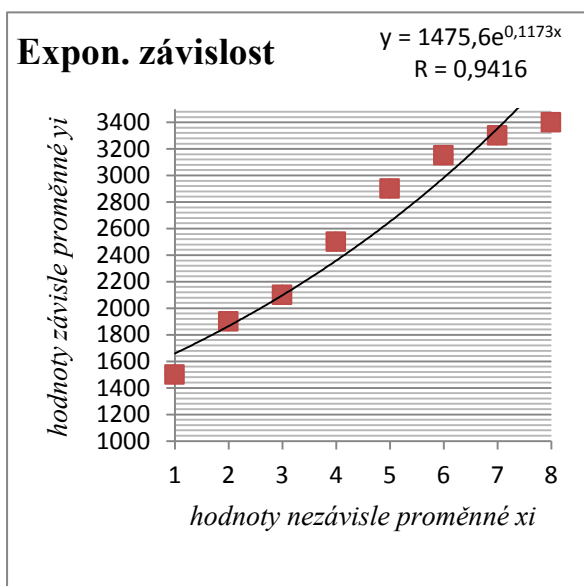
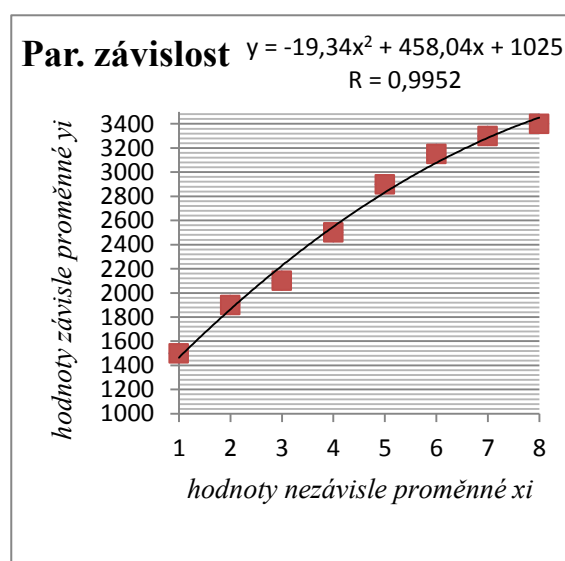
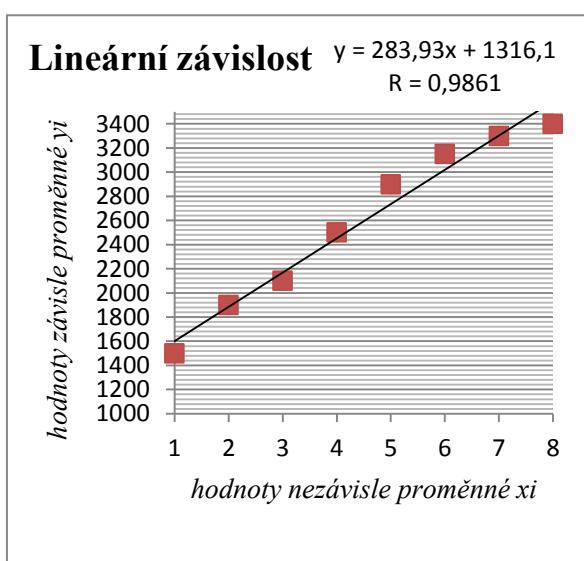
Korelační úkol se snaží určit těsnost daného regresního odhadu, tedy jeho spolehlivost. Termín korelační koeficient se používá pro znázornění spolehlivosti regresního odhadu při vyrovnání přímkou. V ostatních případech používáme označení korelační index (vyrovnání parabolou, exponenciálou, mocninnou funkcí ...). Korelační koeficient nabývá hodnot od -1 až do 1. Čím vyšší hodnota korelačního koeficientu, tím větší je spolehlivost daného regresního odhadu. Sledovaný vztah mezi dvěma veličinami může být buď kladný $y = a+bx$, zvýšení hodnoty nezávislé proměnné x se přímo projeví zvýšením hodnoty závislé proměnné y . Nebo záporný $y = a-bx$, zvýšení hodnoty nezávislé proměnné x povede ke snížení hodnoty závislé proměnné y . Hodnota korelačního koeficientu -1 značí zcela nepřímou korelační závislost. Hodnota $+1$ značí zcela přímou korelační závislost. A hodnota 0 značí nekorelovanost. Stále mluvíme o korelačním koeficientu, jde tedy o vyrovnání přímkou. Nulová hodnota značí, že mezi závisle proměnnou y a nezávisle proměnnou x není zjištěna žádná přímá závislost. Obecně nemůžeme říct, že mezi sledovanými veličinami není žádná statistická závislost. Uvědomme si, že pokud použijeme jinou rovnici křivky pro vyrovnání, mohou nám vyjít úplně odlišné hodnoty korelačního indexu. I zde platí pravidlo, čím vyšší je hodnota korelačního indexu, tím je vyšší spolehlivost regresního odhadu. Hodnoty se mohou pohybovat od 0 do 1, vzorec si odvodíme v 3. kapitole. Uvedme si pár rovnic, které můžeme použít pro vyrovnání.

1. Rovnice přímky: $\hat{y}_i = b_0 + b_1x_i$
2. Rovnice paraboly: $\hat{y}_i = b_0 + b_1x_i + b_2x_i^2$
3. Rovnice hyperboly: $\hat{y}_i = b_0 + \frac{b_1}{x_i}$
4. Rovnice mocninné funkce: $\hat{y}_i = b_0 x_i^{b_1} \rightarrow \log b_0 + b_1 \log x_i$
5. Rovnice exponenciály: $\hat{y}_i = b_0 b_1^{x_i} \rightarrow \log b_0 + x_i \log b_1$

Na příkladu si ukážeme průběhy jednotlivých funkcí, a jak se výběr křivky projeví na spolehlivosti regresního odhadu. Máme tabulku průměrné roční spotřeby elektrické energie v závislosti na počtu členů domácnosti, *tabulka 1.3.1*. Závisle proměnnou je zde roční spotřeba elektrické energie v kilowatthodině, nezávisle proměnnou zase počet členů domácnosti. V našem případě nejspolehlivějším odhadem regresní funkce bude parabolická přímka, protože má nejvyšší korelační index. Hodnota $R=0,9952$ zaručuje téměř stoprocentní spolehlivost regresního odhadu.

xi [počet členů]	yi [roč. spotřeba Kwh]
1	1500
2	1900
3	2100
4	2500
5	2900
6	3150
7	3300
8	3400

Tabulka 1.3.1 Vliv volby typu regresní přímky



1.4 Jednoduchá lineární regrese

Při jednoduché lineární regresi vyrovnáváme empirické hodnoty přímkou. Jedná se o nejčastěji využívanou matematicko-statistickou metodu používanou k analýze závislosti průběhu dvou proměnných. Jednoduchá lineární regrese znázorňuje aproximaci hodnot proměnných (závisle proměnné y a nezávisle proměnné x) přímkou metodou nejmenších čtverců. Jelikož rovnice přímky je dána vztahem $\hat{y}_i = b_0 + b_1 x_i$, snažíme se najít pomocí této metody optimální hodnoty koeficientů b_0 a b_1 . Tím dosáhneme nejspolehlivějšího regresního odhadu. Zjednodušeně řečeno získáme rovnici přímky, která zaručuje nulový součet všech čtverců odchylek empirických hodnot a hodnot vyrovnaných, s nejvyšší hodnotou korelačního koeficientu. Jak přesný a spolehlivý tento odhad je, nám pomůže určit korelační úkol.

Stanovení koeficientů b_0 a b_1 , provádíme pomocí metody nejmenších čtverců.

$$F(b_0; b_1) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = \min$$

Jelikož hledáme optimální parametry b_0 a b_1 , budeme hledat extrém (minimum) funkce dvou proměnných. Abychom našli minimum součtu čtverců, provedeme parciální derivaci reziduálního rozptylu funkce $F(b_0; b_1)$ podle parametrů b_0 a b_1 a dáme je do rovnosti s nulou.

$$\frac{\partial F(b_0; b_1)}{\partial b_0} = \frac{2}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i) (-1) = 0$$

$$\frac{\partial F(b_0; b_1)}{\partial b_1} = \frac{2}{n} \sum_{i=1}^n (y_i - b_0 - b_1 x_i) (-x_i) = 0$$

Tyto vztahy upravíme a obdržíme rovnice:

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

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

Máme soustavu dvou rovnic o dvou neznámých. Dosazovací nebo sčítací metodou si odvodíme koeficienty b_0 a b_1 (tzv. regresní koeficient):

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

Regresní koeficient b_1 udává, jak se projeví jednotková změna nezávisle proměnné x na výsledné hodnotě závisle proměnné y .

$$b_0 = \frac{\sum_{i=1}^n y_i}{n} - b_1 \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - b_1 \bar{x}$$

Spolehlivost regresního odhadu postihuje korelační koeficient. Vzorec pro jeho výpočet odvodil počátkem 20. století anglický matematik a filosof Karl Pearson. Vycházel z takzvané kovariance. Tou se rozumí míra vzájemné vazby mezi dvěma proměnnými (závisle proměnnou y a nezávisle proměnnou x). Kovarianci vyjádřil jako průměrný součin odchylek korelovaných veličin od jejich průměrů.

$$cov(x; y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Korelační koeficient obdržíme, když podělíme kovarianci součinem směrodatných odchylek obou proměnných (s_x ; s_y):

$$r_{yx} = \frac{cov(x; y)}{s_y s_x} = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\{[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2][n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]\}}$$

Ze vzorce je patrné, že při nulové hodnotě kovariance, bude i korelační koeficient nulový. Při nulové hodnotě kovariance tedy můžeme prohlásit obě proměnné za nekorelované (nebo také nezávislé).

1.5 Vícenásobná lineární regresní a korelační analýza

V minulé kapitole jsme si ukázali, jak matematicky vyjádřit průběh vztahu mezi dvěma proměnnými. Máme zde pouze jednu vstupující proměnnou x , která působí na výslednou hodnotu vystupující proměnné y . Pro technické a společenské zkoumání je však

častější větší počet vzájemně působících veličin. Musíme se tedy rozhodnout, jaké faktory jsou podstatné. Snažíme se najít činitele, které významně působí na naši závisle proměnnou y .

V diplomové práci se budu rozhodovat mezi více vstupujícími veličinami. Mým úkolem bude vybrat pro regresní a korelační analýzu jen ty činitele, které jsou statisticky významné a zároveň nejsou ve velké lineární závislosti samy mezi sebou.

Úkolem vícenásobné regresní a korelační analýzy tedy je vybrání podstatných činitelů, které statisticky významně působí na výslednou hodnotu závisle proměnné y , kvantitativní popsání průběhu sledovaného vztahu mezi proměnnými a následné určení spolehlivosti daného regresního odhadu. Jelikož zkoumáme, jak se projeví působení dvou a více vstupujících proměnných x_i na výsledné hodnotě vystupující proměnné y , mluvíme o vícenásobné regresní a korelační analýze. Analogicky dvojnásobná regresní a korelační analýze se zabývá vztahem jedné závisle proměnné y a dvou nezávislých proměnných x_1 a x_2 , trojná vztahem jedné závisle proměnné y a tří nezávislých proměnných x_1, x_2, x_3 atd...

My se budeme zabývat lineární závislostí. Jde o nejčastější typ vyrovnání, navíc i některé nelineární vztahy se dají transformovat na lineární. Regresní funkci pro vícenásobnou analýzu napíšeme ve tvaru:

$$\hat{y}_i = a_{y.123\dots k} + b_{y1.23\dots k}x_1 + b_{y2.13\dots k}x_2 + \dots + b_{yk.12\dots(k-1)}x_k$$

$a_{y.123\dots k}$... absolutní parametr regresní funkce, doplňující člen regresní rovnice, není přímo ve spojitosti s nezávislými proměnnými

$b_{y1.23\dots k}$... dílčí regresní koeficient – vyjadřuje, jak se průměrně změní závisle proměnná y při jednotkové změně nezávisle proměnné x_1 , za předpokladu pozastavení vlivu zbylých proměnných (x_2, x_3, \dots, x_k) na dané stálé úrovni

Pro výpočet dílčích regresních koeficientů $b_{yi.23\dots k}$ použijeme směrodatné proměnné $u_1, u_2, u_3, \dots, u_k$. Směrodatné proměnné:

$$u_y = \frac{y_i - \bar{y}}{s_y}$$

$$u_i = \frac{x_i - \bar{x}}{s_i}$$

\bar{y} ... aritmetický průměr empirických hodnot závisle proměnné y , $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

s_y ... směrodatná odchylka empirických hodnot závisle proměnné y , $s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}$

\bar{x} ... aritmetický průměr empirických hodnot nezávisle proměnné x_i , $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

s_x ... směrodatná odchylka empirických hodnot nezávisle proměnné x_i , $s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$

n ... počet sledování

Jelikož budeme mít zpravidla více hodnot pozorování, uvažujeme, že směrodatné proměnné mají povahu náhodných veličin podléhajících normálnímu rozdělení. Hustota pravděpodobnosti má tedy v tomto případě tvar pro $x \in (-\infty, \infty)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

x ... hodnota náhodné veličiny

μ ... střední hodnota náhodné veličiny normálního rozdělení

σ ... směrodatná odchylka náhodné veličiny normálního rozdělení

S tímto tvarem se však v praxi moc neseťkáváme. Práci nám ulehčí normované normální rozdělení. Transformované normální rozdělení se střední hodnotou $E(U)=0$ a rozptylem $D(U)=1$. Rozdělení je definováno pro teoreticky odvozenou veličinu u , která vznikne transformací původní náhodné veličiny x tak, že od ní odečteme střední hodnotu základního souboru a rozdíl se vydělí směrodatnou odchylkou základního souboru. Zde označení u , definuje teoreticky odvozenou veličinu, nepleťme si to s označením směrodatných proměnných! Hustota pravděpodobnosti má zde tvar pro $x \in (-\infty, \infty)$:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

$$u = \frac{x - \mu}{\sigma}$$

Pro výpočet regresních koeficientů použijeme vlastnost náhodných veličin vyplývající z tohoto normovaného normálního rozdělení, $E(U)=0$ a $D(U)=1$. A to, že aritmetický průměr směrodatných odchylek je rovný nule a jejich rozptyl a směrodatná odchylka rovné 1 $\bar{u} = 0, s^2(u) = 1, s(u) = 1$.

Směrodatné proměnné zavedeme do regresní rovnice:

$$\hat{u}_y = \alpha_{y.123\dots k} + \beta_{y1.23\dots k}u_1 + \beta_{y2.13\dots k}u_2 \dots \beta_{yk.12\dots(k-1)}u_k$$

Znovu použijeme metodu nejmenších čtverců.

$$F(\alpha_{y.123\dots k}; \beta_{y1.23\dots k}; \beta_{y2.13\dots k}; \dots; \beta_{yk.12\dots(k-1)}) = \frac{1}{n} \sum_{i=1}^n (u_{y_i} - \hat{u}_{y_i})^2 =$$

$$\frac{1}{n} \sum_{i=1}^n (u_{y_i} - \alpha_{y.123\dots k} - \beta_{y1.23\dots k}u_{1i} - \beta_{y2.13\dots k}u_{2i} \dots - \beta_{yk.12\dots(k-1)}u_{ki})^2 = \min$$

$\beta_{yi.12\dots k} \dots$ koeficienty beta, slouží k výpočtu dílčích regresních koeficientů a souhrnného korelačního koeficientu, zároveň z něho můžeme vyčíst, do jaké míry se projeví pozastavení vlivů určitých veličin na dané průměrné úrovni

$\alpha_{y.123\dots k} \dots$ absolutní parametr regresní rovnice, naší snahou bude dokázat, že se rovná nule

Postupujeme stejným způsobem jako u jednoduché lineární regrese. Postupně parciálně derivujeme parametry funkce $F(\alpha_{y.123\dots k}; \beta_{y1.23\dots k}; \beta_{y2.13\dots k}; \dots; \beta_{yk.12\dots(k-1)})$ a rovnice dáváme do rovnosti s nulou.

Pro usnadnění zavedeme zápis:

$$\alpha_{y.123\dots k} = \alpha$$

$$\beta_{y1.23\dots k} = \beta_1$$

$$\beta_{y2.13\dots k} = \beta_2$$

.....

$$\beta_{yk.12\dots(k-1)} = \beta_k$$

Získáme $(k+1)$ normálních rovnic:

$$\frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \alpha} = \frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_1} = \frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_2} = \dots = \frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_k} = 0$$

$$\frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \alpha} = \frac{2}{n} \sum_{i=1}^n (u_y - \alpha - \beta_1 u_1 - \beta_2 u_2 - \dots - \beta_k u_k)(-1) = 0$$

$$\frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_1} = \frac{2}{n} \sum_{i=1}^n (u_y - \alpha - \beta_1 u_1 - \beta_2 u_2 - \dots - \beta_k u_k)(-u_1) = 0$$

$$\frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_2} = \frac{2}{n} \sum_{i=1}^n (u_y - \alpha - \beta_1 u_1 - \beta_2 u_2 - \dots - \beta_k u_k)(-u_2) = 0$$

.....

$$\frac{\partial F(\alpha; \beta_1; \beta_2; \beta_k)}{\partial \beta_k} = \frac{2}{n} \sum_{i=1}^n (u_y - \alpha - \beta_1 u_1 - \beta_2 u_2 - \dots - \beta_k u_k)(-u_k) = 0$$

Pokrátíme zlomek $\frac{2}{n}$ a převedeme doleva členy bez parametrů a doprava členy s parametry:

$$\sum_{i=1}^n u_y = n\alpha + \beta_1 \sum_{i=1}^n u_1 + \beta_2 \sum_{i=1}^n u_2 + \dots + \beta_k \sum_{i=1}^n u_k$$

$$\sum_{i=1}^n u_y u_1 = \alpha \sum_{i=1}^n u_1 + \beta_1 \sum_{i=1}^n u_1^2 + \beta_2 \sum_{i=1}^n u_1 u_2 + \dots + \beta_k \sum_{i=1}^n u_1 u_k$$

$$\sum_{i=1}^n u_y u_2 = \alpha \sum_{i=1}^n u_2 + \beta_1 \sum_{i=1}^n u_1 u_2 + \beta_2 \sum_{i=1}^n u_2^2 + \dots + \beta_k \sum_{i=1}^n u_2 u_k$$

.....

$$\sum_{i=1}^n u_y u_k = \alpha \sum_{i=1}^n u_k + \beta_1 \sum_{i=1}^n u_1 u_k + \beta_2 \sum_{i=1}^n u_2 u_k + \beta_3 \sum_{i=1}^n u_3 u_k + \dots + \beta_k \sum_{i=1}^n u_k^2$$

Z podmínky $\bar{u} = 0$ plyne:

$$\sum_{i=1}^n u_y = \sum_{i=1}^n u_1 = \sum_{i=1}^n u_2 \dots = \sum_{i=1}^n u_k = 0$$

Dosadíme do první rovnice:

$$0 = n\alpha + \beta_1 0 + \beta_2 0 + \dots + \beta_k 0$$

$$n\alpha = 0 \rightarrow \alpha = 0$$

Odvoďme si další vztahy:

$$\frac{1}{n} \sum_{i=1}^n u_y^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \bar{y}}{s_y} \right)^2 = \frac{1}{ns_y^2} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{s_y^2} s_y^2 = 1$$

$$\frac{1}{n} \sum_{i=1}^n u_i^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_i} \right)^2 = \frac{1}{ns_i^2} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{s_i^2} s_i^2 = 1$$

$$\frac{1}{n} \sum_{i=1}^n u_i u_j = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{s_i s_j} = \frac{\text{cov}(x_i; x_j)}{s_i s_j} = r_{ij}$$

r_{ij} ... párový korelační koeficient, dán jednoduchou lineární závislostí dvou proměnných, více viz kapitola jednoduchá lineární regrese, $r_{ij} = r_{ji}$, $r_{ii} = r_{jj} = 1$

Použijeme následující odvozené vztahy do normovaných rovnic. Máme už jenom k rovnic, které jsme transformovali podle výše uvedených vztahů na příslušné konstanty r_{ij} , r_{yj} a β_i .

$$r_{y1} = \beta_1 + \beta_2 r_{12} + \dots + \beta_k r_{1k}$$

$$r_{y2} = \beta_1 r_{21} + \beta_2 + \dots + \beta_k r_{2k}$$

.....

$$r_{yk} = \beta_1 r_{k1} + \beta_2 r_{k2} + \dots + \beta_k$$

Máme k rovnic o k neznámých. Dosazovací, sčítací metodou nebo pomocí matic si můžeme odvodit koeficienty beta β_i . Nejsnazším postupem je použití webové aplikace pro výpočet matic. V praktické části diplomové práce budu vytvářet vlastní excelovskou aplikaci pro výpočet regresních koeficientů dvojnásobné a trojnásobné regresní a korelační analýzy. Řeším zde soustavu rovnic pomocí matic. $Ax = b \rightarrow x = A^{-1} * b$

$$A \dots \text{regulární matice} \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix}$$

$$A^{-1} \dots \text{inverzní matice} \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix}^{-1}$$

$$b \dots \text{sloupový vektor} \begin{bmatrix} r_{y1} \\ r_{y2} \\ \dots \\ \dots \\ r_{yk} \end{bmatrix}$$

$$x \dots \text{sloupový vektor neznámých} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \dots \\ \beta_k \end{bmatrix}$$

Koeficienty β_i tedy získáme vynásobením matic:

$$\text{inverze} \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix} * \begin{bmatrix} r_{y1} \\ r_{y2} \\ \dots \\ \dots \\ r_{yk} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \dots \\ \beta_k \end{bmatrix}$$

Dílčí regresní koeficienty $b_{yi.123\dots k}$ obdržíme vynásobením daného koeficientu beta $\beta_{yi.123\dots k}$ zlomkem $\frac{s_y}{s_i}$:

$$b_{yi.123\dots k} = \frac{s_y}{s_i} \beta_{yi.123\dots k}$$

Absolutní parametr regresní rovnice $a_{y.123\dots k}$ odvodíme z první normální rovnice, neboť dílčí regresní koeficienty už známe:

$$a_{y.123\dots k} = \bar{y} - b_{y1.23\dots k} \bar{x}_1 - b_{y2.13\dots k} \bar{x}_2 \dots - b_{yk.12\dots(k-1)} \bar{x}_k$$

Korelace ve vícenásobné RKA:

Dílčí korelační koeficienty nám pomáhají určit korelační závislost dvou veličin při pozastavení vlivu veličin ostatních na určité průměrné úrovni. Pomocí nich můžeme vyčíst vliv jednotlivých veličin na samotný regresní odhad.

$r_{yi.123\dots k}$... *dílčí korelační koeficient, vyjadřuje stupeň těsnosti lineární korelační závislosti y a x_i a tím tedy i přesnost regresního odhadu y a x_i ($b_{yi.12\dots k}$) při pozastavení vlivu zbylých proměnných na určité průměrné úrovni*

Pro jejich výpočet vycházíme ze sdružených regresních koeficientů:

$$r_{yi.12\dots k} = \sqrt{(b_{yi.12\dots k} \cdot b_{iy.12\dots k})}$$

Po dosazení obdržíme:

$$r_{yi.12\dots k} = \frac{r_{yi.12\dots k-1} - r_{yk.12\dots k-1} \cdot r_{ik.12\dots k-1}}{\sqrt{(1 - r_{yk.12\dots k-1}^2)(1 - r_{ik.12\dots k-1}^2)}}$$

Souhrnný korelační koeficient udává těsnost korelační závislosti vystupující proměnné y a vstupujících proměnných $x_1, x_2 \dots x_k$. Udává tedy spolehlivost našeho regresního odhadu.

$r_{y.123\dots k}$... *souhrnný korelační koeficient udává těsnost daného regresního odhadu*

Je dán vzorcem:

$$r_{y.123\dots k} = \sqrt{(\beta_{y1.23\dots k} \cdot r_{y1} + \beta_{y2.13\dots k} \cdot r_{y2} + \dots + \beta_{yk.12\dots k-1} \cdot r_{yk})}$$

2 Identifikace a charakteristika veličin

Jedním z cílů mé diplomové práce je odvození rovnice pro výpočet respektive prognózování hrubého domácího produktu HDP zemí EU v závislosti na hodnotách námi zvolených vstupujících veličin. Po dlouhé úvaze jsem vybral 7 vstupujících veličin, které dle mého uvážení ovlivňují výslednou hodnotu HDP a zároveň nejsou obsaženy v některém ze vzorců pro jeho výpočet (výdajová, důchodová, produkční metoda). Dále prozkoumám závislosti mezi veličinami a odstraním z modelu statisticky nepotřebné veličiny. Sestavím tedy regresní model, jen z těch veličin, které jsou pro náš regresní odhad statisticky významné.

Závisle proměnnou y_i je HDP. Charakterizujme si tedy tento ukazatel a ukažme si způsoby jeho výpočtu.

HDP je suma tržních hodnot veškeré produkce finálních statků, vyprodukovaných výrobními faktory na daném území za dané období (zpravidla rok), bez ohledu na to, kdo tyto výrobní faktory vlastní. Jde tedy o klíčový ukazatel výkonnosti ekonomiky dané země. Rozlišujeme HDP reálné a nominální. U nominálního je produkce oceněna v tržních cenách. To jest v cenách období, ve kterém jsou statky vyrobeny. Velikost nominálního HDP je tedy podmíněno změnám tržních cen a produkce výrobků a služeb. Reálné HDP naopak vyjadřuje celkovou peněžní hodnotu produkce ve stálých cenách. Oprošťujeme ho od vlivu pohybu cenových hladin. Zvolíme ceny některého roku za základ a pomocí nich oceníme produkci následujících i předchozích let. Nesledujeme změny tržních cen výrobků a služeb v daných letech. Výsledná hodnota HDP závisí tedy pouze na objemu produkce.

2.1 Způsoby výpočtu HDP

2.1.1 Výdajová metoda

Protože výdaj za zboží odpovídá tržní ceně tohoto zboží. Uvažujeme, že celková hodnota HDP je dána sumou veškerých agregátních výdajů (výdajů domácností, firem, veřejné správy, zahraničních osob) na daném území za daný časový interval (zpravidla jeden rok). Nominální HDP získáme tedy ze vzorce:

$$HDP = C + I + G + NX$$

C ... spotřeba domácností

I ... investiční výdaje firem, nákup finální produkce

G ... výdaje státu na nákup finální produkce

NX ... čistý export, $NX = \text{export} - \text{import}$

2.1.2 Důchodová metoda

Protože každý výdaj z ekonomického hlediska je zároveň i něčím příjmem (neboli důchodem), můžeme celkové HDP vyjádřit jako součet příjmů domácností, státu a firem. Jelikož odpisy jsou náklady, které nikam neodvádíme, jenom nám snižují zisk pro výpočet daně, jsou připočteny k příjmům. Nominální HDP získáme tedy ze vzorce:

$$HDP = PB + Tx + Pr + Am + In + CO - CG$$

PB ... osobní náklady firem

Tx ... nepřímé daně

Pr ... zisky firem

Am ... odpisy

In ... úroky firem

CO ... čistý smíšený důchod, odměny vlastníkům

CG ... peněžní dotace

2.1.3 Produkční metoda

HDP získáme sečtením přidaných hodnot firem, ze všech odvětví..

$$HDP = \sum_1^x PH$$

PH = výkony (tržby z prodeje vlastních výrobků a služeb, aktivace, změna zásob vlastní výroby) – výkonová spotřeba (spotřeba materiálu, energie a služeb)

Hodnoty HDP zemí EU jsem čerpal z portálu z teritoriálních informací portálu <http://www.businessinfo.cz>, který spravuje Ministerstvo průmyslu a obchodu. Data jsou ze statistik pro rok 2013.

2.2 Výběr a charakteristika regresorů

Snažil jsem se najít všechny proměnné, které, dle mého názoru jsou ve statistické závislosti se závisle proměnnou y . Ty proměnné, které ovlivňují výslednou hodnotu HDP a zároveň nejsou obsaženy v některém ze vzorců pro jeho výpočet. Vybral jsem 7 regresorů:

k	Název regresoru	Jednotky []
1	Počet obyvatel	[1000]
2	Počet nezaměstnaných	[1000]
3	Počet zaměstnaných	[1000]
4	Roční míra inflace	[%]
5	Roční spotřeba elektrické energie	[1000GWh]
6	Počet MSP	[1000]
7	Produktivita práce	[USD/h]

Tabulka 2.2.1 Regresory $x_1, x_2 \dots x_7$

2.2.1 Počet obyvatel

Obyvatelé jsou ti, kteří odebírají finální produkci. Ve většině zemí EU se spotřeba domácností podílí alespoň 50 % na výsledné hodnotě HDP. Lidé jsou důležitým zdrojem ve výrobním procesu. Jejich efektivní využití napomáhá organizacím zvyšovat přidanou hodnotu. Údaje o počtu obyvatel jsem čerpal z teritoriálních informací portálu <http://www.businessinfo.cz>, který spravuje Ministerstvo průmyslu a obchodu. Data jsou ze statistik pro rok 2013.

2.2.2 Počet nezaměstnaných a počet zaměstnaných

Nezaměstnanost je považována za závažný ekonomický problém. Státy s vysokou mírou nezaměstnanosti vynakládají vysoké finanční prostředky na výplaty sociálních důchodů. V rozpočtu států potom nezůstávají dostatečné prostředky na investice, které napomáhají zvyšovat přidanou hodnotu. Navíc není žádným překvapením, že nezaměstnaní mají nižší příjmy než zaměstnaní. Jejich spotřeba je tedy nižší. Pro svou analýzu jsem zvolil jako proměnnou i počet pracujících. Tím chci odstranit vliv generačního rozložení obyvatel v jednotlivých zemích EU. Údaje o počtu zaměstnaných a nezaměstnaných jsem čerpal z portálu OECD, http://stats.oecd.org/Index.aspx?DatasetCode=LFS_D. Data jsou ze statistik pro rok 2013.

2.2.3 Roční míra inflace

Inflace, neboli míra změny cenové hladiny, vytváří nejisté prostředí pro podnikatelské činnosti. Pracovníci i podnikatelé tím mohou být odrazeni od své aktivity a také tvorby úspor. Vysoká inflace v investičním prostředí zvyšuje diskontní sazbu. Mnoho dlouhodobých investičních plánů, kvůli tomu nemusí sehnat investory. Na druhou stranu inflace může vést k nárůstu spotřebních výdajů. S rostoucí inflací roste i finanční ocenění produkce. Údaje o roční míře inflace jsem čerpal z Eurostatu, <http://ec.europa.eu/eurostat/web/national-accounts/statistics-illustrated>. Data jsou pro rok 2013.

2.2.4 Roční spotřeba elektrické energie

Elektrická energie je hlavním zdrojem energie pro průmysl a dopravu. Množství spotřebované elektrické energie je pomyslným odrazem velikosti výrobních kapacit jednotlivých zemí. Efektivní využití energie šetří náklady firem, tím tedy i zvyšuje jejich přidanou hodnotu. Údaje o spotřebě elektrické energie zemí EU jsem čerpal z portálu index mundi, <http://www.indexmundi.com/g/g.aspx?v=81&c=bo&l=en>. Data jsou pro rok 2013.

2.2.5 Počet malých a středních podniků

EU definuje MSP jako podniky s maximálním počtem zaměstnanců do 250. Ve většině zemí EU malé a střední podniky zaměstnávají zhruba dvě třetiny pracujících a vytvářejí více jak polovinu přidané hodnoty. Jsou to podniky, které většinu zisku vkládají do vývoje. Je pro ně typická vysoká míra inovativnosti a pružné reagování na změnu požadavků zákazníků. Údaje o počtu MSP jsem čerpal z portálu European Commission, <http://ec.europa.eu/enterprise/policies/sme.html>. Data jsou ze statistik pro rok 2013

2.2.6 Produktivita práce

Je hlavním ukazatelem výkonnosti výrobních faktorů jednotlivých zemí. Vyjadřuje schopnost výrobního systému transformovat vstupy na výstupy. Je dána vzorcem:

$$\text{Produktivita práce} = \frac{\text{výstupy}}{\text{vstupy}}$$

výstupy ... produkce organizací, bývá vyjádřena v peněžních jednotkách

vstupy ... jsou dány v důsledku využití technologií, pracovních zkušeností, zručností a potřebou kapitálu

Produktivita práce roste se zvyšující se zručností a znalostí pracujících, s užitím dokonalejších strojů, technologií, organizací výroby atd...

Pro naši analýzu jsem vymyslel vzorec produktivity práce:

$$\text{Produktivita práce} = \frac{HDP}{PZ * ((52 * \emptyset POHT) - (PS * \frac{\emptyset POHT}{5}))}$$

HDP ... hrubý domácí produkt [mld USD]

PZ ... počet zaměstnaných [1000]

∅ POHT ... průměrný počet odpracovaných hodin týdně [h/týdně]

PS ... počet státních svátků, dní bez práce [dní]

52 ... počet týdnů v roce

5 ... počet pracovních dnů v týdnu

Jmenovatel rovnice vyjadřuje roční efektivní časový fond pro všechny aktivně pracující obyvatele dané země. Produktivita práce je zde tedy popsána hrubým domácím produktem vytvořeným za jednu odpracovanou hodinu jedním pracovníkem. Níže uvádím tabulku průměrných odpracovaných hodin týdně a počtu dní bez práce pro jednotlivé země EU. Údaje jsem čerpal z teritoriálních informací portálu <http://www.businessinfo.cz>, který spravuje Ministerstvo průmyslu a obchodu. Data jsou ze statistik pro rok 2013.

Země EU	PS [1000]	Ø POHT [h/týdně]
Belgické království	12	35,2
Bulharská republika	10	41,0
Česká republika	12	40,5
Dánské království	12	34,5
Estonská republika	11	39,5
Finská republika	14	36,7
Francouzská republika	11	36,4
Chorvatská republika	13	39,9
Irská republika	10	35,3
Italská republika	11	37,0
Kyperská republika	10	38,9
Litva	12	38,7
Lotyšská republika	12	41,3
Lucembursko	11	36,2
Maďarsko	12	40,1
Maltská republika	14	38,4
Spolková republika Německo	15	34,4
Nizozemské království	9	29,7
Polská republika	12	40,0
Portugalská republika	8	39,1
Rakouská republika	13	37,5
Rumunsko	11	40,5
Řecká republika	10	39,8
Slovenská republika	10	40,4
Slovinská republika	12	39,9
Velká Británie	8	36,4
Španělské království	11	38,3
Švédské království	17	35,7

Tabulka 2.2.6.1 Průměrných odpracovaných hodin týdně a počtu dní bez práce

3 Volba regresního modelu

Jde o proces hledání ideálního počtu veličin pro náš regresní odhad. Neexistuje žádné kouzelné pravidlo, které by nám pomohlo pro dané empirické šetření najít nejvhodnější model. Nicméně můžeme, alespoň podle určitých metod a pravidel snížit počet v úvahu přicházejících variant. Jinými slovy, sestavíme regresní model jen z těch veličin, které jsou statisticky významné, použijeme metodu linearity a metody multikolinearity.

Ale pozor, nejde jenom o to odstranit z modelu přebytečné veličiny, které by zatěžovaly náš výpočet. Může se stát, že opomineme některou z důležitých vstupujících veličin. Výběr je potom nedostatečně reprezentující a výsledky analýzy nejsou zcela vypovídající. Je proto vždy lepší začínat s mnohorozměrným pozorováním a postupně odstraňovat statisticky nepotřebné veličiny. Samozřejmě omezenost zdrojů nám neumožňuje sledovat všechny faktory, které působí na náš problém. Z výše uvedeného tedy vyplývá, že statistik by se měl snažit najít co možno největší počet veličin, které jsou zároveň časově a nákladově přijatelné. A následně pomoci určitých metod a pravidel odstranit přebytečné veličiny. Sami jistě uznáte, že to není nic lehkého.

Dalším krokem je výběr typu regresní funkce. Když už máme vybrán počet vstupujících proměnných, musíme si odpovědět na otázku, „jakou formou je co nejvhodněji vyjádřit“? Volíme regresní funkci, jejíž tvar se co možno nejvíce podobá charakteru průběhu empirických hodnot, získaných z pozorování. Výběr tvaru funkce vychází zpravidla z dřívějších zkušeností a poznatků daného statistika. Já se budu zabývat lineární závislostí, jelikož jde o nejčastější typ vyrovnávání. Navíc lze některé nelineární vztahy transformovat na vztahy lineární. Budu tedy muset dokázat, že proměnné jsou v lineárním vztahu.

V datech by neměly být odlehlé hodnoty. Regresní analýza je na ně citlivá. I jedna odlehlá hodnota může ovlivnit odhady parametrů rovnice. Odlehlé hodnoty můžeme odhalit buď vizuálně, prozkoumáním bodových grafů, nebo podle testů (Grubbsův test, Dixonův test). V mé práci použiji Grubbsův test.

Kvalita regresního modelu se posuzuje dle schopnosti vyhovět požadavkům dané úlohy. Vezmeme si například model, který prognózuje návštěvnost letního kina v závislosti na počasí, srážkách atd. Pouhým porovnáním skutečných hodnot návštěvnosti od hodnot modelových získáme přehled o přesnosti daného modelu. Pokud nejsou velké rozdíly mezi těmito hodnotami a směrodatné chyby regresního odhadu splňují požadavek minima,

považujeme model za kvalitní. Co je velký a co malý rozdíl, závisí na každém konkrétním případě a uvážení příslušného statistika.

3.1 Postup volby regresního modelu

Při sestavování regresního modelu budu postupovat následujícím způsobem:

1. Identifikuji všechny faktory, které mohou mít vliv na naši závisle proměnnou y (HDP). Samozřejmě vyberu jen ty faktory $x_1, x_2 \dots x_k$, ke kterým půjdou sehnat data. Jelikož pracuji s údaji ze sekundárních zdrojů, jsou mé možnosti omezené.
2. Pomocí metody linearitity popíši, jak velký vliv má každá nezávisle proměnná $x_1, x_2 \dots x_k$, na závisle proměnnou y . Ty s malým vlivem odstraním z modelu.
3. Pomocí metod multikolinearity odstraním nezávislé proměnné $x_1, x_2 \dots x_k$, mezi kterými je velká lineární závislost.
4. Odstraním pomocí Grubbsova testu všechny odlehle hodnoty nezávisle proměnných $x_1, x_2 \dots x_n$.
5. Provedu testy reziduí. Vždy pro dvojice proměnných, závisle proměnná y a některou ze zbylých nezávisle proměnných ($x_1, x_2 \dots x_k$). Rezidua by měla být náhodná, nezávislá a podléhat normálnímu rozdělení $N(0; \sigma^2)$. Rozptyl σ^2 musí být konstantní.
6. Provedu regresní a korelační úkon.

3.1.1 Metoda linearitity

Tato metoda nám pomáhá odstranit z regresního modelu ty nezávislé proměnné, které mají se závisle proměnnou malou statistickou závislost. Náš regresní model předpokládá, že závisle proměnná y bude záviset lineárně na n nezávisle proměnných $x_1, x_2 \dots x_n$. Musím tedy dokázat, že průběhy dvojic proměnných y a x_i mají lineární povahu. Pokud vztahy mezi dvojicí proměnných y a x_i nejsou lineární, pokusím se data statisticky transformovat (logaritmovat, odmocňovat ...), tak abychom naplnili požadavek linearitity.

Postup:

1. Vypočtu všechny párové regresní koeficienty r_{yi} .
2. Pokud některý z $r_{yi} < 0,6$, pokusím se data transformovat.
3. Pokud ani transformace nepomůže, nezávisle proměnnou x_i se slabou statistickou závislostí odstráním z modelu.

3.1.2 Multikolinearita

Multikolinearitou se rozumí vzájemná lineární závislost mezi nezávislými proměnnými. Značí ji vysoká hodnota párového korelačního koeficientu mezi dvěma vstupujícími proměnnými x_i a x_j . Důvody multikolinearity mohou být různé. Relativně nejjednodušším a nejčastějším případem je zatěžování regresního modelu nadbytečnými vstupujícími proměnnými. V mé práci se snažím pomoci statistických metod tyto nadbytečné proměnné identifikovat a následně odstranit z regresního modelu. Použiji dvě metody:

Metodu postupného vyřazování

Zde odstráním vždy jednu z dvojice vstupujících proměnných mezi, kterými je silná závislost, neboli hodnota párového korelačního koeficientu $r_{ij} \geq 0,8$. Do výpočtu vstupují pouze proměnné, které mají významnou statistickou závislost se závisle proměnnou y_i . Ty už známe z metody linearity.

Postup:

1. Vypočtu všechny párové regresní koeficienty r_{yj} a r_{ij} .
2. Vyberu nezávisle proměnnou x_i , která má nejvyšší korelační závislost se závisle proměnnou y_i (nejvyšší hodnotu r_{yj}).
3. Jako další sleduji nezávisle proměnné x_j , které mají vysokou korelační závislost s mou vybranou proměnnou x_i z druhého kroku. Pokud je korelační koeficient mezi vybranou proměnnou vyšší než 0,8 $\rightarrow r_{ij} \geq 0,8$, vyřazuji proměnnou x_j z modelu.

Metodu Stepwise

Zde proměnné přiřazují do modelu postupně, máme tedy více modelů. Každý nový model má o jednu proměnnou navíc. V každém kroku vypočtu reziduální součty čtverců RSS a provedu F -test. Postup se aplikuje tak dlouho, dokud je přínos přidání proměnné statisticky významný. Dokud se nám statisticky významně snižuje reziduální součet čtverců RSS.

Vycházíme z rozkladu čtverců. Variabilitu vystupující veličiny můžeme také vyjádřit pomocí součtu čtverců odchylek empirických hodnot od jejich průměru. Celková suma čtverců TSS se rovná:

$$TSS = \sum_1^n (y_i - \bar{y}) * (y_i - \bar{y}) = \sum_1^n (y_i - \bar{y})^2$$

Dále celkovou sumu čtverců můžeme rozložit na dvě položky.

$$TSS = MSS + RSS$$

MSS neboli modelová suma čtverců je část z celkové sumy čtverců, která je vysvětlena závislostí vystupující veličiny na regresorech. Je dána jako součet čtverců odchylek vyrovnaných hodnot od průměru empirických hodnot.

$$MSS = \sum_1^n (\hat{y}_i - \bar{y}) * (\hat{y}_i - \bar{y}) = \sum_1^n (\hat{y}_i - \bar{y})^2$$

RSS neboli reziduální suma čtverců je část z celkové sumy čtverců, kterou mimo jiné používá metoda nejmenších čtverců. Je dána jako součet čtverců odchylek vyrovnaných hodnot od empirických.

$$RSS = \sum_1^n (y_i - \hat{y}_i) * (y_i - \hat{y}_i) = \sum_1^n (y_i - \hat{y}_i)^2$$

Pomoci těchto rozkladů můžeme také určit těsnost regresního odhadu. Koefficient (index) determinace R^2 udává, do jaké míry regresní model vysvětluje variabilitu vystupující veličiny.

$$R^2 = \frac{MSS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

Odmocníme a získáme korelační index R :

$$R = \sqrt{R^2} = \sqrt{\frac{MSS}{TSS}} = \sqrt{\frac{TSS - RSS}{TSS}} = \sqrt{1 - \frac{RSS}{TSS}}$$

Korelační index může tedy nabývat hodnot od 0 do 1. Ze vzorce je patrné, že když je RSS roven nule pak $R = 1$, regresní odhad je 100% spolehlivý. Naopak když $RSS = TSS$, pak $R = 0$, regresní odhad je 0% spolehlivý (tedy nekorelovaný).

Myšlenkou této metody je zjistit, zdali se přidání regresoru statisticky významně projeví na výsledné hodnotě RSS, respektive zdali se hodnota RSS statisticky významně sníží. Pokud se po přidání regresoru hodnota RSS nesníží nebo sníží statisticky nevýznamně, regresor vypouštíme z regresního odhadu.

Přidáním i -tého regresoru buď snížíme hodnotu výsledného RSS, nebo ji zachováme stejnou:

$$\Delta RSS = RSS_{(k-1)} + RSS_k$$

$$\Delta RSS \geq 0$$

i ... počet regresorů, $i = 1, 2, 3 \dots k$

Kritérium pro posouzení statistické významnosti je takzvané parciální F . Pro i -tý regresor F_i :

$$F_i = \frac{\Delta RSS}{RSS_{(k-1)}/(n-k)} = \frac{RSS_{(k-1)} + RSS_k}{RSS_{(k-1)}/(n-k)}$$

Když $F_i < F_{1,(n-k-1)}(1 - \alpha)$ je po přidání i -tého regresoru snížení i -tého statisticky nevýznamné. Regresor vypouštíme z regresního odhadu.

k ... současný počet regresorů

n ... počet pozorování

$F_{1,(n-k-1)}(1 - \alpha)$... Hodnota pravděpodobnosti Fisherova rozdělení pro 1 a $(n - k - 1)$ stupně volnosti na dané zvolené hladině významnosti α

Postup:

1. Vypočtu všechny párové regresní koeficienty r_{yj} a r_{ij} .
2. Vyberu nezávisle proměnnou x_i (první regresor), který má nejvyšší korelační závislost se závisle proměnnou y_i (nejvyšší hodnotu r_{yj}).
3. Provedeme pro tyto dvě proměnné regresní úkon (jednoduchá lineární RKA), a vypočteme RSS_k .
4. Vybereme další nezávisle proměnnou x_j (druhý regresor), který má s prvním regresorem nejnižší párovou korelační závislost (nejnižší hodnotu r_{ij}). Znovu provedeme regresní úkon (dvojnásobná lineární RKA), vypočteme RSS_k a provedeme F -test. Pokud je přínos regresoru statisticky významný, přidáme regresor do odhadu. Analogicky postupujeme pro zbylé regresory. Pokud přínos není statisticky významný, regresor vyzárují z regresního odhadu a testují zbylé regresory.

3.1.3 Grubbsův test

Odlehlé hodnoty mohou ovlivnit výsledky RKA. I jedna odlehlá hodnota se projeví na odhadu parametrů rovnice, proto je důležité tyto hodnoty identifikovat a následně je vyloučit ze souboru hodnot.

Postup:

1. Seřadím empirické hodnoty vzestupně ($x_{(1)} < x_{(2)} < \dots < x_{(n)}$).
2. Vypočtu aritmetický průměr $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ a směrodatnou odchylku $s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ ze všech hodnot zkoumané veličiny.
3. Vypočtu testovací kritérium pro nejnižší a nejvyšší hodnotu.

$$T_1 = \frac{\bar{x} - x_{(1)}}{s_x}$$

$$T_n = \frac{x_{(n)} - \bar{x}}{s_x}$$

4. Vypočtené hodnoty testovacích kritérií porovnáám s tabulkovou kritickou hodnotou dané hladiny významnosti α . Pokud $T_{1(n,\alpha)} > T_{krit} \rightarrow$ vyloučím hodnotu $x_{(1)}$ ze souboru. Musím provést nový výpočet pro zbylé hodnoty. Analogicky postupuji i pro maximální hodnotu $x_{(n)}$, když $T_{n(n,\alpha)} > T_{krit} \rightarrow$ vyloučím hodnotu $x_{(n)}$ ze souboru.

n	$\alpha = 0,05$	$\alpha = 0,01$
3	1,412	1,414
4	1,689	1,723
5	1,869	1,955
6	1,996	2,130
7	2,093	2,265
8	2,172	2,374
9	2,237	2,464
10	2,294	2,540
11	2,343	2,606
12	2,387	2,663
13	2,426	2,714
14	2,461	2,759
15	2,493	2,800
16	2,523	2,837
17	2,551	2,871
18	2,577	2,903
19	2,600	2,932
20	2,623	2,959
21	2,644	2,984
22	2,664	3,008
23	2,683	3,030
24	2,701	3,051
25	2,717	3,071

Tabulka 3.1.3.1. Kritických hodnot Grubbsova testu

3.1.4 Analýza reziduí

Analýzu reziduí provádím vždy pro proměnnou y a některou z nezávisle proměnných $x_1, x_2 \dots x_k$. Chci dokázat, že rezidua mají povahu náhodných veličin, jejichž rozdělení bude při větších výběrech podléhat normálnímu rozdělení se střední hodnotou blížící se k nule. Rezidua musí tedy plnit podmínky:

1. Rezidua jsou náhodná a nezávislá.
2. Rezidua mají normální rozdělení $N(0; \sigma^2)$.
3. Rozptyl reziduí σ^2 je konstantní.

Provádím tedy testování o náhodnosti reziduí. Máme dvě hypotézy:

H_0 : rezidua jsou náhodná

H_1 : rezidua nejsou náhodná

Vycházím z testového kritéria:

$$U = \frac{S + 0,5 - \frac{n-1}{2}}{\sqrt{\frac{n-1}{4}}}$$

n ... počet sledování

S ... počet kladných rozdílů ($e_{i+1} - e_i$), $e_i = \text{rezidua} = (y_i - \hat{y}_i)$

Nulovou hypotézu přijímám, když testové kritérium je menší než 1,96. Odpovídá to hodnotě Studentova rozdělení na hladině významnosti $\alpha = 0,05 \rightarrow u_{0,975} = 1,96, u_{0,025} = -1,96$.

$$U = \frac{S + 0,5 - \frac{n-1}{2}}{\sqrt{\frac{n-1}{4}}} < 1,96 \dots \text{přijímáme } H_0$$

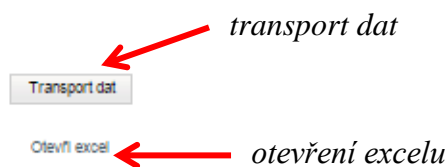
4 Analýza závislost HDP na zvolených regresorech

4.1 Příručka použití aplikace

K analýze poslouží aplikace, kterou jsem vytvořil [PŘ1]. Aplikace uživatelům usnadní zkoumání závislosti, pomůže najít ideální počet veličin pro volbu regresního modelu a následně provede regresní a korelační úkon.

4.1.1 Zápis dat

Data můžeme zadat dvěma způsoby. Buď na webové stránce <http://student.fsid.cvut.cz/~sangajak/diplomovápráce.html> do připraveného databázového formuláře. Data poté odešleme do excelovské aplikace pomocí tlačítka transport dat, viz *Obrázek 4.1.1.1*. Nebo si stáhneme ze stránek excelovskou aplikaci a data zapíšeme rovnou tam, viz *Obrázek 4.1.1.1*. Máme možnost zapsat data až pro 30 měření a 11 proměnných (1 závisle proměnná y_i a 10 nezávisle proměnných $x_1, x_2 \dots x_{10}$, viz [PŘ1] list “Vstupní oblast dat“. Aby aplikace fungovala, je třeba povolit makra. Excel se vás na povolení sám zeptá. Dále pokud nemáme, nainstalujeme v Soubor→Možnosti →Doplňky Analytické nástroje – VBA.



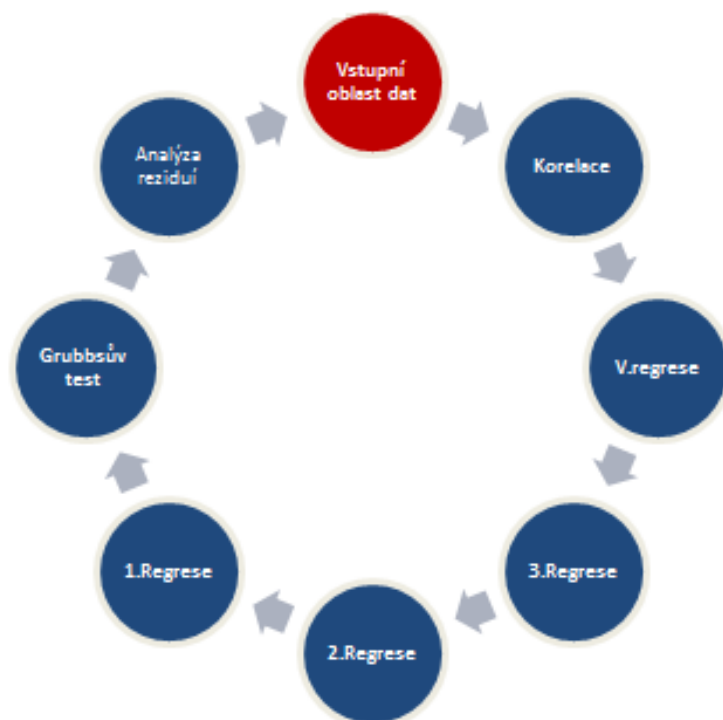
Obrázek 4.1.1.1 Zápis dat

Po zadání dat klikneme v [PŘ1] na tomtéž listě na tlačítko “Zadejte názvy proměnných“ otevře se nám uživatelský formulář, viz *Obrázek 4.1.1.2*. Zde zadáme názvy proměnných a jednotek, poté potvrdíme stisknutím tlačítka “ok“.

Zadejte názvy proměnných:	
Závisle proměnná	Jednotky []
1.regresor	Jednotky []
2.regresor	Jednotky []
3.regresor	Jednotky []
4.regresor	Jednotky []
5.regresor	Jednotky []
6.regresor	Jednotky []
7.regresor	Jednotky []
8.regresor	Jednotky []
9.regresor	Jednotky []
10.regresor	Jednotky []

Obrázek 4.1.1.2 Uživatelský formulář názvy proměnných

Pokud chcete načíst původní hodnoty, klikněte na tlačítko “Načíst původní hodnoty“. Pokud chcete odstranit prázdné řádky v tabulce, klikněte na “Smazat prázdné řádky“. Mezi listy v [PŘ1] můžeme přepínat pomocí uživatelské lišty, viz Obrázek 4.1.1.2.



Obrázek 4.1.1.2 Uživatelská lišta

4.1.2 Test linearity

Na listě “Korelace“ klikněte na tlačítko “Korelace“. Aplikace se nás zeptá na počet měření, viz *Obrázek 4.1.2.1*. A na počet proměnných, viz *Obrázek 4.1.2.2*. Počtem proměnných se myslí součet všech proměnných. Tedy když máme 10 regresorů, zadáme počet proměnných 11 (1 závisle proměnná + 10 nezávisle proměnných). Počet měření 1 až 30.

Obrázek 4.1.2.1 Počet měření

Obrázek 4.1.2.2 Počet proměnných

Po zadání těchto dvou parametrů se nám zpusť výpočet párových korelačních koeficientů r_{yj} a r_{ij} . Výsledky jsou zde na listě zobrazeny v tabulce, viz *Obrázek 4.1.2.3*. Nyní můžeme koeficienty analyzovat, a provést metodu linearity a metodu multikolinearity (metodu postupného vyzařování), viz 3. kapitola.

	Závisle proměnná	1.regresor	2.regresor	3.regresor	4.regresor	5.regresor	6.regresor	7.regresor	8.regresor	9.regresor	10.regresor
Závisle proměnná	1,00										
1.regresor	0,99	1,00									
2.regresor	0,99	1,00	1,00								
3.regresor	0,99	1,00	1,00	1,00							
4.regresor	0,99	1,00	1,00	1,00	1,00						
5.regresor	0,98	0,94	0,94	0,94	0,94	1,00					
6.regresor	0,68	0,77	0,77	0,77	0,77	0,52	1,00				
7.regresor	-0,37	-0,32	-0,32	-0,32	-0,32	-0,43	0,00	1,00			
8.regresor	0,27	0,13	0,13	0,13	0,13	0,45	-0,52	-0,43	1,00		
9.regresor	0,99	1,00	1,00	1,00	1,00	0,94	0,77	-0,32	0,13	1,00	
10.regresor	0,99	1,00	1,00	1,00	1,00	0,94	0,77	-0,32	0,13	1,00	1,00

Obrázek 4.1.2.3 Výsledky párových korelačních koeficientů r_{yj} a r_{ij}

4.1.3 Test multikolinearity

Párové korelační koeficienty r_{ij} již známe, není tedy problém provést metodu postupného vyřazování, viz 3. kapitola.

Na listě “V.regrese“ zpustíme tlačítkem “Regrese“ vícenásobnou regresi pro námi zadané hodnoty z listu “Vstupní oblast dat“. Znovu se nás aplikace zeptá na počet měření a na počet proměnných. Po zadání nám aplikace vygeneruje regresní rovnici, regresní koeficienty, souhrnný korelační koeficient, determinační koeficient, adjustovaný determinační koeficient, vyrovnané hodnoty, analýzu rozptylu a testy statistické významnosti regresního modelu.

Metoda Stepwise je zde vyložena pomocí adjustovaného determinačního koeficientu. Pokud je adjustovaný korelační koeficient, neboli nastavená hodnota spolehlivosti R, s přidáním regresoru vyšší, je vliv regresoru statisticky významný, viz *Obrázek 4.1.3.1*. Začneme s jednoduchou lineární závislostí. Přidáme do listu “Vstupní oblast dat“ pouze naměřená data pro závisle proměnnou y_i a 1. regresor x_{1i} . Provedme regresi, zase přes list “V.regrese“. Postupně přidávejme hodnoty pro další regresory a sledujme, jak se projeví vliv přidání regresoru na výslednou hodnotu adjustovaného determinačního koeficientu. Regresory přidávejme v pořadí, které určíme na základě výsledků analýzy lineární metody a metody postupného vyřazování. Pokud je adjustovaný determinační koeficient s přidáním regresoru vyšší, je vliv přidání nového regresoru statisticky významný.

Test významnosti regresorů		
adjust. determinační koeficient	59,0%	2 regresorů
Pokud je adj. determinační koeficient s přidáním regresoru vyšší, je vliv nového regresoru statisticky významný		

Obrázek 4.1.3.1 Adjustovaný determinační koeficient

Regresní model je statisticky významný, když parciální F získané z analýzy rozptylu (Anova) je větší než hodnota pravděpodobnosti Fisherova rozdělení pro $(k-1)$ a $(n-k-1)$ stupně volnosti na dané zvolené hladině významnosti α , viz *Obrázek 4.1.3.2*.

$$F > F_{\alpha} (k-1, n-k-1) \dots \text{regresní model je statisticky významný}$$

k ... současný počet regresorů

n ... počet pozorování

ANOVA					
	Rozdíl	SS	MS	F	Významnost F
Regrese	2	26,06190476	13,03095238	3,392004958	0,137581305
Rezidua	4	15,36666667	3,841666667		
Celkem	6	41,42857143			

Obrázek 4.1.3.2 Analýza Anova

Test významnosti regresního modelu:

H0 - vícenásobný lineární regresní model není významný ($R^2 = 0$) $F < F_{0,95}(2, 4)$
H1 - vícenásobný lineární regresní model je významný ($R^2 \neq 0$) $F > F_{0,95}(2, 4)$

determinační koeficient	0,629	62,9%	
korelační koeficient	0,793	79,3%	

testové kritérium	F	✖ 3,39	
kritická hodnota	$F_{0,95}(2, 4)$	6,94	

Obrázek 4.1.3.2 Test významnosti regresního modelu

4.1.4 Grubbsův test

Test provedeme na listě “Grubbsův test“. Data seřadíme od nejmenšího po největší kliknutím na tlačítko “Seřadit data“. Poté už jen stačí porovnat testovací kritérium s kritickou hodnotou, viz *Obrázek 4.1.4.1*. Pokud $T_{1(n,\alpha)} > T_{krit}$ → vyloučíme hodnotu $x_{(1)}$ ze souboru. Musíme provést nový výpočet pro zbylé hodnoty. Analogicky postupujeme i pro maximální hodnotu $x_{(n)}$, když $T_{n(n,\alpha)} > T_{krit}$ → vyloučíme hodnotu $x_{(n)}$ ze souboru.

Testové kritérium	Kritické hodnoty	$\alpha = 0,05$
	$T_{\alpha}(n, .05, \alpha)$	
$T_{(1)} =$ ✓ 1,136		1,996
$T_{(n)} =$ ✓ 1,784		1,996

Testové kritérium	Kritické hodnoty	$\alpha = 0,01$
	$T_{\alpha}(n, .01, \alpha)$	
$T_{(1)} =$ ✓ 1,136		2,130
$T_{(n)} =$ ✓ 1,784		2,130

Obrázek 4.1.4.1 Grubbsův test

Na listě je výpočet pro všech deset regresorů.

4.1.5 Analýza reziduí

Na listě “*Analýza reziduí*“ stačí pouze porovnat testovací kritérium s hodnotou 1,96, viz *Obrázek 4.1.5.1*. Pokud testovací kritérium $U < 1,96$ mají rezidua povahu náhodných veličin. Na listě je výpočet pro všechny dvojice $y_i : x_1, x_2 \dots x_{10}$.

$$U = \frac{S + 0,5 - \frac{n-1}{2}}{\sqrt{\frac{n-1}{4}}} < 1,96$$

<u>Testové kritérium</u>	<u>Kritické hodnoty</u> $\alpha = 0,05$
U = ✔ 1,265	1,960

Obrázek 4.1.5.1 Analýza reziduí

4.1.6 Regresní a korelační analýza

Regresní a korelační analýzu až pro 10 regresorů provedeme na listě “*V.regrese*“. Nebo na listech “*1.regrese, 2.regrese a 3 regrese*“, kde je navíc výpočet pro jeden dva nebo tři regresory podrobně rozepsán podle kroků z 1. kapitoly. Data pro RKA zadáváme vždy do listu “*Vstupní oblast dat*“. Aplikace nám vygeneruje regresní rovnici, regresní koeficienty, souhrnný korelační koeficient, determinační koeficient, adjustovaný determinační koeficient, vyrovnané hodnoty, analýzu rozptylu a testy statistické významnosti regresního modelu.

4.2 Analýza HDP

4.2.1 Zápis dat

Jak již jsem výše psal, budu analyzovat vliv 7 regresorů $x_1, x_2 \dots x_7$ na výslednou hodnotu závisle proměnné y_i , hrubého domácího produktu zemí EU. EU je politická a ekonomická unie tvořena 28 evropskými státy. Mám tedy vždy 28 hodnot pro každou proměnnou, viz *Obrázek 4.1.5.1*. Tyto data jsem zadal do [PŘ1] do listu “*Vstupní oblast dat*“. Následně jsem postupovat dle příručky použití aplikace, viz výše. Mou snahou je najít ideální počet veličin pro regresní odhad. A následně provést regresní a korelační úkon.

Země Eu	Y _i	X _{1i}	X _{2i}	X _{3i}	X _{4i}	X _{5i}	X _{6i}	X _{7i}
	HDP	PO	PN	PZ	MI	SEE	PMSP	PP
	[mld USD]	[1000]	[1000]	[1000]	[%]	[1000 GWh]	[1000]	[USD/h]
Belgické království	524,80	11 203,99	416,35	4 485,00	1,20	84,78	520,70	6,42
Bulharská republika	54,48	7 245,68	338,13	3 687,24	0,40	28,30	302,19	3,54
Česká republika	208,80	10 512,42	367,53	4 846,00	1,40	59,26	1 007,12	5,23
Dánské království	335,90	5 627,24	201,92	2 622,00	0,50	32,70	212,87	3,29
Estonská republika	24,88	1 315,82	57,96	599,00	3,20	7,43	60,38	0,67
Finská republika	267,30	5 451,27	219,00	2 404,00	2,20	83,09	222,03	3,02
Francouzská republika	2 806,00	65 856,61	2 818,05	25 508,00	1,00	460,90	2 598,02	36,33
Chorvatská republika	57,87	4 246,70	339,74	2 019,54	2,30	18,87	145,90	2,16
Irská republika	232,10	4 604,03	299,10	1 818,00	0,50	26,10	144,37	2,61
Italská republika	2 149,00	60 782,67	3 105,06	21 985,00	4,10	309,90	3 718,24	32,99
Kyperská republika	21,91	858,00	71,50	405,17	-0,40	3,22	40,80	0,44
Litva	45,93	2 943,47	148,81	1 486,45	0,40	10,30	134,11	1,53
Lotyšská republika	30,96	2 001,47	76,72	1 035,20	1,20	6,22	87,62	0,98
Lucembursko	60,13	549,68	14,83	236,00	0,00	6,45	29,39	0,31
Maďarsko	133,40	9 879,00	448,43	3 906,00	1,70	42,57	525,92	4,97
Maltská republika	9,64	425,38	16,54	219,78	1,70	1,99	30,49	0,23
Spol. republika Německo	3 730,00	80 780,00	2 259,00	39 529,00	1,00	544,50	2 201,14	47,92
Nizozemské království	853,50	16 829,29	590,27	8 184,00	1,60	112,50	802,09	11,29
Polská republika	525,90	38 495,66	1 788,10	15 313,00	2,00	132,20	1 474,95	19,40
Portugalská republika	227,30	10 427,30	851,80	4 158,00	0,40	48,27	774,83	5,29
Rakouská republika	428,30	8 507,79	214,95	4 099,00	0,40	65,57	307,33	4,59
Rumunsko	189,60	19 942,64	808,78	10 270,46	2,10	51,46	426,30	9,89
Řecká republika	242,20	10 992,59	1 348,08	3 560,00	3,20	59,53	653,94	5,68
Slovenská republika	97,71	5 415,95	385,50	2 318,00	1,00	28,76	391,80	2,68

Slovenská republika	47,96	2 061,09	101,81	888,00	-0,90	14,70	115,60	1,04
Velká Británie	2 678,00	64 308,26	2 438,13	28 709,00	7,50	344,70	1 724,62	35,05
Španělské království	1 393,00	46 507,76	6 041,81	17 002,00	1,50	267,50	2 252,78	24,38
Švédské království	579,70	9 644,86	407,60	4 553,00	1,50	132,10	665,82	5,56

Obrázek 4.1.5.1 Data pro analýzu

HDP ... hrubý domácí produkt pro rok 2013 v miliardách USD

PO ... počet obyvatel pro rok 2013 v 1000

PN ... počet nezaměstnaných pro rok 2013 v 1000

PZ ... počet zaměstnaných pro rok 2013 v 1000

MI ... míra inflace pro rok 2013 v %

SEE ... spotřeba elektrické energie pro rok 2013 v 1000 GWh

PMSP ... počet malých a středních podniků pro rok 2013 v 1000

PP ... produktivita práce pro rok 2013 v USD/h

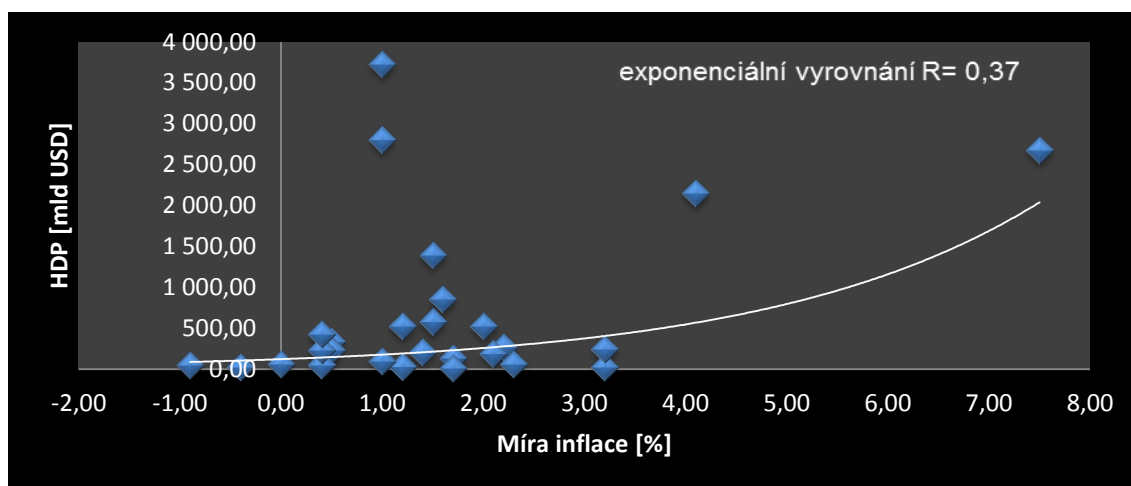
4.2.2 Metoda linearity

Požadavek na linearitu nenaplnuje pouze průběh závislosti mezi lineárně závislou y_i a 4. regresorem x_4 , MI, viz *Obrázek 4.2.2.1*. Průběh vztahu mezi těmito dvěma proměnnými jsem dále vyrovnal parabolou, mocninou a exponenciální křivkou. Ani v jednom případě korelační index nepřesáhl hodnotu 0,5, viz *Obrázek 4.2.2.2* a *Obrázek 4.2.2.3*. Nebudu proto data pro tento regresor transformovat a regresor vyloučím z regresního odhadu. Nejvyšší párový korelační koeficient má se závisle proměnnou regresor SEE, ten nyní podrobíme analýze multikolinearity. Výpočet párových korelačních koeficientů r_{y_i} , r_{ij} zpustíme na listě “korelace”.

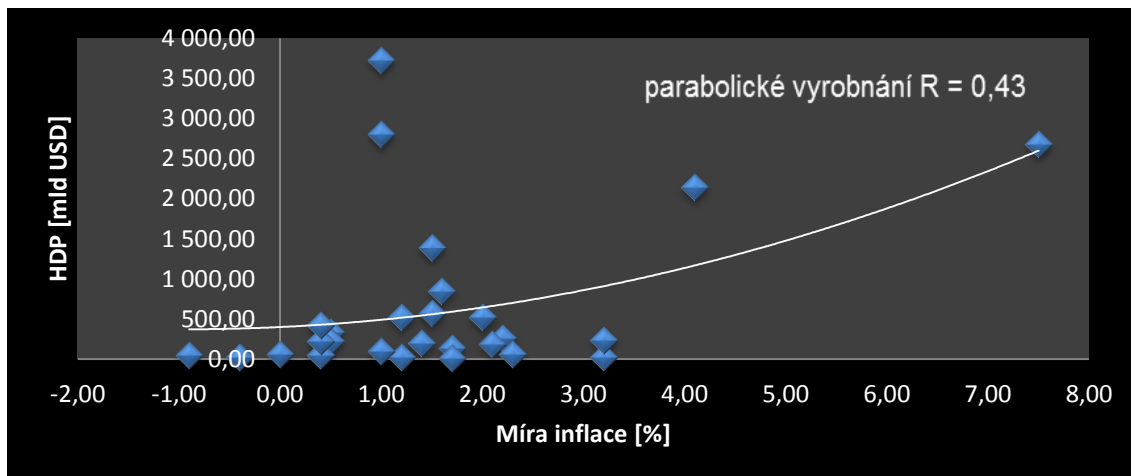
Výsledné párové koeficienty

	HDP	PO	PN	PZ	MI	SEE	PMSP	PP
HDP	1,00							
PO	0,96	1,00						
PN	0,68	0,79	1,00					
PZ	0,97	0,99	0,72	1,00				
MI	0,40	0,46	0,36	0,44	1,00			
SEE	0,99	0,97	0,73	0,97	0,36	1,00		
PMSP	0,83	0,91	0,83	0,84	0,42	0,86	1,00	
PP	0,97	1,00	0,76	0,99	0,44	0,98	0,89	1,00

Obrázek 4.2.2.1 Tabulka párových korelačních koeficientů r_{ij} a r_{y_i}



Obrázek 4.2.2.2 Vyrovnání exponenciálou



Obrázek 4.2.2.3 Vyrovnání parabolou

4.2.3 Metoda multikolinelarity

Regresor SEE je ve velké lineární závislosti s regresory PO, PZ, PP, viz *Obrázek 4.2.3.1*. Závislost mezi těmito dvojicemi nezávisle proměnných je tak vysoká, že užití regresorů PO, PZ, PP by pouze zatěžovalo náš regresní odhad. Odstráním tedy i tyto regresory. Zbývající regresory podrobím analýze Stepwise.

Výsledné párové koeficienty						
	PO	PN	PZ	SEE	PMSP	PP
SEE	0,97	0,73	0,97	1,00	0,86	0,98

Obrázek 4.2.3.1 Párové regresní koeficienty r_{ij} s SEE

Na listě “*Vstupní oblast dat*“ odstráním hodnoty pro regresory, které jsem vyřadil v předchozích krocích. Zbudou 4 proměnné. Jejich hodnoty zapíši do vstupní tabulky na listě “*Vstupní oblast dat*“ v pořadí:

	Název proměnné	Jednotky []
1	HDP ... y_i	[mld USD]
2	Spotřeba elektrické energie ... x_1	[1000 GWh]
3	Počet nezaměstnaných ... x_2	[1000]
4	Počet MSP ... x_3	[1000]

Tabulka 4.2.3.1 Zbývající proměnné

Poté změním názvy proměnných a jednotek, viz *Obrázek 4.2.3.2*.

Vícenásobná regrese

Zadejte názvy proměnných:

Závisle proměnná	Jednotky []
HDP	mld USD
1.regresor: SEE	1000 GWh
2.regresor: PN	1000
3.regresor: PMSP	1000
4.regresor: []	[]
5.regresor: []	[]
6.regresor: []	[]
7.regresor: []	[]
8.regresor: []	[]
9.regresor: []	[]
10.regresor: []	[]

OK

Obrázek 4.2.3.2 Zadání názvů proměnných

Na listě “2.Regrese“ zpustím výpočet regresního a korelačního úkonu dvojnásobné RKA pro proměnné:

	Název proměnné	Jednotky []
1	HDP ... y_i	[mld USD]
2	Spotřeba elektrické energie ... x_1	[1000 GWh]
3	Počet nezaměstnaných ... x_2	[1000]

Tabulka 4.2.3.2 Proměnné pro dvojnásobnou RKA

V tomto kroku si nebudu všimnout výsledků regresního úkonu. Zaměřím pozornost pouze na výsledky hodnot koeficientů spolehlivosti a testu významnosti regresního modelu, viz *Obrázek 4.2.3.3* a *Obrázek 4.2.3.4*.

Regresní statistika	
Násobné R	0,9659329
Hodnota spolehlivosti R	0,933026368
Nastavená hodnota spolehlivosti R	0,927668478
Chyba stř. hodnoty	265,3016755
Pozorování	28

Obrázek 4.2.3.3 Koefficienty spolehlivosti pro dvojnásobnou R

Test významnosti regresního modelu:			
H0 - vícenásobný lineární regresní model není významný ($R^2 = 0$)			$F < F_{0,95}(2, 25)$
H1 - vícenásobný lineární regresní model je významný ($R^2 \neq 0$)			$F > F_{0,95}(2, 25)$
determinační koeficient	0,93	93,3%	
korelační koeficient	0,97	96,6%	
testové kritérium	F	✓ 174,14	
kritická hodnota	$F_{0,95}(2, 25)$	3,385	

Obrázek 4.2.3.4 Test významnosti regresního modelu pro dvojnásobnou RKA

Na listě "3.Regrese" zpustím výpočet regresního a korelačního úkonu trojnásobné RKA pro proměnné:

	Název proměnné	Jednotky []
1	HDP... y_i	[mld USD]
2	Spotřeba elektrické energie ... x_1	[1000 GWh]
3	Počet nezaměstnaných ... x_2	[1000]
4	Počet MSP ... x_3	[1000]

Tabulka 4.2.3.3 Proměnné pro trojnásobnou RKA

Znovu zaměřím pozornost na výsledky hodnot koeficientů spolehlivosti a testu regresního modelu, viz Obrázek 4.2.3.5 a Obrázek 4.2.3.6.

Regresní statistika	
Násobné R	0,988282072
Hodnota spolehlivosti R	0,976701453
Nastavená hodnota spolehlivosti R	0,973789135
Chyba stř. hodnoty	159,7043953
Pozorování	28

Obrázek 4.2.3. 5 Koeficienty spolehlivosti pro trojnásobnou RK

Test významnosti regresního modelu:			
H0 - vícenásobný lineární regresní model není významný ($R^2 = 0$)			$F < F_{0,95}(3, 24)$
H1 - vícenásobný lineární regresní model je významný ($R^2 \neq 0$)			$F > F_{0,95}(3, 24)$
determinační koeficient	0,977	97,7%	
korelační koeficient	0,988	98,8%	
testové kritérium	F	✓ 335,37	
kritická hodnota	$F_{0,95}(3, 24)$	3,01	

Obrázek 4.2.3.6 Test významnosti regresního modelu pro trojnásobnou RKA

Přidání regresoru PMSP zvýšilo výslednou hodnotu souhrnného korelačního koeficientu o více než 2 procenta. Nastavená hodnota spolehlivosti R , neboli adjustovaný korelační koeficient, také vzrostl. Považujeme tedy vliv přidání regresoru PMSP za statisticky významný. K dalšímu zkoumání jdou tedy tyto 4 proměnné (y_i a x_1, x_2, x_3):

	Název proměnné	Jednotky []
1	HDP ... y_i	[mld USD]
2	Spotřeba elektrické energie ... x_1	[1000 GWh]
3	Počet nezaměstnaných ... x_2	[1000]
4	Počet MSP ... x_3	[1000]

Tabulka 4.2.3.4 Proměnné k dalšímu zkoumání

Toto srovnání můžeme provést i přes list “V.regrese”.

4.2.4 Grubbsův test

Na listě “Grubbsův test“ aplikace hlásí u všech tří regresorů odlehle hodnoty, viz Obrázek 4.2.4.1 a Obrázek 4.2.4.2 a Obrázek 4.2.4.3.

Testové kritérium	Kritické hodnoty $\alpha = 0,05$
$T_{(1)} =$ ✘ 3,093	$T_{28(28,0,05)\alpha}$ 2,753
$T_{(28)} =$ ✔ 0,739	2,753

Testové kritérium	Kritické hodnoty $\alpha = 0,01$
$T_{(1)} =$ ✔ 3,093	$T_{28(28,0,01)}$ 3,119
$T_{(28)} =$ ✔ 0,739	3,119

Obrázek 4.2.4.1 Grubbsův test pro proměnnou x_1

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,05$
	$T_{28(28,0,05)\alpha}$	
$T_{(1)} =$	$\times 3,871$	2,753
$T_{(28)} =$	$\checkmark 0,697$	2,753

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,01$
	$T_{28(28,0,01)}$	
$T_{(1)} =$	$\times 3,871$	3,119
$T_{(28)} =$	$\checkmark 0,697$	3,119

Obrázek 4.2.4.2 Grubbsův test pro proměnnou x_2

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,05$
	$T_{28(28,0,05)\alpha}$	
$T_{(1)} =$	$\times 3,224$	2,753
$T_{(28)} =$	$\checkmark 0,810$	2,753

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,01$
	$T_{28(28,0,01)}$	
$T_{(1)} =$	$\times 3,224$	3,119
$T_{(28)} =$	$\checkmark 0,810$	3,119

Obrázek 4.2.4.3 Grubbsův test pro proměnnou x_3

Postupně jsem odstraňoval hodnoty jednotlivých regresorů vzestupně a opakovaně prováděl Grubbsův test na hladině významnosti $\alpha = 0,05$. V *Tabulce 4.2.4.1* jsou uvedeny červeně hodnoty, které testem neprošly a bíle hodnoty, které prošly. Země EU alespoň s jednou červenou hodnou jsem odstranil z modelu. Zbylo mi 11 zemí, viz *Tabulka 4.2.4.2*.

Země EU	X_{1i}	X_{2i}	X_{3i}
	SEE	PN	PMSP
	[1000 GWh]	[1000]	[1000]
Belgické království	84,78	416,35	520,70
Bulharská republika	28,30	338,13	302,19
Česká republika	59,26	367,53	1 007,12
Dánské království	32,70	201,92	212,87
Estonská republika	7,43	57,96	60,38
Finská republika	83,09	219,00	222,03
Francouzská republika	460,90	2 818,05	2 598,02
Chorvatská republika	18,87	339,74	145,90
Irská republika	26,10	299,10	144,37
Italská republika	309,90	3 105,06	3 718,24
Kyperská republika	3,22	71,50	40,80
Litva	10,30	148,81	134,11
Lotyšská republika	6,22	76,72	87,62
Lucembursko	6,45	14,83	29,39
Maďarsko	42,57	448,43	525,92
Maltská republika	1,99	16,54	30,49
Spolková republika Německo	544,50	2 259,00	2 201,14
Nizozemské království	112,50	590,27	802,09
Polská republika	132,20	1 788,10	1 474,95
Portugalská republika	48,27	851,80	774,83
Rakouská republika	65,57	214,95	307,33
Rumunsko	51,46	808,78	426,30
Řecká republika	59,53	1 348,08	653,94
Slovenská republika	28,76	385,50	391,80
Slovinská republika	14,70	101,81	115,60
Velká Británie	344,70	2 438,13	1 724,62
Španělské království	267,50	6 041,81	2 252,78

Švédské království	132,10	407,60	665,82
--------------------	--------	--------	--------

Tabulka 4.2.4.1 Hodnoty, které neprošli testem

Země EU	y_i	x_{1i}	x_{2i}	x_{3i}
	HDP	SEE	PN	PMSP
	[mld USD]	[1000 GWh]	[1000]	[1000]
Francouzská republika	2 806,00	460,90	2 818,05	2 598,02
Italská republika	2 149,00	309,90	3 105,06	3 718,24
Maďarsko	133,40	42,57	448,43	525,92
Spolková republika Německo	3 730,00	544,50	2 259,00	2 201,14
Nizozemské království	853,50	112,50	590,27	802,09
Polská republika	525,90	132,20	1 788,10	1 474,95
Portugalská republika	227,30	48,27	851,80	774,83
Rumunsko	428,30	51,46	808,78	426,30
Řecká republika	242,20	59,53	1 348,08	653,94
Velká Británie	2 678,00	344,70	2 438,13	1 724,62
Španělské království	1 393,00	267,50	6 041,81	2 252,78

Tabulka 4.2.4.2 Hodnoty, které prošly testem

4.2.5 Analýza reziduí

Na listě “Analýza reziduí“ jsem zkontroloval pro jednotlivé dvojice proměnných, vždy závisle proměnná y_i a některý z regresorů x_1, x_2, x_3 testy o náhodnosti reziduí, viz Obrázek 4.2.5.1 a Obrázek 4.2.5.2 a Obrázek 4.2.5.3. Před samotnou analýzou reziduí jsem odstranil na listě “Vstupní oblast dat“ ty odlehle hodnoty, které neprošli předchozím testem.

y_i a x_1	
<u>Testové kritérium</u>	<u>Kritické hodnoty</u> $\alpha = 0,05$
U =	1,960
✓0,671	

Obrázek 4.2.5.1 Analýza reziduí pro y_i a x_1

y_i a x_2

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,05$
U =	-0,224	1,960

Obrázek 4.2.5.3 Analýza reziduí pro y_i a x_2

y_i a x_3

<u>Testové kritérium</u>	<u>Kritické hodnoty</u>	$\alpha = 0,05$
U =	✓0,224	1,960

Obrázek 4.2.5.3 Analýza reziduí pro y_i a x_3

Testy prošli všechny dvojice proměnných. Regresní model tedy bude obsahovat vždy 11 hodnot pro každou ze 4 proměnných (y_i a x_1, x_2, x_3):

	Název proměnné	Jednotky []
1	HDP ... y_i	[mld USD]
2	Spotřeba elektrické energie ... x_1	[1000 GWh]
3	Počet nezaměstnaných ... x_2	[1000]
4	Počet MSP ... x_3	[1000]

Tabulka 4.2.5.3 Proměnné k RKA

4.2.6 Regresní a korelační úkon

Zbyly mi pouze 3 regresory, provádím tedy trojnásobnou RKA. Mohl jsem si vybrat, jestli výpočet zpusťím na listě “3.regrese“, kde je postup výpočtu rozepsán dopodrobna podle 1. kapitoly diplomové práce, nebo na listě “V.regrese“. Aplikace nám vygeneruje koeficienty beta, regresní koeficienty, regresní rovnici, souhrnný korelační koeficient, determinační koeficient, adjustovaný determinační koeficient, vyrovnané hodnoty, analýzu rozptylu a testy statistické významnosti regresního modelu.

Výsledky regresního a korelačního úkonu:

1. Koeficienty beta

β_1	1,03561
β_2	-0,11518
β_3	0,013364

Tabulka 4.2.6.1 Koeficienty beta

2. Regresní koeficienty a souhrný korelační koeficient

$a_{y,123}$	-30,346
$b_{y1,23}$	7,266
$b_{y2,13}$	-0,090
$b_{y3,12}$	0,016
$r_{y,123 \dots k}$	0,9876

Tabulka 4.2.6.2 Regresní koeficienty a korel. koeficient

3. Regresní rovnice

$$\hat{y}_i = -30,346 + 7,266x_1 - 0,090x_2 + 0,016x_3$$

4. Analýza rozptylu (Anova) + kvantily normálního rozdělení

ANOVA					
	Rozdíl	SS	MS	F	Významnost F
Regrese	3	15530986,53	5176995,509	92,67169419	5,35351E-06
Rezidua	7	391046,7902	55863,82718		
Celkem	10	15922033,32			

Tabulka 4.2.6.3 Analýza rozptylu

	Koeficienty	Chyba stř. hodnoty	Dolní 95%	Horní 95%	Dolní 95,0%	Horní 95,0%
Hranice	-30,34633167	134,2670483	-347,8374502	287,1447869	-347,8374502	287,1447869
Soubor X 1	7,266239428	0,648329486	5,733183803	8,799295054	5,733183803	8,799295054
Soubor X 2	-0,090023657	0,064519773	-0,242588678	0,062541363	-0,242588678	0,062541363
Soubor X 3	0,016058422	0,129896799	-0,291098699	0,323215543	-0,291098699	0,323215543

Tabulka 4.2.6.4 Pravděpodobnostní kvantily NR

5. Test významnosti regresního modelu

Test významnosti regresního modelu:			
H0 - vícenásobný lineární regresní model není významný ($R^2 = 0$)			$F < F_{0,95}(3, 7)$
H1 - vícenásobný lineární regresní model je významný ($R^2 \neq 0$)			$F > F_{0,95}(3, 7)$
determinační koeficient	0,975	97,5%	
korelační koeficient	0,988	98,8%	
testové kritérium	F	92,67	
kritická hodnota	$F_{0,95}(3, 7)$	4,35	

Tabulka 4.2.6.5 Test významnosti regresního modelu

6. Vyrovnané hodnoty

Země EU	y_i	\hat{y}_i
	HDP	HDP
	[mld USD]	[mld USD]
Francouzská republika	2 806,00	3 106,69
Italská republika	2 149,00	2 001,64
Maďarsko	133,40	247,05
Spol. republika Německo	3 730,00	3 758,10
Nizozemské království	853,50	746,85

Polská republika	525,90	792,96
Portugalská republika	227,30	256,16
Rumunsko	428,30	277,61
Řecká republika	242,20	291,36
Velká Británie	2 678,00	2 282,53
Španělské království	1 393,00	1 405,64
Suma	15 166,60	15 166,60

Tabulka 4.2.6.6 Vyrovnané hodnoty

5 Vyhodnocení RKA

Z původně sedmi vybraných regresorů, zbyly pouze tři pro samotnou regresní a korelační analýzu. Odstraněné regresory by pouze zatěžovaly náš regresní odhad. Z výsledku je patrné, že výsledná hodnota hrubého domácího produktu je ve velké závislosti s hodnotou spotřebované elektrické energie. Mezi těmito veličinami je vysoká souvztažnost. Vliv zbylých regresorů je nepatrný. Pouze u regresorů počet nezaměstnaných a počet malých a středních podniků se dá vliv považovat za statisticky významný.

Výsledná hodnota souhrnného korelačního koeficientu 0,9876 značí vysokou těsnost korelační závislosti hrubého domácího produktu na zbylých třech regresorech (spotřeba elektrické energie, počet nezaměstnaných, počet malých a středních podniků). Očekávám tedy vysokou spolehlivost mého regresního odhadu.

I když je spolehlivost regresního odhadu vysoká, nepředpokládám, že by se regresní rovnice dala použít ke skutečnému prognózování hodnot HDP zemí EU. Když porovnáme výsledné vyrovnané hodnoty HDP s těmi skutečnými, vidíme, že rozdíly nejsou nepatrné. Rozdíly jsou v řádech desítek až stovek miliard USD. To nelze zanedbat. Nicméně rovnice může posloužit k hlubší analýze jednotlivých proměnných v závislosti na působení daných teritoriálních vlivů prostředí.

6 Závěr

Hlavním cílem práce bylo vytvoření aplikace, která uživatelům usnadní práci v procesu hledání ideálního počtu veličin pro daný regresní model a následné provedení regresní a korelační analýzy. K tomu poslouží mou vytvořená aplikace [PŘ1].

V práci je nastolen metodický postup zkoumání a podrobné popsání, jak tento postup praktikovat v aplikaci [PŘ1]. Funkčnost aplikace byla prokázána na zkoumaném příkladu.

Publikace

- [P1] Karel Zvára: Regresní analýza. Praha, Academia Praha 1989
- [P2] Jan Kožíšek – Barbora Stieberová: Management jakosti 1 a 2. Praha, ČVUT 2010
- [P3] Pavel Hebák – Josef Hustopecký: Vícerozměrné statistické metody s aplikacemi. Praha, STNL 1989
- [P4] Jan Kožíšek: Statistika. Praha, ČVUT 2006
- [P5] Lucie Šimonová: Průvodce k programu STRATHGRAPHICS. Ostrava, VŠB 2005

Přílohy (CD)

- [PŘ1] DP_APLIKACE.XSLM

Seznam obrázků

<i>Obrázek 4.1.1.1 Zápis dat</i>	38
<i>Obrázek 4.1.1.2 Uživatelský formulář názvy proměnných</i>	39
<i>Obrázek 4.1.1.2 Uživatelská lišta</i>	39
<i>Obrázek 4.1.2.1 Počet měření</i>	40
<i>Obrázek 4.1.2.2 Počet proměnných</i>	40
<i>Obrázek 4.1.2.3 Výsledky párových korelačních koeficientů r_{yj} a r_{ij}</i>	40
<i>Obrázek 4.1.3.1 Adjustovaný determinační koeficient</i>	41

<i>Obrázek 4.1.3.2 Analýza Anova</i>	42
<i>Obrázek 4.1.3.2 Test významnosti regresního modelu</i>	42
<i>Obrázek 4.1.4.1 Grubbsův test</i>	42
<i>Obrázek 4.1.5.1 Analýza reziduí</i>	43
<i>Obrázek 4.1.5.1 Data pro analýzu</i>	45-46
<i>Obrázek 4.2.2.2 Vyrovnání exponenciálou</i>	47
<i>Obrázek 4.2.2.3 Vyrovnání parabololy</i>	48
<i>Obrázek 4.2.3.1 Párové regresní koeficienty r_{ij} s SEE</i>	48
<i>Obrázek 4.2.3.2 Zadání názvů proměnných</i>	49
<i>Obrázek 4.2.3.3 Koeficienty spolehlivosti pro dvojnásobnou R</i>	50
<i>Obrázek 4.2.3.4 Test významnosti regresního modelu pro dvojnásobnou RKA</i>	50
<i>Obrázek 4.2.3.5 Koeficienty spolehlivosti pro trojnásobnou RK</i>	51
<i>Obrázek 4.2.3.6 Test významnosti regresního modelu pro trojnásobnou RKA</i>	51
<i>Obrázek 4.2.4.1 Grubbsův test pro proměnnou x_1</i>	52
<i>Obrázek 4.2.4.2 Grubbsův test pro proměnnou x_2</i>	53
<i>Obrázek 4.2.4.3 Grubbsův test pro proměnnou x_3</i>	53
<i>Obrázek 4.2.5.1 Analýza reziduí pro y_i a x_1</i>	55
<i>Obrázek 4.2.5.3 Analýza reziduí pro y_i a x_2</i>	56
<i>Obrázek 4.2.5.3 Analýza reziduí pro y_i a x_3</i>	56

Seznam tabulek

<i>Tabulka 2.3.1 Vliv volby typu regresní přímky</i>	14
<i>Tabulka 2.2.1 Regresory $x_1, x_2 \dots x_7$</i>	26
<i>Tabulka 2.2.6.1 Průměrných odpracovaných hodin týdně a počtu dní bez práce</i>	29
<i>Tabulka 3.1.3.1. Kritických hodnot Grubbsova testu</i>	36
<i>Tabulka 4.2.3.1 Zbylé proměnné</i>	48
<i>Tabulka 4.2.3.2 Proměnné pro dvojnásobnou RKA</i>	49
<i>Tabulka 4.2.3.3 Proměnné pro trojnásobnou RKA</i>	51
<i>Tabulka 4.2.6.1 Koeficienty beta</i>	57
<i>Tabulka 4.2.6.2 Regresní koeficienty a korel. koeficient</i>	57
<i>Tabulka 4.2.6.3 Analýza rozptylu</i>	57
<i>Tabulka 4.2.6.4 Pravděpodobnostní kvantily NR</i>	58
<i>Tabulka 4.2.6.5 Test významnosti regresního modelu</i>	59
<i>Tabulka 4.2.6.6 Vyrovnané hodnoty</i>	59-60