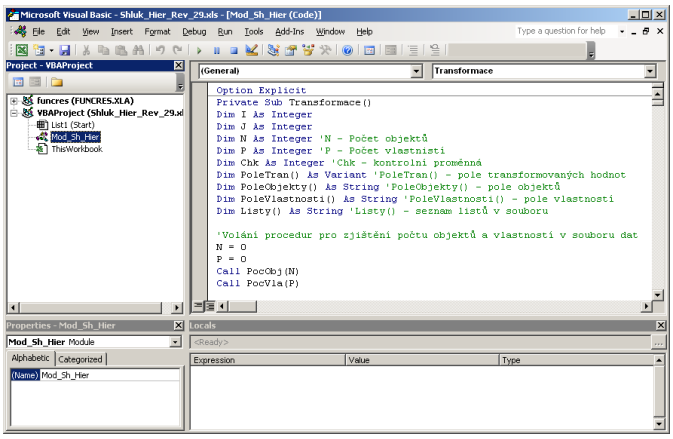
# NADSTAVBOVÝ MODUL MOHSA V1

Nadstavbový modul pro hierarchické shlukování se jmenuje Mod_Sh_Hier (MOHSA V1) je součástí souboru Shluk_Hier.xls. Tento soubor je přístupný na http://jonasova.upce.cz, a je určen pro nekomerční účely, výuku a jako podpora pro zpracování studentských závěrečných prací.




Obrázek 1 Vývojové prostředí VBA [zdroj: autor]

Nadstavbový modul obsahuje procedury realizující hierarchické shlukování na základě událostí podle požadavků uživatele. Těmito událostmi jsou nejčastěji kliknutí na tlačítko v listu. Například kliknutím na tlačítko „Kvantitativní data – Transformace dat“ se spustí procedura *Transformace()*. VBA explicitně nevyžaduje deklaraci proměnných, ale jejich vyžadování je možné vynutit. Protože, podle mého názoru, nedeklarování proměnných vnáší do programového kódu chyby a zmatek, je deklarování proměnných vynuceno příkazem Option Explicit (viz obrázek 1, řádek 1). Každé důležité místo nebo každý dílčí celek v programovém kódu a také každá proměnná, která má nějaký specifický význam, jsou v modulu popsány komentáři, tj. zkrácenými popisky zelené barvy, které stručně vysvětlují význam daného kódu nebo proměnné (viz obrázek 1). Nadstavbový modul je navržen tak, že zpracovává dva druhy dat – kvantitativní a dichotomická data.

V případě kvantitativních dat je postup následující:

- transformace dat – vznikne list s názvem „Transformace“
- výpočet vzdáleností - vznikne list s názvem „Vzdálenosti“
- shlukování - vznikne list s názvem „Shlukování“
- vytvoření dendrogramu - vznikne list s názvem „Dendrogram“

## 1 Popis procedur a algoritmů nadstavbového modulu

V následujících podkapitolách jsou popsány procedury a algoritmy nadstavbového modulu pro hierarchické shlukování a to zvlášť pro zpracování kvantitativních dat a zvlášť pro zpracování dichotomických dat.

## 1.1 Předzpracování kvantitativních dat

**1. krok – transformace dat**

V nadstavbovém modulu je tato činnost ošetřena procedurou s názvem *Transformace()*. Tato procedura je spuštěna kliknutím na tlačítko „Kvantitativní data – Transformace dat“. Po úspěšné transformaci dat je vytvořen list s názvem „Transformace“, kde jsou uložena transformovaná data. Před spuštěním procedury transformace dat musí uživatel zvolit způsob transformace pomocí roletového menu napravo od tlačítka pro spuštění transformace. Nadstavbový modul nabízí tyto možnosti transformace dat:

- kontrola dat,
- standardizace,
- normalizace,
- standardizace + normalizace,
- normalizace + standardizace.

V případě volby „kontroly dat“ jsou data pouze zkontrolována a do listu „Transformace“ nakopírována bez úpravy. Tato volba je zde proto, aby uživatel měl možnost provést hierarchické shlukování i takových dat, která má již předzpracovaná jiným způsobem, například jiným SW.

***Popis procedury transformace dat:***

Název procedury: *Transformace()*

Výpočty v proceduře Procedura volá proceduru *PocObj(N)*, která vrací počet objektů vstup. tabulky – proměnná N. Procedura volá proceduru *PocVla(P)*, jenž vrací počet vlastností vstup. tabulky – proměnná P. Procedura volá proceduru *Kontrola(N, P, Chk, PoleTran())*, které jsou předány proměnné *N* a *P*, a tato procedura vrací informaci o výsledku kontroly dat – proměnnou *Chk* a pole *PoleTran()* s načtenými a zkontrolovanými hodnotami ze vstupní tabulky.

Procedura dále podle volby uživatele volá proceduru *Standardizace(N, P, PoleTran())*, nebo *Normalizace (N, P, PoleTran())*, a tyto procedury vrací upravené pole *PoleTran()* se standardizovanými, nebo normalizovanými hodnotami.

Výstup procedury:

Procedura vytvoří list s názvem „Zkratky“ s tabulkou zkratek objektů a vlastností. Procedura vytvoří list s názvem „Transformace“ s tabulkou transformovaných hodnot.

Při zpracování jak kvantitativních tak dichotomických dat je vždy na začátku prvního kroku prováděna kontrola dat. U kvantitativních dat je prováděna kontrola zda tabulka neobsahuje chybějící hodnoty a zda je obsah buněk číslo, u dichotomických dat je prováděna kontrola zda tabulka obsahuje pouze číselné hodnoty 0 a 1. V opačném případě program dál nepokračuje a uživatel je upozorněn na chybu v datech.

Protože názvy objektů a vlastností jsou často dlouhé textové řetězce, navíc různé délky, jsou v průběhu programu používány místo těchto názvů jejich zkratky. Objekty jsou značeny O1, O2, …, On, kde n je počet objektů a vlastnosti jsou značeny VL1, VL2, …, VLp, kde p je počet vlastností. Tabulka zkratek přiřazených jednotlivým objektům a vlastnostem je uložena na listu s názvem „Zkratky“, který je vygenerován vždy po prvním kroku zpracování dat.

**2. krok – výpočet vzdáleností**

V nadstavbovém modulu je tato činnost ošetřena procedurou s názvem *Vzdalenost()*. Tato procedura je spuštěna kliknutím na tlačítko „Vzdálenost“ které je umístěno na listu „Transformace“. Po výpočtu vzdálenosti je vytvořen list s názvem „Vzdálenosti“ s tabulkou vzdáleností mezi objekty. Před spuštěním procedury výpočtu vzdáleností musí uživatel zvolit způsob výpočtu vzdáleností mezi objekty pomocí roletového menu napravo od tlačítka „Vzdálenost“.

Nadstavbový modul nabízí tyto možnosti výpočtu vzdáleností:

- Euklidovská vzdálenost,
- Čtvercová Euklidovská vzdálenost,
- Hammingova vzdálenost,
- Čebyševova vzdálenost.

***Popis procedury výpočtu vzdáleností:***

Název procedury: *Vzdalenost ()*

Vstupní parametry procedury: *N* – počet objektů *P* – počet vlastností

Výpočty v proceduře Procedura podle volby uživatele volá proceduru *EuklVzdal(N, P, PoleVzdal()), EuklCtver(N, P, PoleVzdal()), HammVzdal(N, P, PoleVzdal())* nebo *CebysVzdal(N, P, PoleVzdal())*, kterým jsou předány proměnné *N* a *P* a pole *PoleVzdal()*, a tyto procedury vrací pole *PoleVzdal()* s hodnotami vzdáleností mezi objekty.

Výstup procedury:

Procedura vytvoří list s názvem „Vzdálenosti“ s tabulkou vzdáleností mezi objekty.

**3. krok – shlukování**

Nadstavbový modul provádí hierarchické shlukování objektů aglomerativní metodou. Aglomerativní přístup je charakteristický tím, že se vychází od jednotlivých objektů, tj. na začátku algoritmu jsou jednotlivé objekty brány jako shluk s jedním objektem. Jejich postupným slučováním se buduje hierarchický systém, až se dospěje ke konečnému sloučení všech objektů do jednoho shluku. Toto slučování se provádí tak, že se v jednotlivých iteracích spojují dva shluky s nejmenší vzájemnou vzdáleností. V nadstavbovém modulu je tato činnost shlukování ošetřena procedurou s názvem *Shluky()*.

Proces shlukování se odstartuje tlačítkem „Shlukování“ v listu „Vzdálenosti“ a výstupem z tohoto procesu je tabulka, ze které je zřejmé jaké shluky byly spojeny v jednotlivých krocích a při jaké vzdálenosti ke sloučení došlo. Před spuštěním procedury shlukování musí uživatel zvolit metodu, která

bude použita pro výpočet vzdáleností mezi shluky. Tato volba se provádí pomocí roletového menu napravo od tlačítka „Shlukování“.

Nadstavbový modul nabízí tyto možnosti výpočtu vzdáleností mezi shluky:

- metoda nejbližšího souseda,
- metoda nejvzdálenějšího souseda,
- metoda průměrné vzdálenosti,
- centroidní metoda,
- mediánová metoda,
- Wardova-Wishartova metoda.

Postup shlukování aglomerativní metodou se skládá z několika úkonů, které jsou sami o sobě natolik složité, že kdyby se prováděli všechny v jedné proceduře, tak by tato procedura byla velmi nepřehledná. Proto samotná procedura *Shluky()* nevykonává žádné výpočty, ale „pouze“ zajišťuje volání dalších procedur a výpis hodnot do listů. Procedury, které jsou procedurou *Shluky()* volány jsou *PocObjShl()*, *MinHodnota()*, *LegTab()* a *ShVypocet()*.

Postup shlukování lze zjednodušeně popsat takto:

Určení shluků, které budou v daném kroku sloučeny na základě minimální hodnoty. Toto zajišťuje procedura *MinHodnota()*. Sloučení shluků a přepočítání tabulky vzdáleností. Toto zajišťuje procedura *ShVypocet()*. Pro výpočet vzdálenosti mezi nově vzniklým shlukem a ostatními shluky je použita Lance-Williams formule, ve které se podle zvolené metody přepočítávají koeficienty , , , $_{A\ B}$ . Pro výpočet těchto koeficientů je potřeba mít informaci o počtu objektů v jednotlivých shlucích v daném kroku. Tuto informaci zajišťuje procedura *PocObjShl()*.

Přepočítané tabulky vzdáleností v jednotlivých krocích se vkládají do listu „Vzdálenosti“ pod sebe. Aby byli dobře čitelné, je potřeba vytvořit legendu pro jednotlivé řádky a sloupce. Tuto legendu zajišťuje procedura *LegTab()*.

***Popis procedury shlukování:***

Název procedury: *Shluky ()*

Vstupní parametry procedury:

*N* – počet objektů

Tabulka vzdáleností mezi objekty

Výpočty v proceduře :

Procedura nevykonává žádné výpočty, ale zajišťuje volání dalších procedur, které potřebné výpočty provádějí. Volané procedury jsou *PocObjShl()*, *MinHodnota()*, *LegTab()* a *ShVypocet*.

Výstup procedury:

Procedura vytvoří list s názvem „Shlukování“ s výstupní tabulkou po shlukování.

Procedura vytvoří list s názvem „Tabulka vzdáleností“ s výstupní tabulkou po shlukování.

**4. krok – vytvoření dendrogramu**

Vytvoření dendrogramu se odstartuje tlačítkem „Vytvořit Dendrogram“ v listu „Shlukování“. Po kliknutí na toto tlačítko je vytvořen list s názvem „Dendrogram“ který obsahuje vytvořený dendrogram. Výstupem z procesu shlukování je tabulka, která dává přehled o tom, jaké shluky resp. objekty byly v každém kroku spojeny, a také informaci při jaké vzdálenosti spojení proběhlo. Tuto tabulku nelze použít pro vykreslení dendrogramu a musí být upravena. Úpravu tabulky pro vykreslení dendrogramu realizuje procedura s názvem *UpTabDend.*

***Vykreslení dendrogramu***

Vykreslení dendrogramu realizuje procedura Dendrogram() a je založeno na vkládání objektů typu Shape do listu „Dendrogram“. Příkaz *Worksheets("Dendrogram").Shapes.AddLine()* vloží linku a příkaz *Worksheets("Dendrogram").Shapes.AddLabel()* vloží text do listu Dendrogram. AddLine a AddLabel jsou metody objektu Shapes a v závorkách se uvádějí parametry - souřadnice vložení objektu a text, jenž má být vložen.

***Popis procedury vykreslení dendrogramu***

Název procedury: *Dendrogram ()*

Vstupní parametry procedury:

*M* – počet kroků shlukování.

Procedura volá proceduru *UpTabDend()*, která vrací upravenou výstupní tabulku z procesu shlukování. Výpočty v proceduře Výpočty počátečních a koncových souřadnic objektů Shape.

Výstup procedury:

Vykreslený dendrogram na list s názvem „Dendrogram“.

## 1.2 Zpracování dichotomických dat

**1. krok – vytvoření asociační tabulky**

V případě, že jsou objekty charakterizovány dichotomickými znaky, se ve shlukové analýze míra podobnosti nazývá koeficientem asociace. Při určování prvků v matici podobností objektů $O_i$ a $O_j$ bude pozorována shoda či neshoda výsledků u $p$ proměnných V nadstavbovém modulu je pro zahájení práce s dichotomickými daty určena procedura s názvem *StartDich()*. Tato procedura je spuštěna kliknutím na tlačítko „Dichotomická data – Asociační tabulka“. Po spuštění této procedury je provedena kontrola, zda jsou data skutečně dichotomické, tj. zda matice čísel obsahuje pouze 0 a 1. V případě, že jsou data pořádku, je vytvořen list s názvem „Asoc_Tab“, kde je uložena asociační tabulka.

***Popis procedury pro zahájení práce s dichotomickými daty:***

Název procedury: *StartDich()*

Vstupní parametry procedury:

Vstupní tabulka s dichotomickými daty.

Procedura volá proceduru *PocObj(N)*, která vrací počet objektů vstup tabulky – proměnná *N*. Procedura volá proceduru *PocVla(P)*, jenž vrací počet vlastností vstup. tabulky – proměnná *P*.
Výpočty v proceduře Procedura volá dále proceduru *AsocTab(N, P, AsocMat())*, která provádí vlastní výpočet asociační tabulky a vrací upravené třírozměrné pole *AsocMat()* s touto asociační tabulkou.
Výstup procedury:
Procedura vytvoří list s názvem „Zkratky“ s tabulkou zkratek objektů. Procedura vytvoří list s názvem „Asoc_Tab“ s asociační tabulkou.

**2. krok – výpočet vzdáleností**

Míry vzdálenosti pro dichotomická data. Výpočet vzdáleností je možné provést více způsoby. Tento nadstavbový modul realizuje výpočet tak, že provádí výpočet koeficientů asociace z asociační tabulky. Protože koeficienty asociace vyjadřují míru podobnosti, je pro proces shlukování potřeba míru podobnosti převést na míru nepodobnosti (funkce založené na vzdálenosti objektů jsou prakticky míry nepodobnosti). V nadstavbovém modulu je výpočet koeficientů asociace ošetřen procedurou s názvem *KoefAsoc()*. Tato procedura je spuštěna kliknutím na tlačítko „Výpočet koeficientů asociace“ které je umístěno na listu „Asoc_tab“. Po výpočtu koeficientů je vytvořen list s názvem „Koef_asociace“ s tabulkou koeficientů asociace. Před spuštěním procedury výpočtu koeficientů asociace musí uživatel zvolit způsob jejich výpočtu pomocí roletového menu napravo od tlačítka „Výpočet koeficientů asociace“. Nadstavbový modul nabízí tyto možnosti výpočtu koeficientů asociace: Sokalův-Michenerův koeficient asociace, Rogersův-Tanimonoův koeficient asociace, Nepojmenovaný 1 koeficient asociace.
Protože tyto koeficienty nabývají hodnot v intervalu <0,1>, je možný jejich převod na míru nepodobnosti takto: míra nepodobnosti = 1 - míra podobnosti. Tento převod je v nadstavbovém modulu ošetřen procedurou s názvem *VzdalDich()*. Tato procedura je spuštěna kliknutím na tlačítko „Vzdálenost“, které je umístěno na listu „Koef_asociace“. Po výpočtu je vytvořen list s názvem „Vzdálenosti“ s tabulkou vzdáleností (měr nepodobností) mezi objekty.

***Popis procedury výpočtu koeficientů asociace:***

Název procedury: *KoefAsoc()*
Vstupní parametry procedury:
Vstupní asociační tabulka *N* – počet objektů Výpočty v proceduře, podle zvolené metody provádí procedura výpočty koeficientů asociace.
Výstup procedury:
Procedura vytvoří list s názvem „Koef_asociace“ s tabulkou koeficientů asociace.

**3. krok – shlukování / 4. krok – vytvoření dendrogramu**

Poté co je vytvořen list vzdáleností mezi objekty zadanými dichotomickými daty se pokračuje ve shlukové analýze stejným způsobem jako v případě práce s kvantitativními daty. Po stisknutí tlačítka „Shlukování“ na listu „Vzdálenost“ dojde k provedení hierarchického shlukování objektů podle

zvolené metody. Dále je možné vygenerovat dendrogram kliknutím na tlačítko „Vytvořit Dendrogram“ na listu „Shlukování“. Shlukování a vytvoření dendrogramu v případě práce s dichotomickými daty je prováděno stejnými procedurami jako v případě práce s kvantitativními daty.

# 2 Metodické pokyny k použití modulu

V této kapitole je popsaný způsob práce s nadstavbovým modulem pro provádění shlukové analýzy hierarchickými (aglomerativními) shlukovacími metodami. Nadstavbový modul je primárně určený jako učební pomůcka pro předmět Zpracování dat metodami shlukové analýzy a jeho funkční možnosti jsou přizpůsobeny rozsahu výuky tohoto předmětu. Soubor Shluk_Hier.xls a nadstavbový modul je možné používat v MS Office Excel 2003, MS Office Excel 2007 a MS Office Excel 2010.
Po spuštění souboru Shluk_Hier.xls se otevře aplikace Excel s jedním listem s názvem Start. Dále je možné pokračovat zpracováním buď kvantitativních dat, nebo dichotomických dat.

## 2.1 Práce s kvantitativními daty

Návod jak postupovat při práci s kvantitativními daty bude ukázán pomocí následujícího příkladu.

***Příklad 1:***

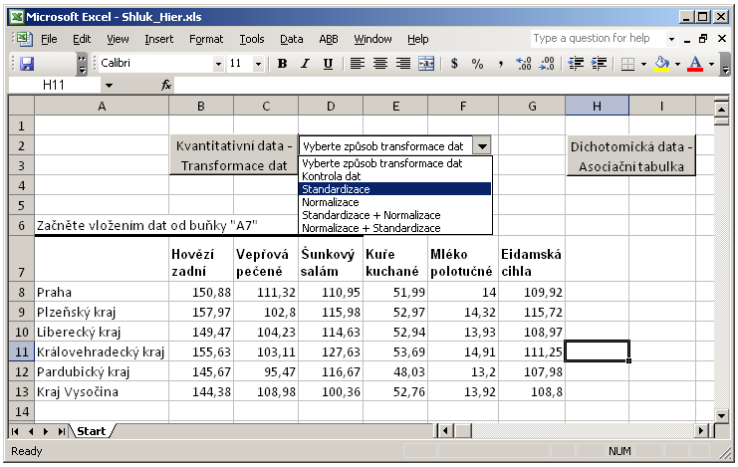
Tabulka 1 vyjadřuje ceny vybraných potravin v obchodech v jednotlivých krajích. Ve kterých krajích jsou ceny potravin nejvíce podobné? Použijte shlukovou analýzu, hierarchickou metodu, aglomerativní přístup. Volte Euklidovskou vzdálenost a metodu nejbližšího souseda.

Tabulka 1 Příklad pro kvantitativní data [zdroj: autor]

| | **Hovězí zadní** | **Vepřová pečeně** | **Šunkový salám** | **Kuře kuchané** | **Mléko polotučné** | **Eidamská cihla** |
|---|---|---|---|---|---|---|
| Praha | 150,88 | 111,32 | 110,95 | 51,99 | 14 | 109,92 |
| Plzeňský kraj | 157,97 | 102,8 | 115,98 | 52,97 | 14,32 | 115,72 |
| Liberecký kraj | 149,47 | 104,23 | 114,63 | 52,94 | 13,93 | 108,97 |
| Královehradecký kraj | 155,63 | 103,11 | 127,63 | 53,69 | 14,91 | 111,25 |
| Pardubický kraj | 145,67 | 95,47 | 116,67 | 48,03 | 13,2 | 107,98 |
| Kraj Vysočina | 144,38 | 108,98 | 100,36 | 52,76 | 13,92 | 108,8 |

***Řešení:***

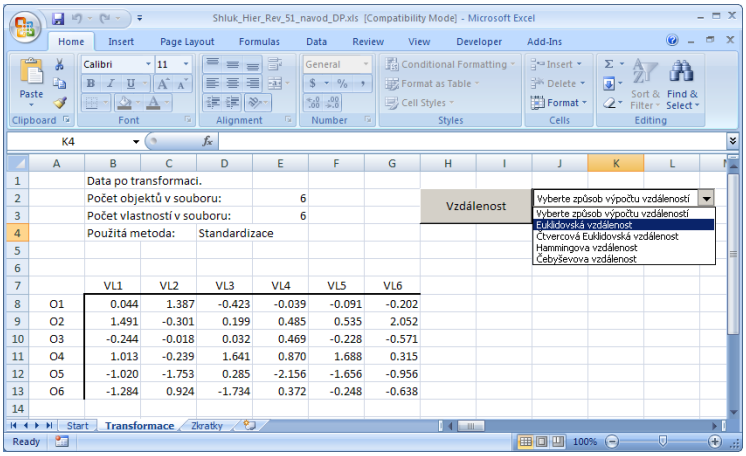
Tabulku je potřeba zkopírovat do listu „Start“ tak, že levý horní roh tabulky bude na buňce A7, sloupec A obsahuje názvy objektů shlukování a řádek 7 obsahuje názvy vlastností těchto objektů. Data začínají na buňce B8.




Obrázek 2 Příklad 1 – zkopírování dat do souboru Shluk_Hier.xls [zdroj: autor]

Protože se jedná o kvantitativní data, prvním krokem při jejich zpracování je transformace dat. Volba způsobu transformace se provede pomocí roletového menu vedle tlačítka Kvantitativní data – transformace dat. Nadstavbový modul umožňuje provést standardizaci, normalizaci, standardizaci a následně normalizaci dat nebo obráceně normalizaci a následně standardizaci dat. V případě, že není požadována žádná transformace dat, provede se volba Kontrola dat.

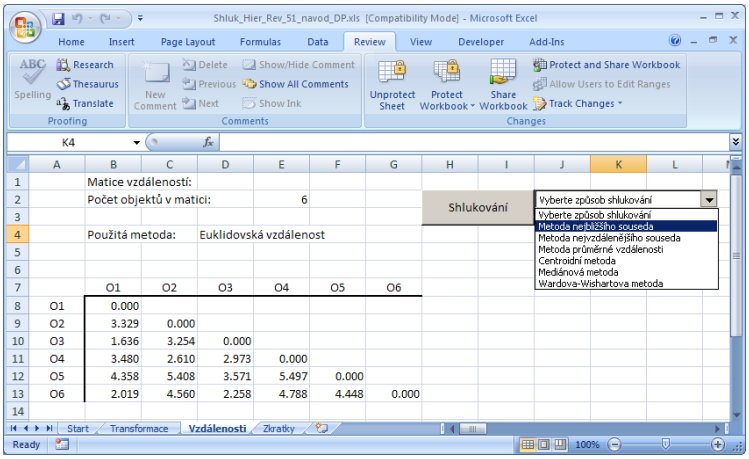
V příkladu je zvolena standardizace dat, aby rozdílná cena jednotlivých produktů neovlivňovala jejich vliv na rozdíl cen v krajích. Kliknutím na tlačítko Kvantitativní data – transformace dat dojde k provedení transformace dat. Vzniknou nové listy s názvem „Transformace“ a „Zkratky“. Na listu Transformace jsou transformovaná data, na listu Zkratky je legenda k vlastnostem VL1 … VL6 a k objektům O1 … O6. Viz obrázek 3.




Obrázek 3 Příklad 1 – výsledek transformace dat [zdroj: autor]

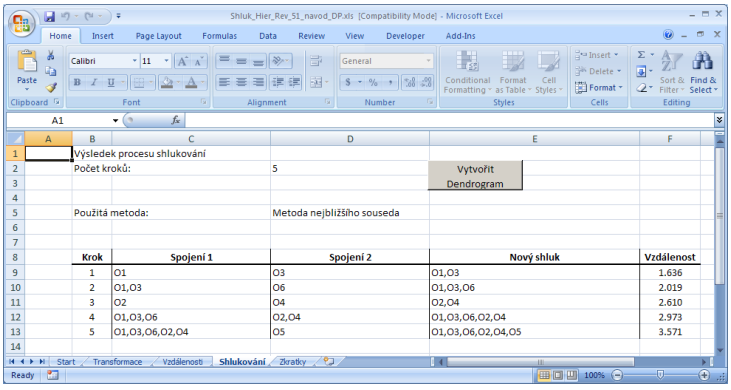
Dalším krokem je výpočet vzdálenosti mezi objekty. Volba způsobu výpočtu vzdálenosti se provede pomocí roletového menu vedle tlačítka Vzdálenost. Nadstavbový modul umožňuje výpočet Euklidovské vzdálenosti, čtvercové euklidovské vzdálenosti, Hammingovy vzdálenosti a Čebyševovy vzdálenosti. V příkladu je zvolena Euklidovská vzdálenost. Kliknutím na tlačítko Vzdálenost dojde k

provedení výpočtu vzdálenosti. Vznikne nový list s názvem „Vzdálenosti“. Na tomto listu je tabulka vzdáleností mezi objekty. Viz obrázek 4. Po výpočtu vzdálenosti následuje proces shlukování. Nadstavbový modul provádí shlukování hierarchickou metodou, aglomerativním přístupem. Proto je nyní potřeba zadat způsob výpočtu vzdálenosti mezi shluky. Měření podobnosti shluků. Volba způsobu výpočtu vzdálenosti se provede pomocí roletového menu vedle tlačítka Shlukování. Nadstavbový modul umožňuje výpočet vzdálenosti metodou nejbližšího souseda, nejvzdálenějšího souseda, průměrné vzdálenosti, centroidní metodou, mediánovou metodou a Ward-Wishartovou metodou.



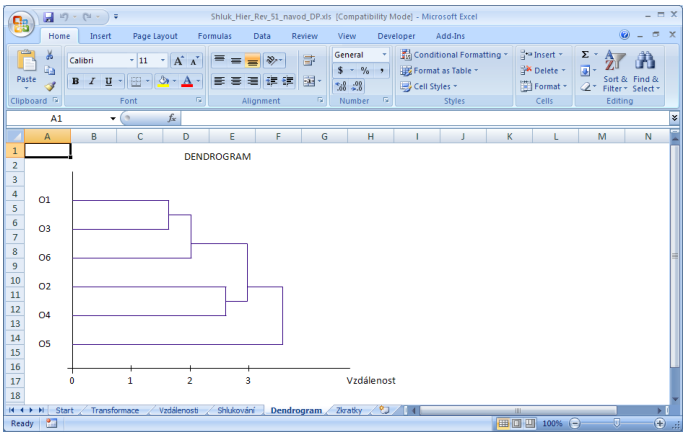

Obrázek 4 Příklad 1 – list se vzdálenostmi mezi objekty [zdroj: autor]

V příkladu je zvolena metoda nejbližšího souseda. Kliknutím na tlačítko Shlukování dojde ke spuštění procesu shlukování. Vznikne nový list s názvem „Shlukování“. Na tomto listu je tabulka, která vystihuje, jaké shluky byly spojeny v jednotlivých krocích a při jaké vzdálenosti ke spojení došlo. Viz obrázek 5.




Obrázek 5 Příklad 1 – list s výsledkem procesu shlukování [zdroj: autor]

Pro grafickou reprezentaci výsledku shlukování se používá stromový graf, který se nazývá dendrogram. Ten je možné vygenerovat kliknutím na tlačítko Vytvořit Dendrogram. Vznikne nový list s názvem „Dendrogram“. Dendrogram pro tento příklad zobrazuje obrázek 6.




Obrázek 6 Příklad 1 – vytvoření dendrogramu [zdroj: autor]

## 2.2 Práce s dichotomickými daty

Návod jak postupovat při práci s dichtomickými daty bude demonstrován pomocí dalšího příkladu.

***Příklad 2:***

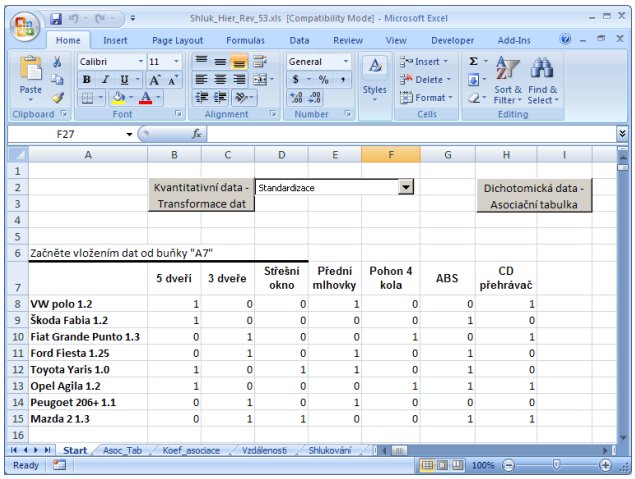
Tabulka 6 uvádí některé vlastnosti v základní výbavě vybraných automobilů s obsahem motoru 1,0 až 1.3. Které automobily jsou si z hlediska těchto vlastností nejvíce podobné, a které lze zařadit do stejných skupin? Použijte shlukovou analýzu, hierarchickou metodu, aglomerativní přístup. Volte Sokalův a Michenerův koeficient asociace a Wardova-Wishartova metodu pro výpočet vzdálenosti mezi shluky.

Tabulka 6 Příklad pro dichotomická data [zdroj: autor]

| | 5 dveří | 3 dveře | Střešní okno | Přední mlhovky | Pohon 4 kola | ABS | CD přehrávač |
|---|---|---|---|---|---|---|---|
| VW polo 1.2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| Škoda Fabia 1.2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Fiat Grande Punto 1.3 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| Ford Fiesta 1.25 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Toyota Yaris 1.0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| Opel Agila 1.2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Peugoet 206+ 1.1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| Mazda 2 1.3 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

***Řešení:***

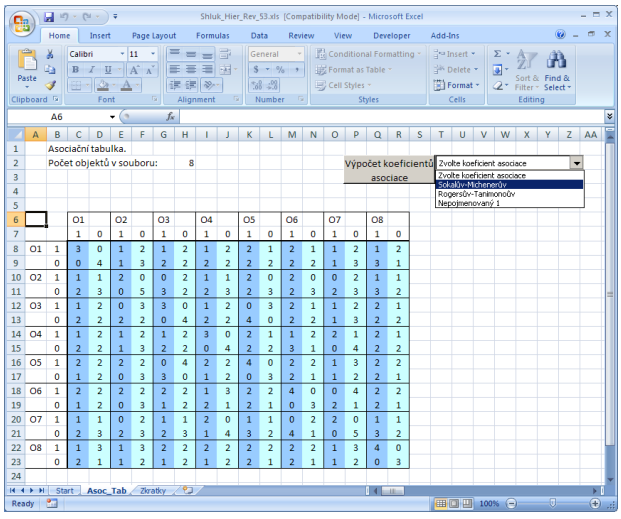
Tabulku zkopírujeme do listu „Start“ analogicky jako v případě práce s kvantitativními daty. Viz obrázek 7.




Obrázek 7 Příklad 2 – zkopírování dat do souboru Shluk_Hier.xls [zdroj: autor]

Při práci s dichotomickými daty, stejně tak jako při práci s kvantitativními daty, je nutné určit vzdálenost mezi objekty. Nadstavbový modul realizuje výpočet této vzdálenosti tak, že nejprve vytvoří asociační tabulku. Po té se z asociační tabulky vypočítají koeficienty asociace. Protože koeficienty asociace vyjadřují míru podobnosti, je potřeba pro proces shlukování hierarchickou metodou tuto míru podobnosti převést na míru nepodobnosti, tj. vzdálenost mezi objekty.

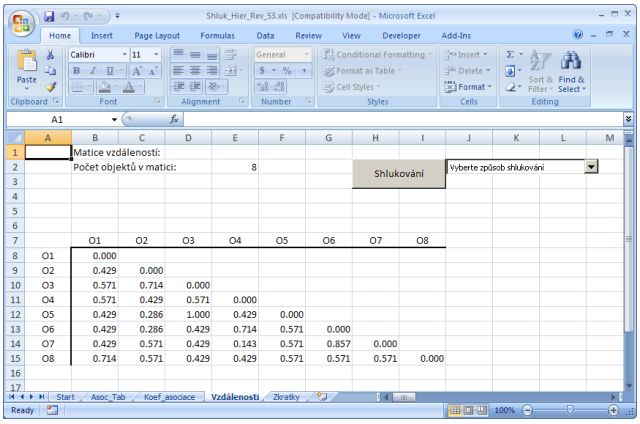
Kliknutím na tlačítko Dichotomická data – Asociační tabulka, dojde nejprve ke kontrole dat a v případě, že jsou dichotomická data v pořádku (soubor hodnot nabývá pouze 0 a 1), dojde k vytvoření nového listu s názvem „Asoc_Tab“, viz obrázek 8. Zároveň dojde k vytvoření listu s názvem „Zkratky“ s tabulkou zkratek. Poté následuje výpočet koeficientů asociace. Volba typu koeficientu asociace se provede pomocí roletového menu vedle tlačítka Výpočet koeficientů asociace. Nadstavbový modul umožňuje výpočet těchto koeficientů asociace: Sokalův-Michenerův koeficient asociace, Rogersův-Tanimonoův koeficient asociace a Nepojmenovaný 1 koeficient asociace.




Obrázek 8 Příklad 2 – vytvoření listu s asociační tabulkou [zdroj: autor]

V příkladu je zvolen Sokalův-Michenerův koeficient asociace. Kliknutím na tlačítko Výpočet koeficientů asociace dojde k provedení výpočtu koeficientů asociace. Vznikne nový list s názvem „Koef_asociace“. Na tomto listu je tabulka koeficientů asociace.

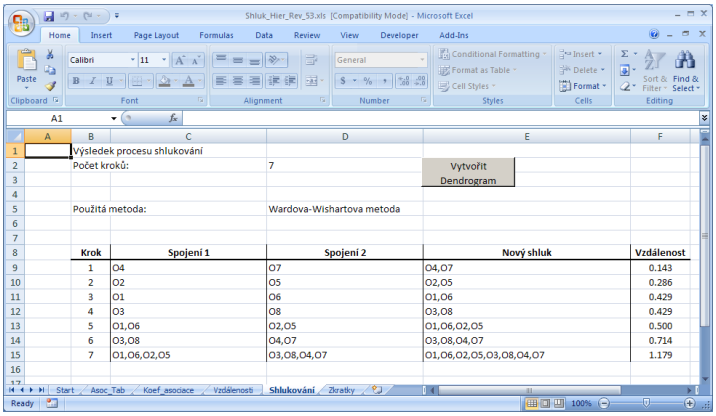
Dalším krokem je přepočet koeficientů asociace na vzdálenost mezi objekty Kliknutím na tlačítko Vzdálenost dojde k provedení výpočtu vzdálenosti. Vznikne nový list s názvem „Vzdálenosti“. Na tomto listu je tabulka vzdáleností mezi objekty. Viz obrázek 9.



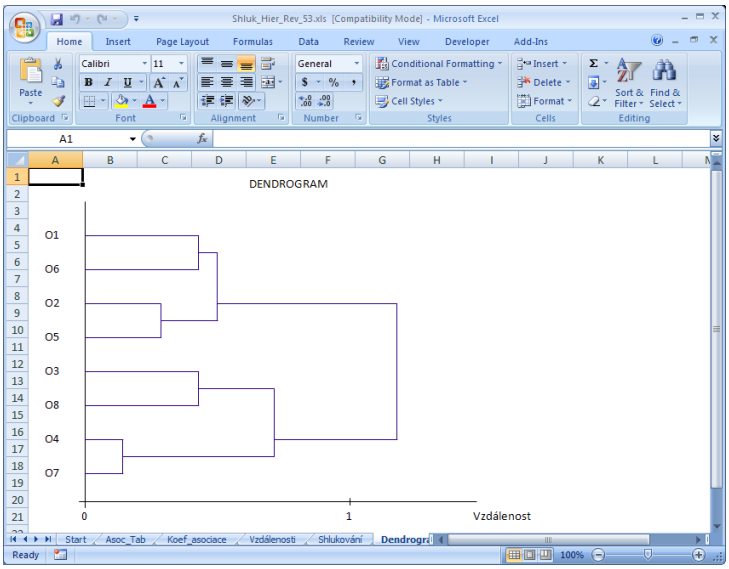

Obrázek 9 Příklad 2 – list se vzdálenostmi mezi objekty [zdroj: autor]

Nyní se již postupuje stejným způsobem jako při zpracovávání kvantitativních dat. Po výpočtu vzdálenosti následuje proces shlukování. Je potřeba zadat způsob výpočtu vzdálenosti mezi shluky. Ta se provede pomocí roletového menu vedle tlačítka Shlukování. V příkladu je zvolena Wardova-Wishartova metoda. Kliknutím na tlačítko Shlukování dojde ke spuštění procesu shlukování. Vznikne nový list s názvem „Shlukování“. Na tomto listu je tabulka, která vystihuje, jaké shluky byly spojeny v jednotlivých krocích a při jaké vzdálenosti ke spojení došlo. Viz obrázek 10.

Pro grafickou reprezentaci výsledku shlukování slouží stromový graf, který se nazývá dendrogram. Ten je možné vygenerovat kliknutím na tlačítko Vytvořit Dendrogram. Vznikne nový list s názvem „Dendrogram“. Dendrogram pro tento příklad zobrazuje obrázek 11.




Obrázek 10 Příklad 2 – list s výsledkem procesu shlukování [zdroj: autor]




Obrázek 11 Příklad 2 – vytvoření dendrogramu [zdroj: autor]

# 3 Řešení známých problémů s nadstavbovým modulem

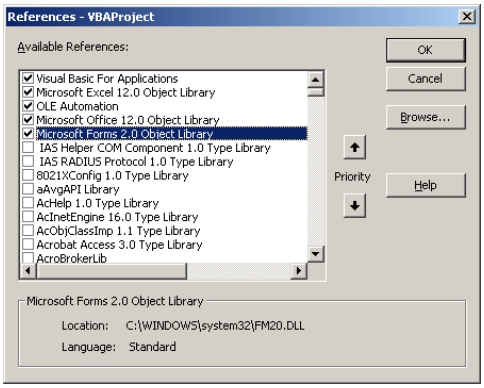
## 3.1 Povolení maker

Pro správnou funkci nadstavbového modulu je zapotřebí mít v aplikaci Excel povolená makra. Úprava nastavení spouštění maker se provádí volbou tlačítka Office → Možnosti aplikace Excel → Centrum zabezpečení → Nastavení Centra zabezpečení →Nastavení maker.

## 3.2 Error in loading DLL

Nadstavbový modul je testován v těchto verzích aplikace Excel: Excel 97-2003, Excel 2007, Excel 2010. Může se stát, že při různých pokusech s modulem, zejména je-li modul používán v různých verzích aplikace Excel a zároveň je spouštěn MS Visual Basic editor, dojde k vyvolání této chybové hlášky: „Error in loading DLL“.

V tomto případě je potřeba při otevřeném nadstavbovém modulu spustit MS Visual Basic editor, vybrat menu Tools / References a v otevřeném okně upravit nastavení DLL knihoven. Viz obrázek 12.




Obrázek 12 Nastavení knihoven DLL [zdroj: autor]

### 3.3 Vykreslování dendrogramu

Vykreslování dendrogramu může trvat dlouhou dobu (v závislosti na rychlosti počítače), nebo může skončit neúspěchem, tj. Visual Basic skript nahlásí chybu a bude ukončen. Tento problém je nejčastěji způsoben chybnou, nebo dokonce žádnou transformací vstupních dat.

Hodnoty na ose X se nemění dynamicky. Osa X je popsána celými čísly od nuly do vzdálenosti, při které došlo k sloučení všech objektů do jednoho shluku (Má-li tato hodnota desetinná místa, pak je popis osy zaokrouhlen dolu na celé číslo). Jestliže je tato vzdálenost příliš vysoká, vypisování všech hodnot do dendrogramu je časově náročné. Dalším důvodem neúspěšného vykreslení dendrogramu může být jeho velikost. Nadstavbový modul vykresluje dendrogram takovým způsobem, že z důvodu jeho čitelnosti je nejmenší vzdálenost sloučení shluků realizována spojem délky 100 pixelů. Jestliže je poměr mezi nejmenší vzdáleností sloučení shluků a největší vzdáleností sloučení shluků příliš vysoký, může dojít k situaci, že program nemůže vykreslit dendrogram, protože se nevejde na daný list. Řešením těchto situací je volba vhodného způsobu transformace dat, případně vhodná volba výpočtu vzdáleností mezi objekty a shluky.

Kontakt:
jaroslav.lohynsky@student.upce.cz
hana.jonasova@upce.cz