evropský
sociální
fond v ČR

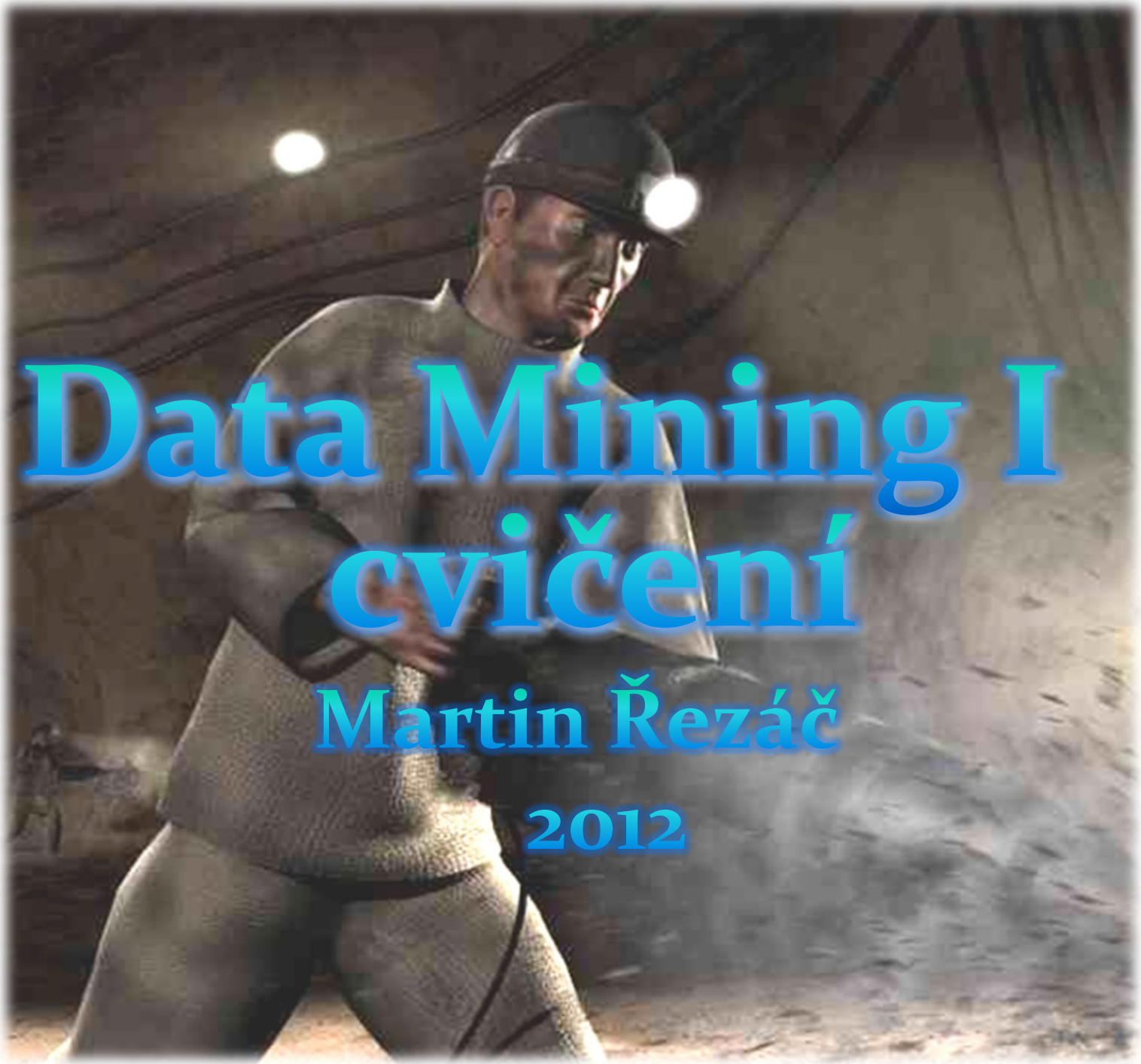
EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

2007-13
OP Vzdělávání
pro konkurenceschopnost

UNIVERSITAS MASARYKIANA BRUNENSIS

INVESTICE
DO ROZVOJE
VZDĚLÁVÁNÍ



# Data Mining I
# cvičení

Martin Řezáč

2012

# Cvičení 1

# Instalační soubory, licenční podmínky

- Instalační soubory SASu jsou k dispozici všem studentům a učitelům ÚMS PřF MU na webu ÚMS v zabezpečené zóně (přístup pod loginem a heslem do domény).
- Před vlastním zobrazením stránky s inst. soubory je nutné odsouhlasit licenční podmínky.

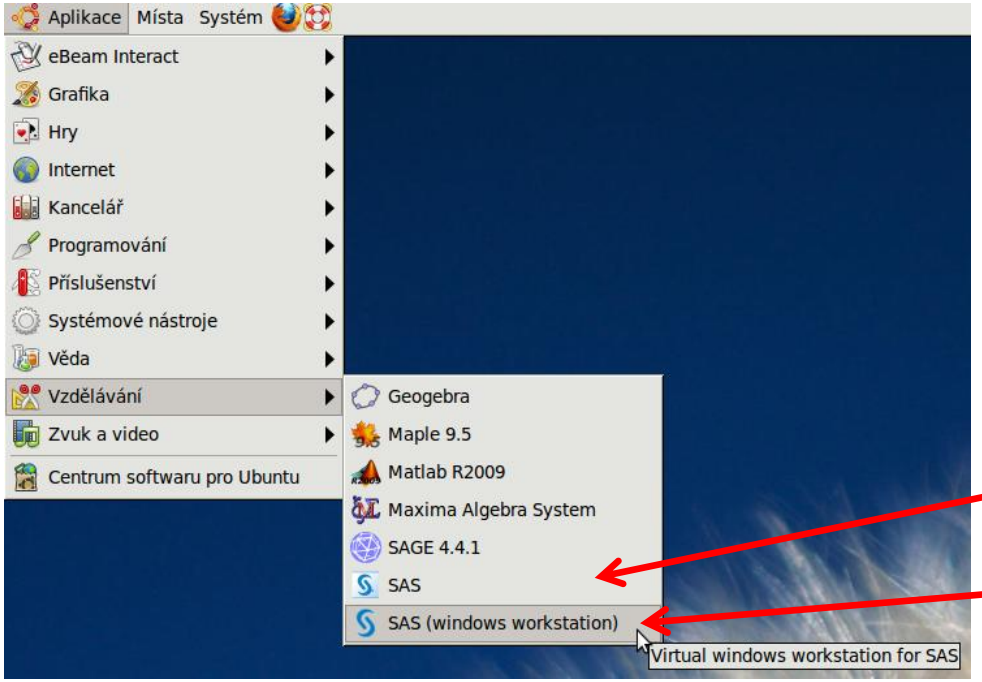

# Instalační soubory, licenční podmínky

- Po odsouhlasení licenčních podmínek jsou k dispozici zkomprimované instalační depa pro OS Windows 32/64bit a Linux 32/64bit.

# Práce v SAS

- Pro studenty (i vyučující) je dostupný SAS na 12 PC ve verzi Linux+Windows a dalších 24 PC ve verzi Linux.
- Výuka probíhá ve verzi Windows (virtuální pod linuxem)

Screenshot (výřez) pracovní plochy:

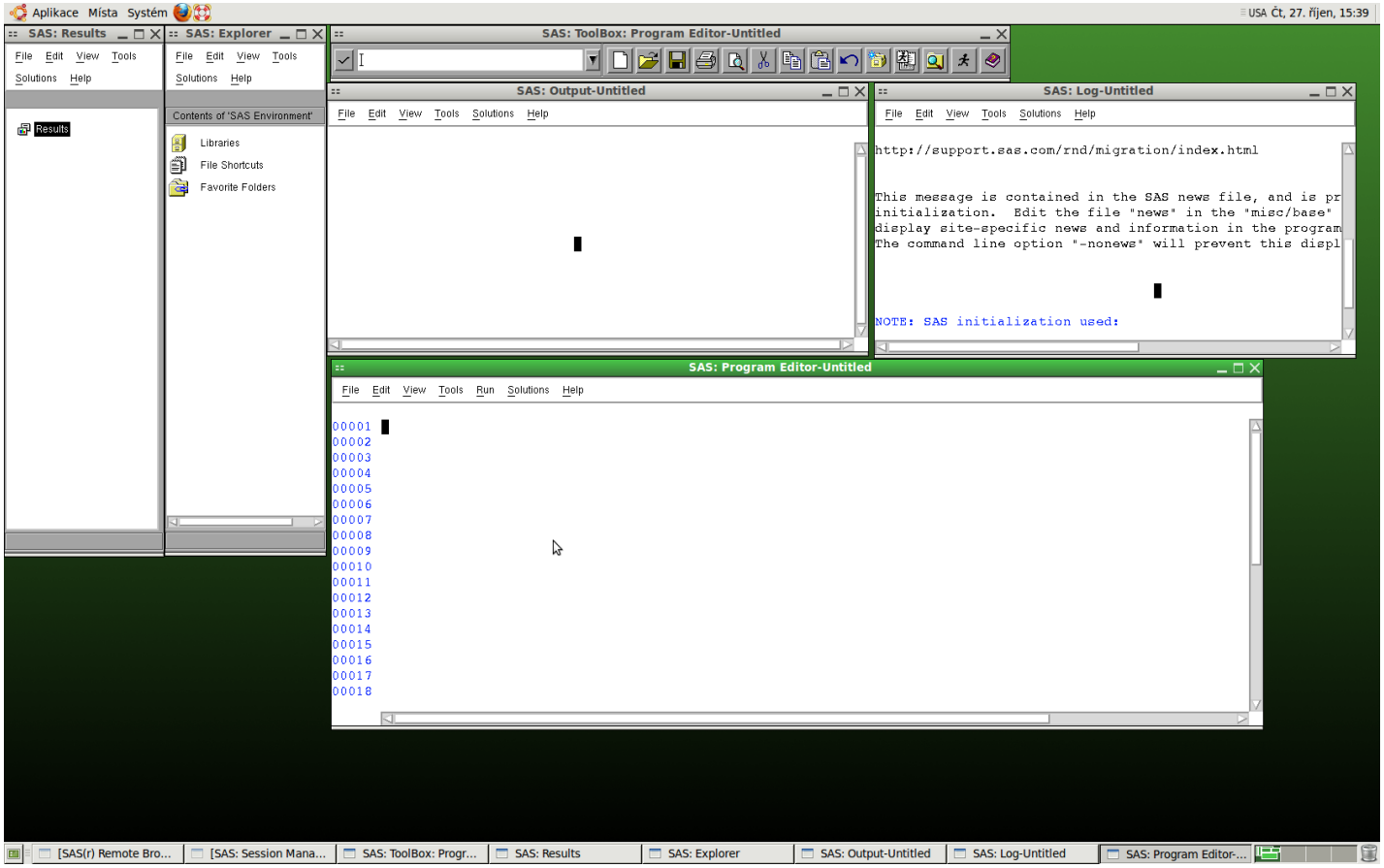



link na vezi linux

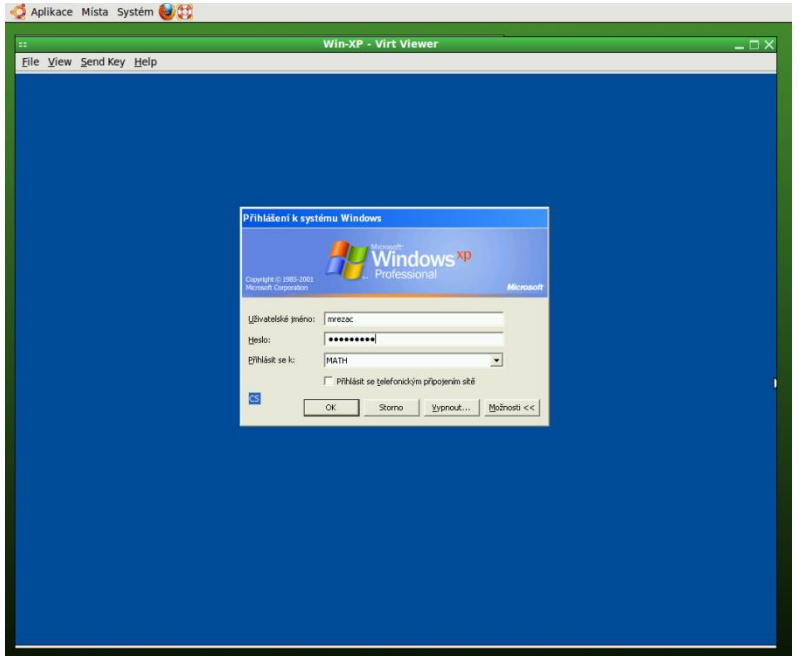
link na verzi „windows“

# Práce v SAS – verze linux

- K dispozici SAS 9.2
- Po spuštění se otevře 6 oken (Results, Explorer, Toolbox, Output, Log, Program Editor)
- Uživatelský komfort je na velmi nízké úrovni, nicméně vše je funkční a pracovat se v „tom“ dá.



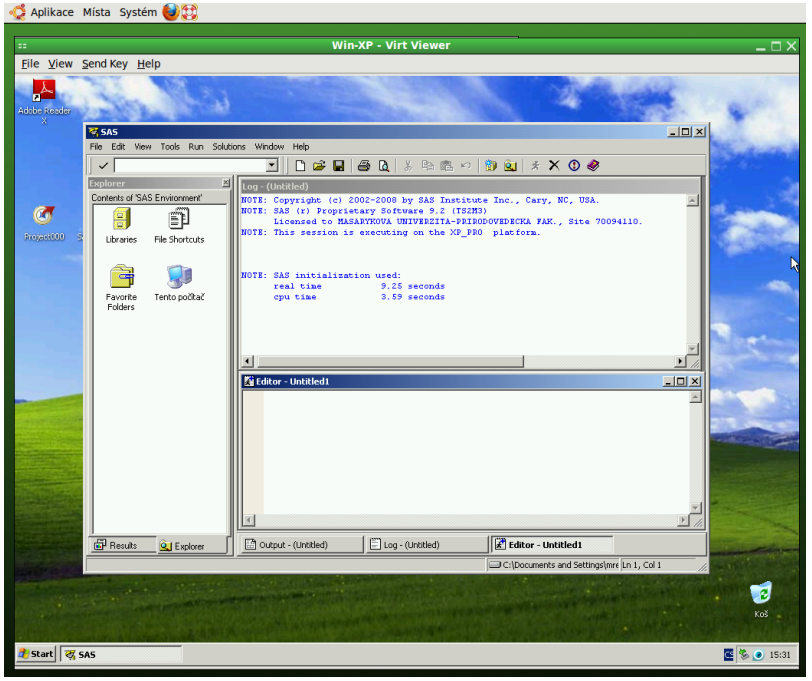
# Práce v SAS – verze windows



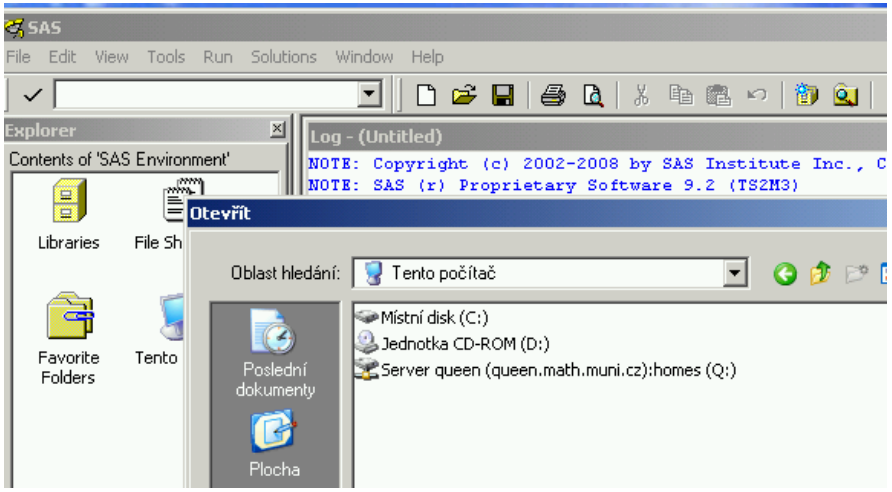

- Po spuštění Windows ve virtuálním prostředí je třeba se přihlásit do domény.

- Po přihlášení je k dispozici:
  - SAS 9.2
  - SAS Enterprice Guide 4.3
  - IML Studio 3.3

# Práce v SAS – verze windows

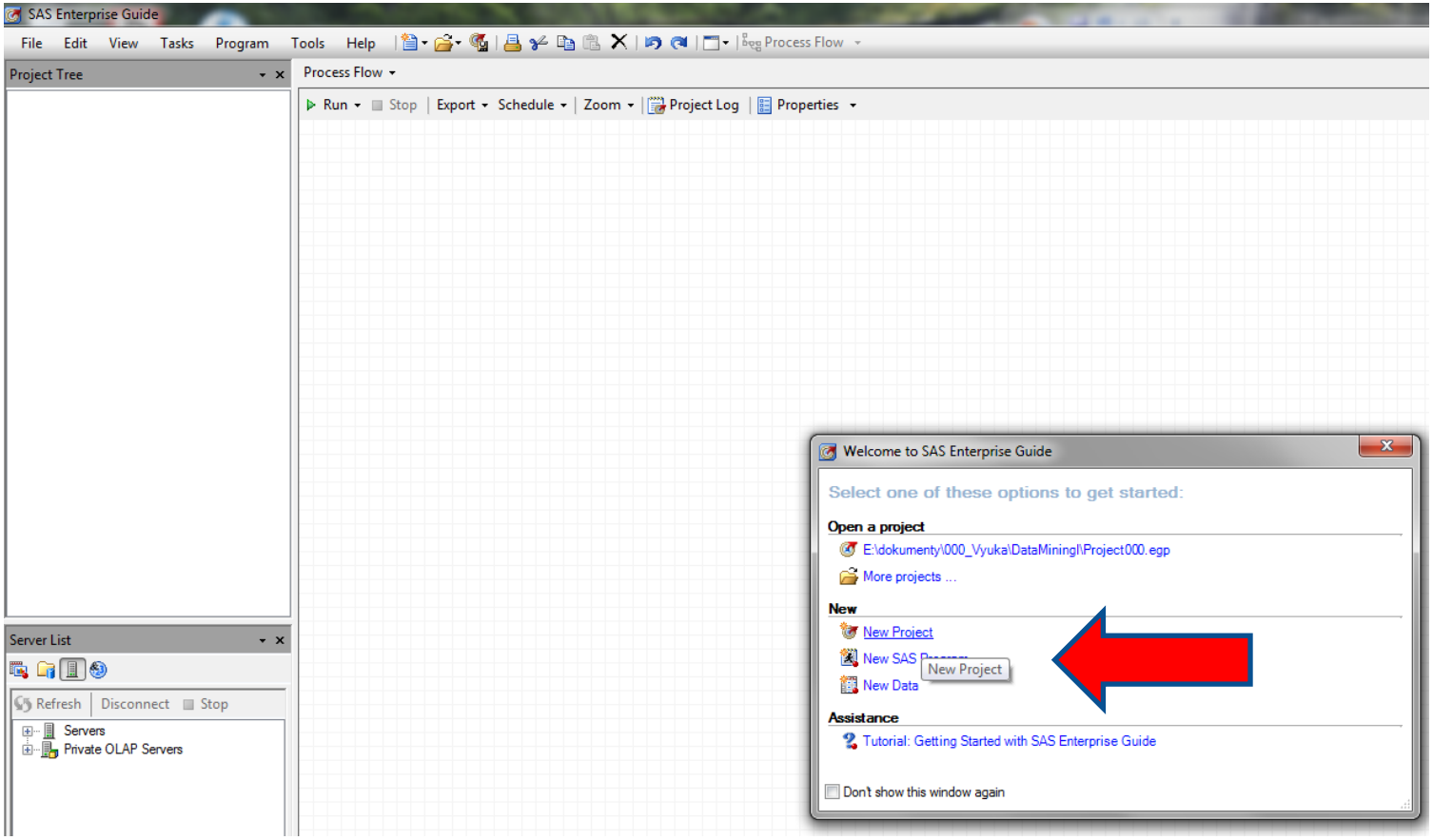



- Vlastní práce v SASu se pak nijak neliší od práce v „klasických“ windows.



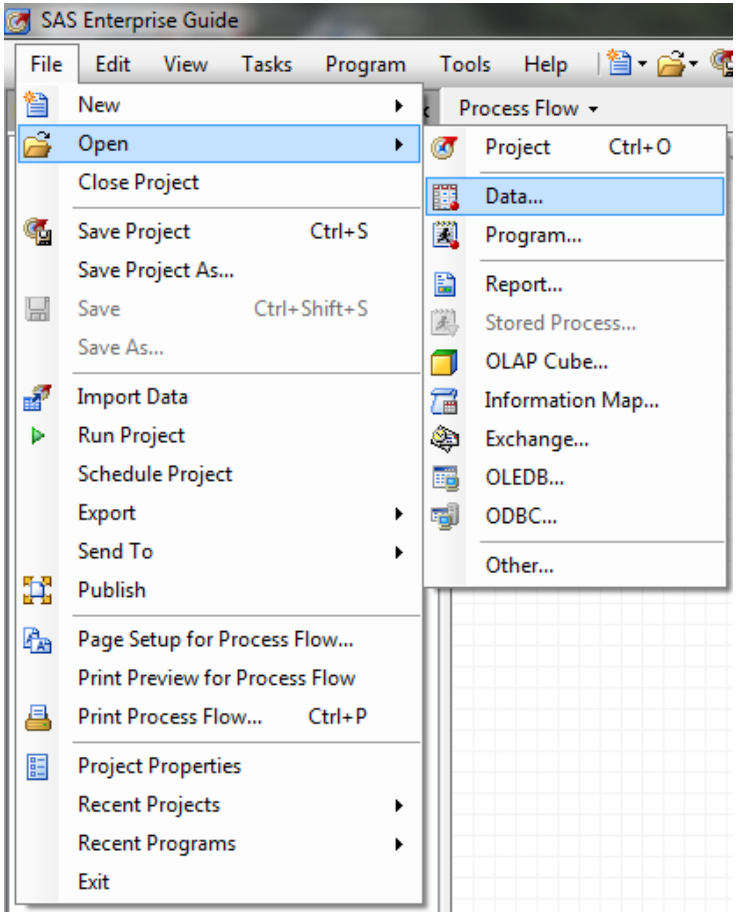

- Ukládat kódy a datové tabulky lze jak na lokálním disku tak na síti v rámci domény.

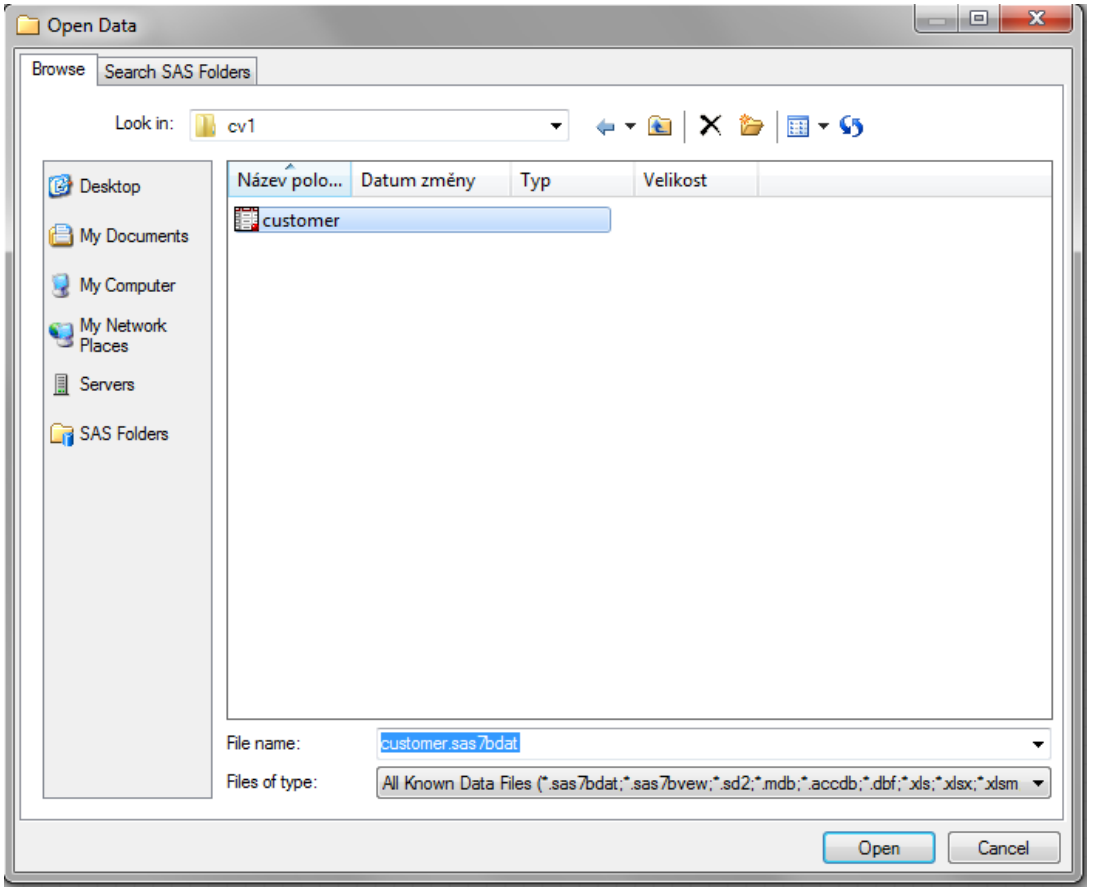
# Enterprice Guide




- Nejprve je třeba vytvořit nový projekt.

# Enterprice Guide

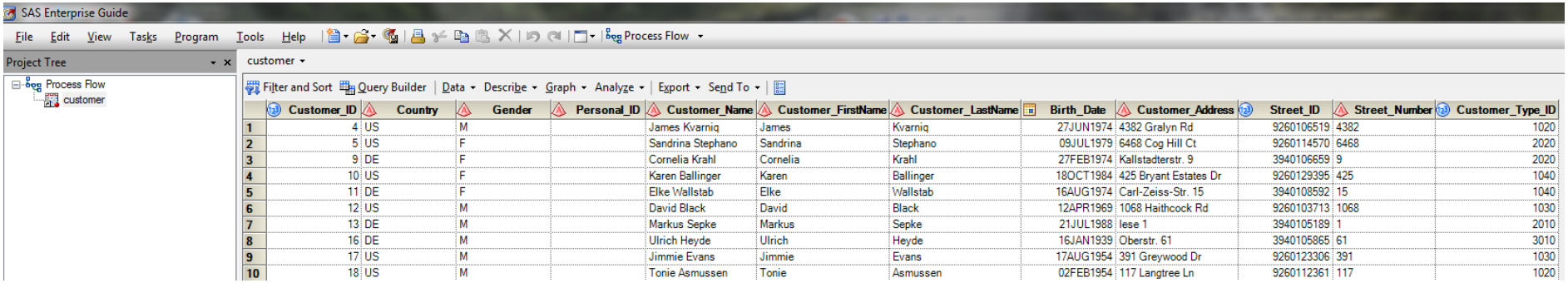



- Načteme data (customer.sas7bdat – studijni materialy v ISu)

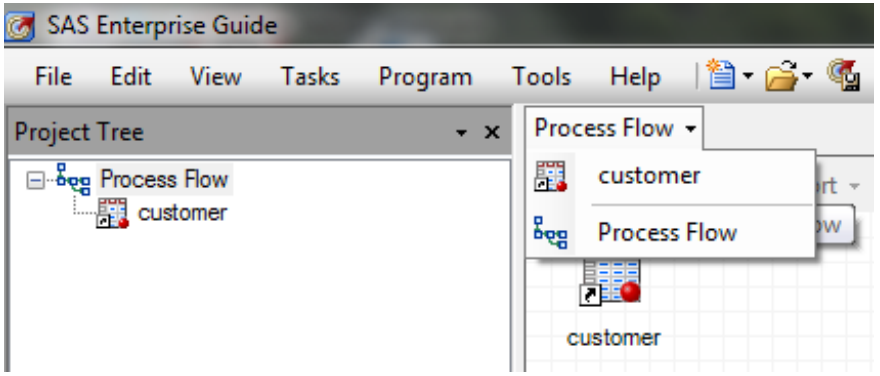
# Enterprice Guide

- Zobrazí se datová tabulka:



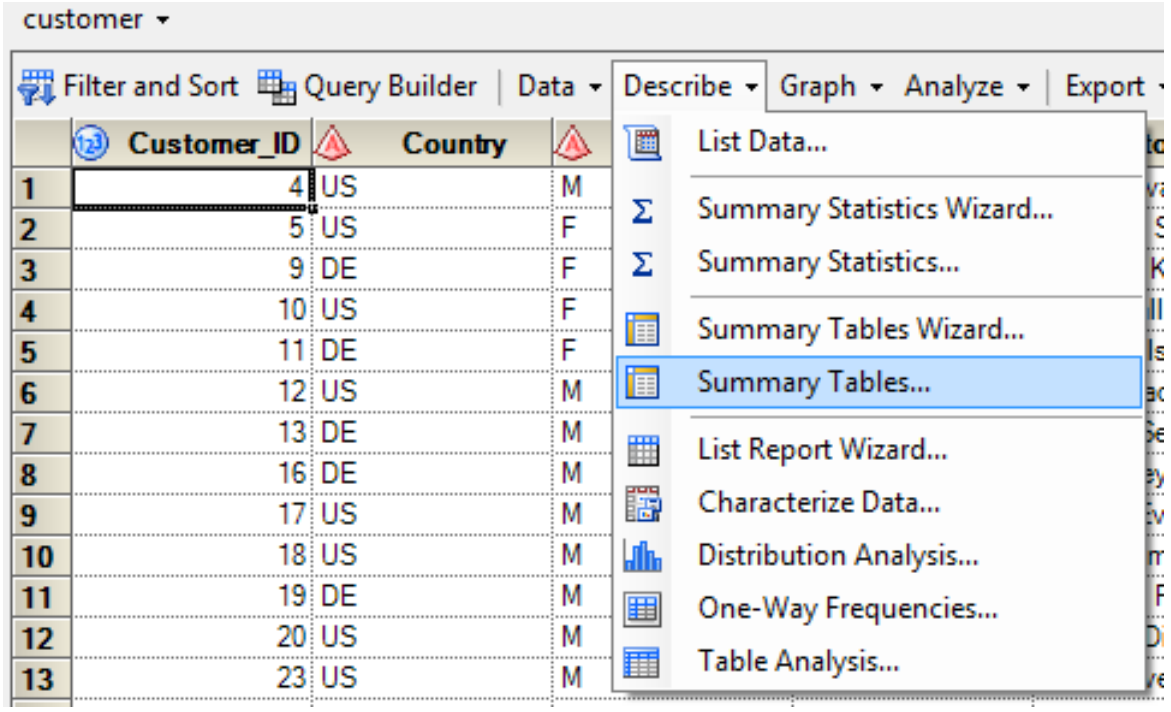

| | Customer_ID | Country | Gender | Personal_ID | Customer_Name | Customer_FirstName | Customer_LastName | Birth_Date | Customer_Address | Street_ID | Street_Number | Customer_Type_ID |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | US | M | | James Kvarniq | James | Kvarniq | 27JUN1974 | 4382 Gralyn Rd | 9260106519 | 4382 | 1020 |
| 2 | 5 | US | F | | Sandrina Stephano | Sandrina | Stephano | 09JUL1979 | 6468 Cog Hill Ct | 9260114570 | 6468 | 2020 |
| 3 | 9 | DE | F | | Cornelia Krahl | Cornelia | Krahl | 27FEB1974 | Kallstadterstr. 9 | 3940106659 | 9 | 2020 |
| 4 | 10 | US | F | | Karen Ballinger | Karen | Ballinger | 18OCT1984 | 425 Bryant Estates Dr | 9260129395 | 425 | 1040 |
| 5 | 11 | DE | F | | Elke Wallstab | Elke | Wallstab | 16AUG1974 | Carl-Zeiss-Str. 15 | 3940108592 | 15 | 1040 |
| 6 | 12 | US | M | | David Black | David | Black | 12APR1969 | 1068 Haithcock Rd | 9260103713 | 1068 | 1030 |
| 7 | 13 | DE | M | | Markus Sepke | Markus | Sepke | 21JUL1988 | lese 1 | 3940105189 | 1 | 2010 |
| 8 | 16 | DE | M | | Ulrich Heyde | Ulrich | Heyde | 16JAN1939 | Oberstr. 61 | 3940105865 | 61 | 3010 |
| 9 | 17 | US | M | | Jimmie Evans | Jimmie | Evans | 17AUG1954 | 391 Greywood Dr | 9260123306 | 391 | 1030 |
| 10 | 18 | US | M | | Tonie Asmussen | Tonie | Asmussen | 02FEB1954 | 117 Langtree Ln | 9260112361 | 117 | 1020 |

- Lze přepnout zpět na Process Flow



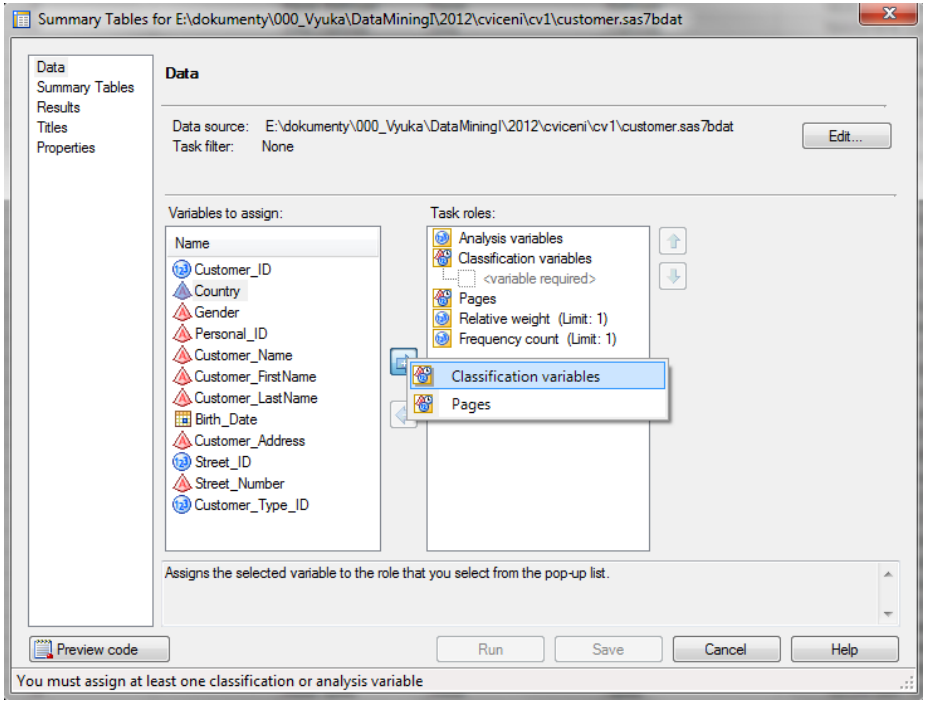
# Enterprice Guide

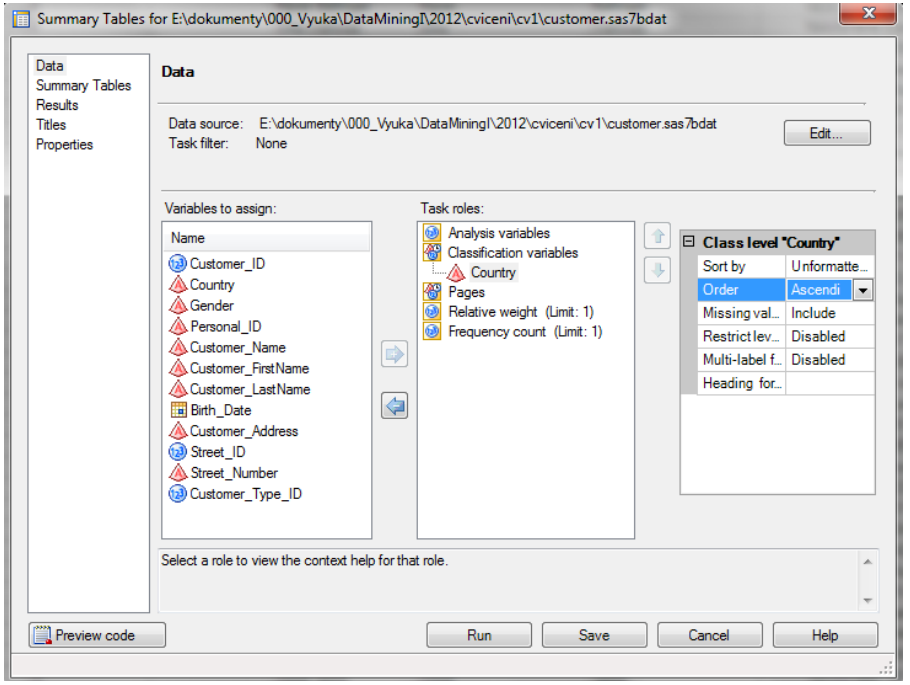
- V záložkách si lze vybrat z řady úloh (kont./frekvenční tabulky, grafy, ANOVA, regrese,...):



# Enterprice Guide

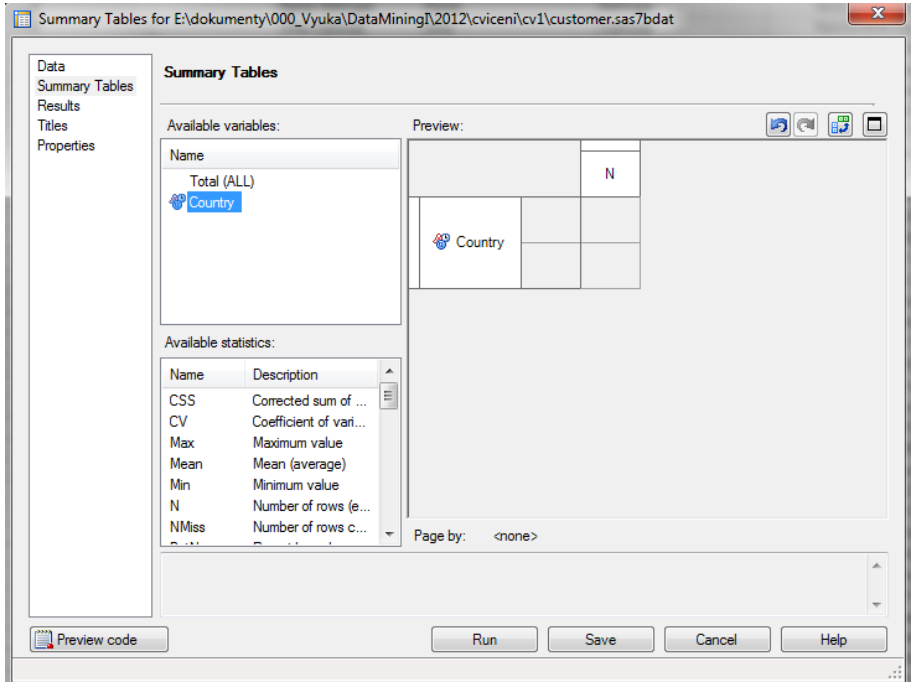
- V záložce „Data“ vybereme proměnné a přiřadíme jim role:
- Např. prom. „Country“ označíme jako „Clasissification“ proměnnou.
- Dále je možné volit např. způsob setřídění výstupu.

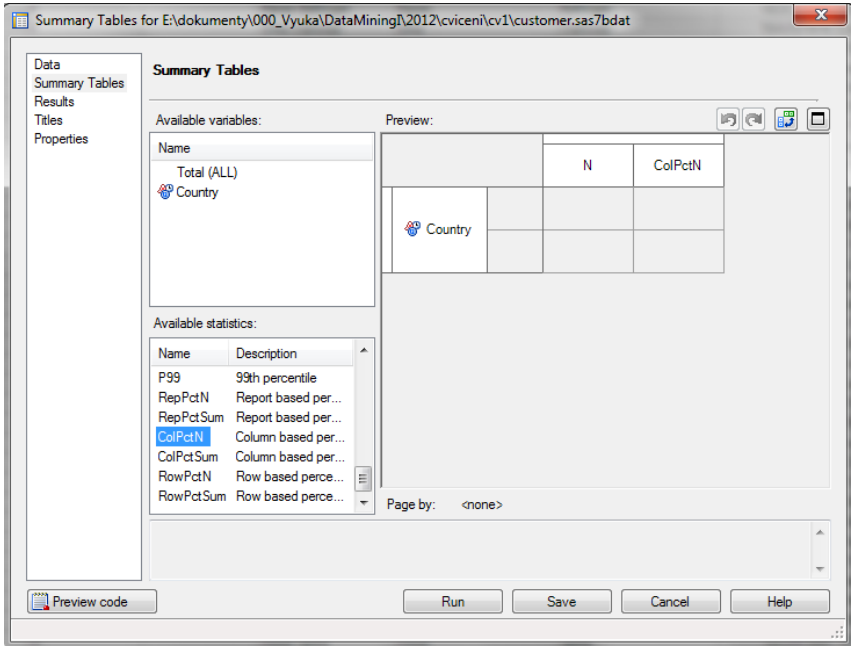


# Enterprice Guide

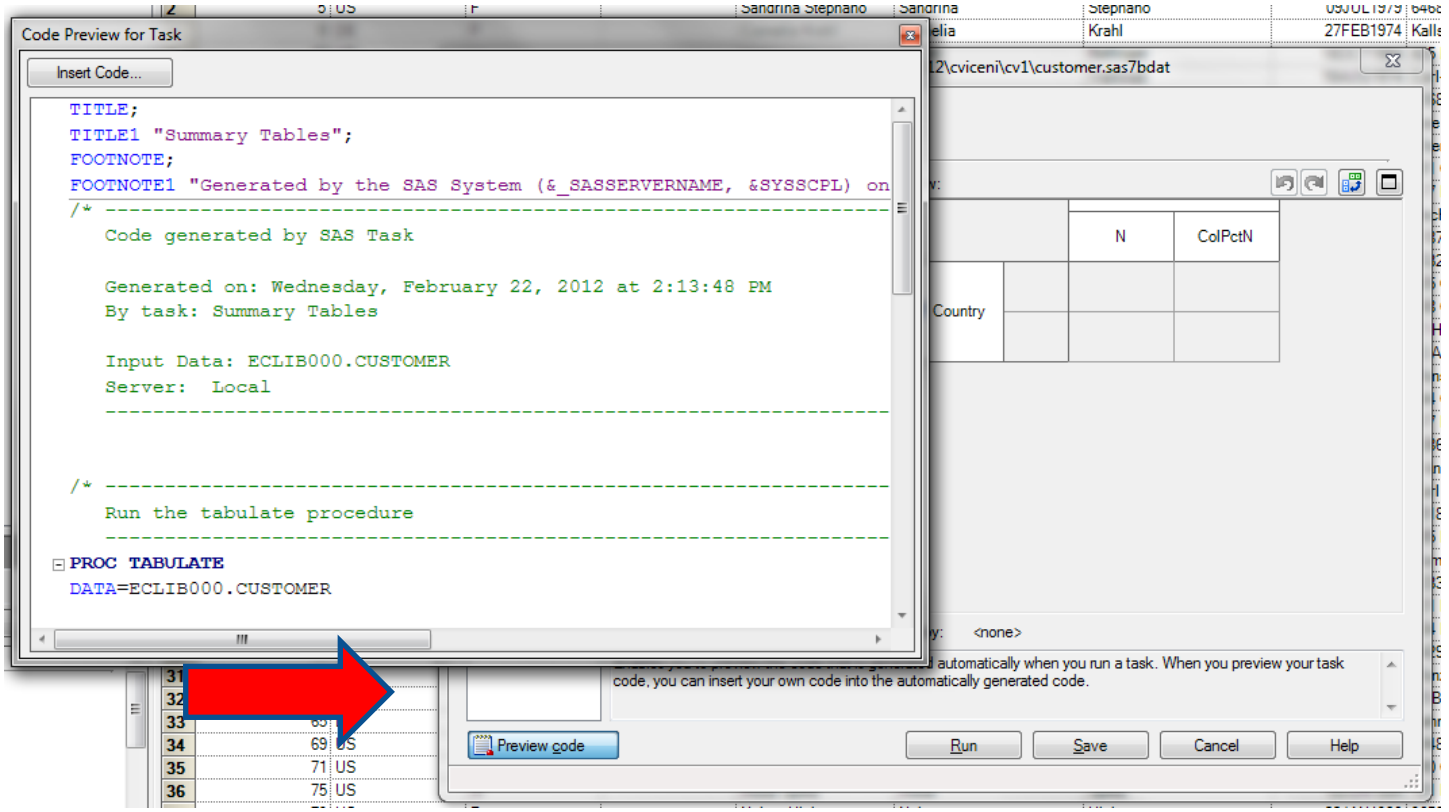
- V záložce „Summary Tables“ nadizajnujeme kontingenční tabulku:



# Enterprice Guide

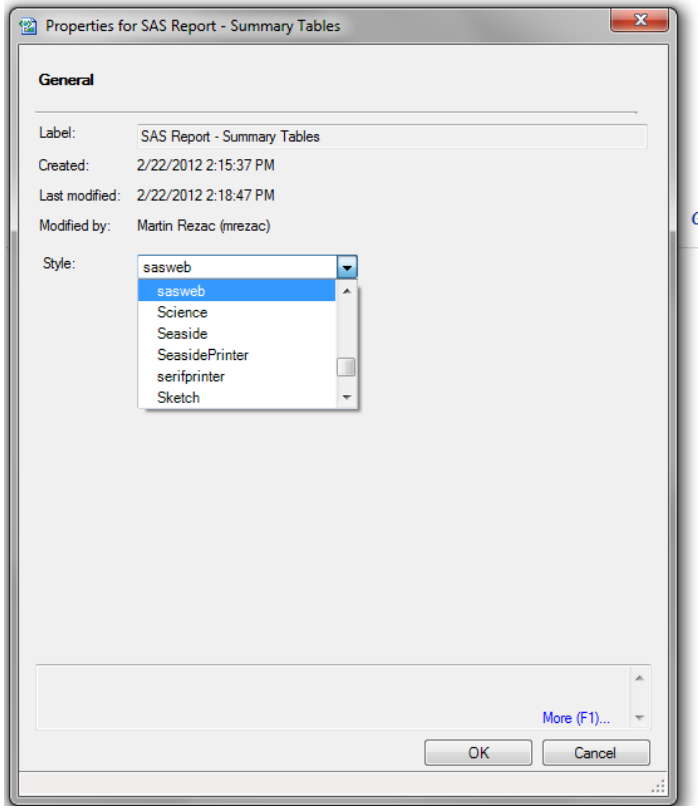
- Po kliknutí na „Preview code“ se zobrazí okno se SASovským kódem, který lze upravovat nebo zkopírovat a použít v programovacím prostředí SASu.



# Enterprice Guide

- Po kliknutí na „Preview code“ se zobrazí okno se SASovským kódem, který lze upravovat nebo zkopírovat a použít v programovacím prostředí SASu.

- V záložce „Properties“ lze měnit styl …např. na „sasweb“




**Summary Tables**

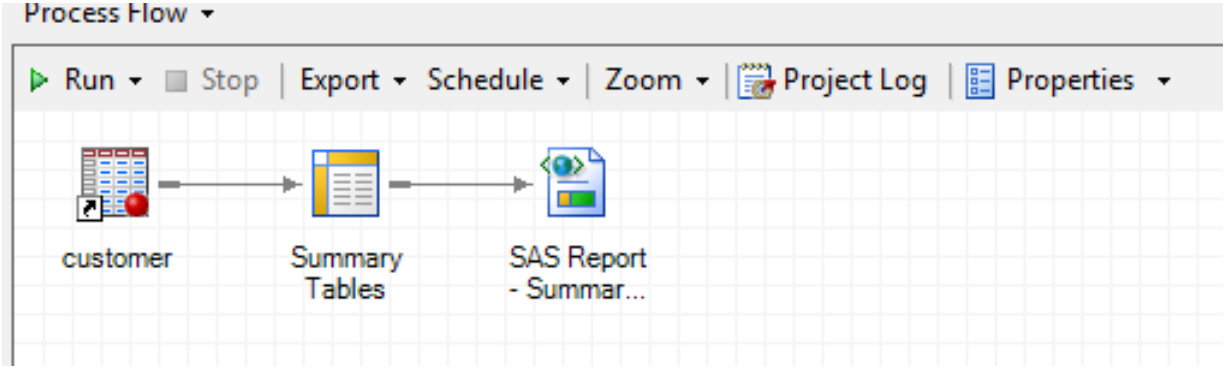
| Customer Country | N | ColPctN |
|---|---|---|
| AU | 8 | 10.39 |
| CA | 15 | 19.48 |
| DE | 10 | 12.99 |
| IL | 5 | 6.49 |
| TR | 7 | 9.09 |
| US | 28 | 36.36 |
| ZA | 4 | 5.19 |

*Generated by the SAS System ('Local', X64_VSPRO) on 22. únor 2012 at 2:15:37 PM*

Page Break

# Enterprice Guide

- V Process Flow přibude uzel pro zvolenou úlohu (Summary Tables) a uzel s výsledky.



# Úkoly

1. Vytvořte kontingenční tabulku pro prom. Country a Gender obsahující absolutní a relativní četnosti včetně řádkově a sloupcově podmíněných relativních četností.

2. Vytvořte koláčový graf pro prom. Country se zobrazením relativních četností.

3. Přeneste příslušné kódy z úkolů 1 a 2 do programovacího prostředí a vygenerujte stejnou tabulku a graf.

4. V Helpu nebo na support.sas.com zjistěte další možnosti úpravy grafu (3D, barvy, fonty písma…)

# Cvičení 2

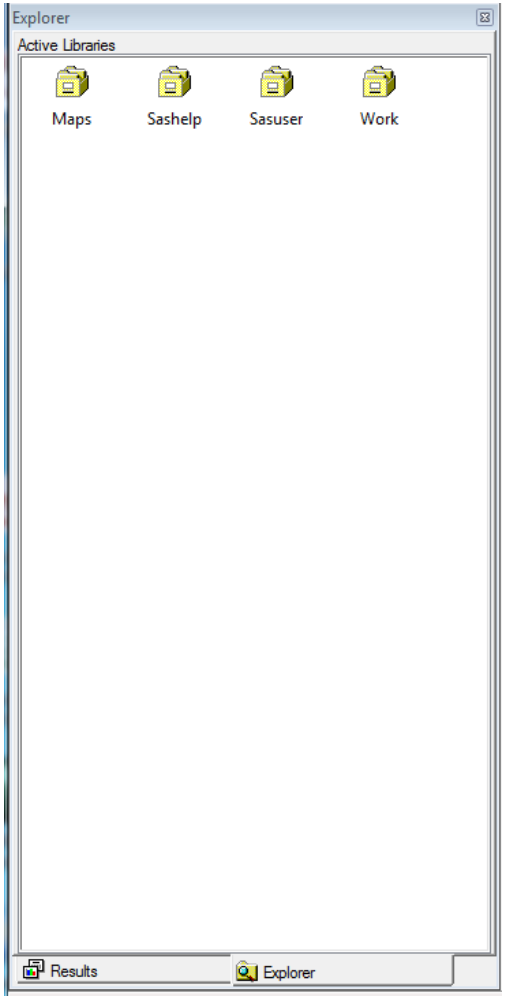
# Libname

Slouží pro namapování knihovny
- typicky jde o adresář na pevném disku.

```
Libname _234567 "D:\dukumenty\prace\vyuka\Data_Mining_1";
Libname dm1 "z:\dm1\data";
```

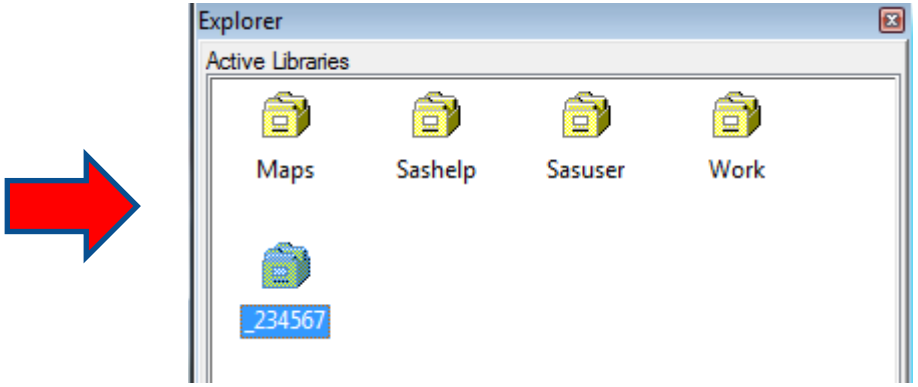
```
1    libname a23456789 "D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez";
ERROR: a23456789 is not a valid SAS name.
ERROR: Error in the LIBNAME statement.

2    libname a234567 "D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez";
NOTE: Libref A234567 was successfully assigned as follows:
      Engine:        V9
      Physical Name: D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez

3    libname _234567 "D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez";
NOTE: Libname _234567 refers to the same physical library as A234567.
NOTE: Libref _234567 was successfully assigned as follows:
      Engine:        V9
      Physical Name: D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez
4    libname 234567 "D:¥dukumenty¥prace¥MU¥vyuka¥Data_Mining_2¥soutez";
ERROR: 234567 is not a valid SAS name.
ERROR: Error in the LIBNAME statement.
```

# Libname



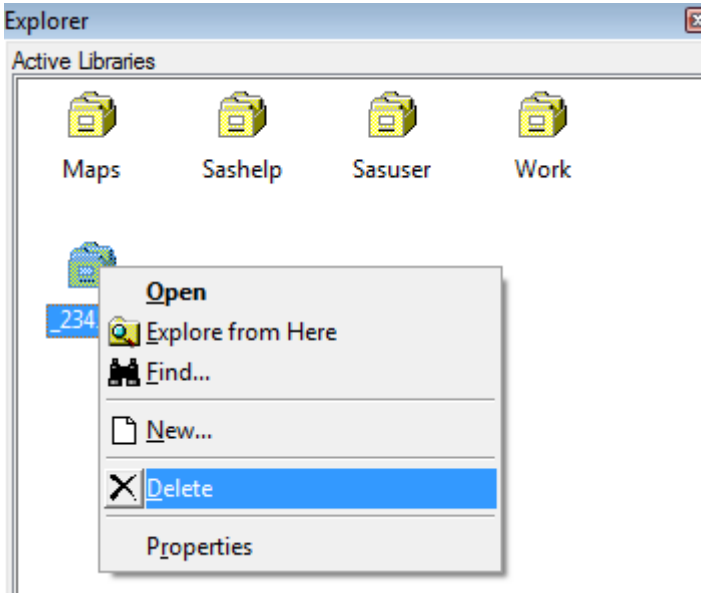

- Základní knihovny jsou Maps, Sashelp, Sasuser a Work

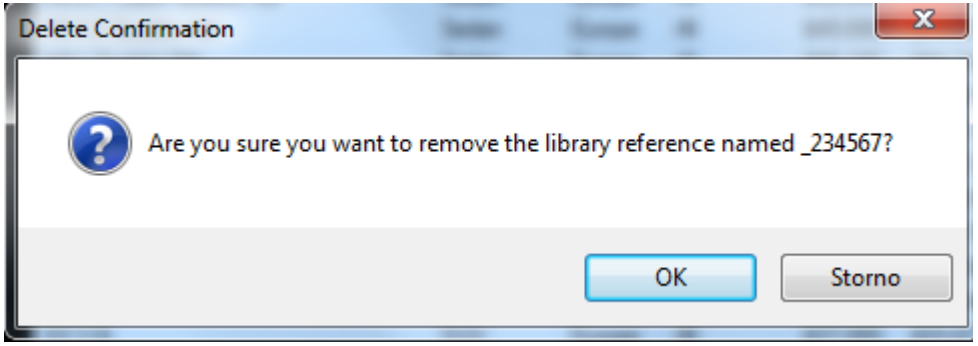
```
Libname _234567 "D:\dukumenty\Data_Mining_1";
```



# Libname

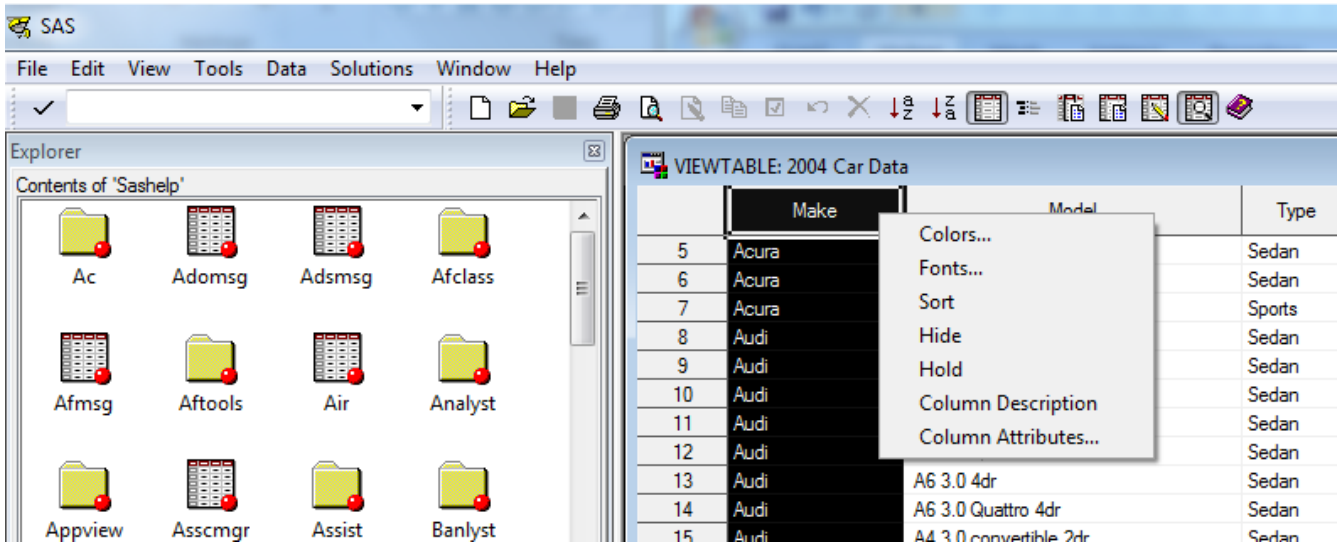
- „Knihovnu“ lze smazat ….smaže se pouze odkaz (na disku se nic fyzicky nemaže), přesto je třeba akci potvrdit.

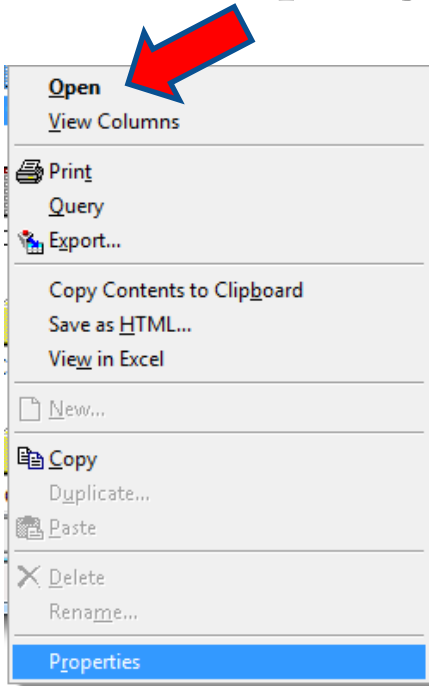


# Datové tabulky

- Datové tabulky v knihovně lze zobrazit pomocí ViewTable (dvojklik na tabulku nebo „Open“ v menu vyvolaném pravým tlačítkem myši nad vybranou tabulkou)

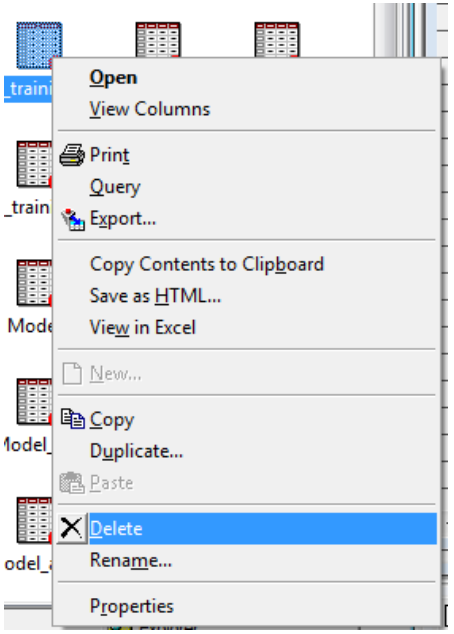


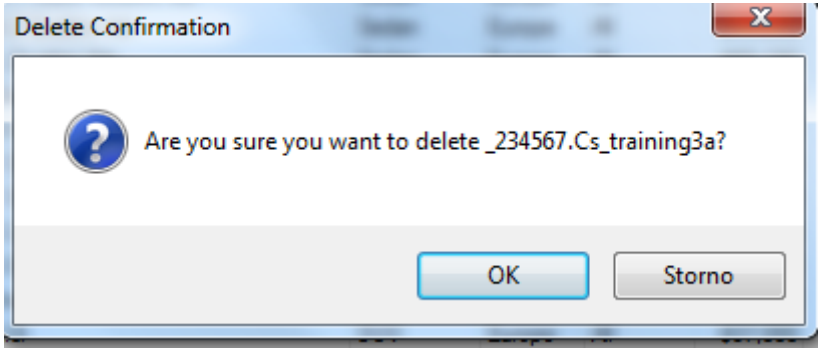




- Lze také zobrazit vlastnosti vybrané tabulky (obecné vlastnosti, seznam sloupců, jejich formáty,...)

# Datové tabulky

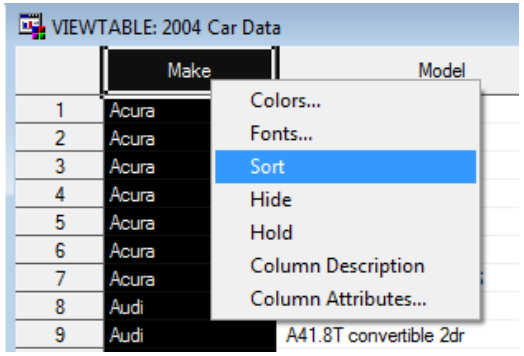
- Tabulku lze kopírovat do schránky a následně uložit (Paste) do jiné knihovny.
- Lze také (v rámci dané knihovny) provést duplikaci nebo přejmenování tabulky.
- Tabulku lze smazat .... je třeba akci potvrdit
  - Po potvrzení se tabulka fyzicky z disku smaže!!!

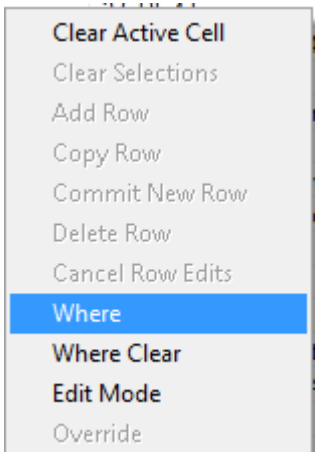


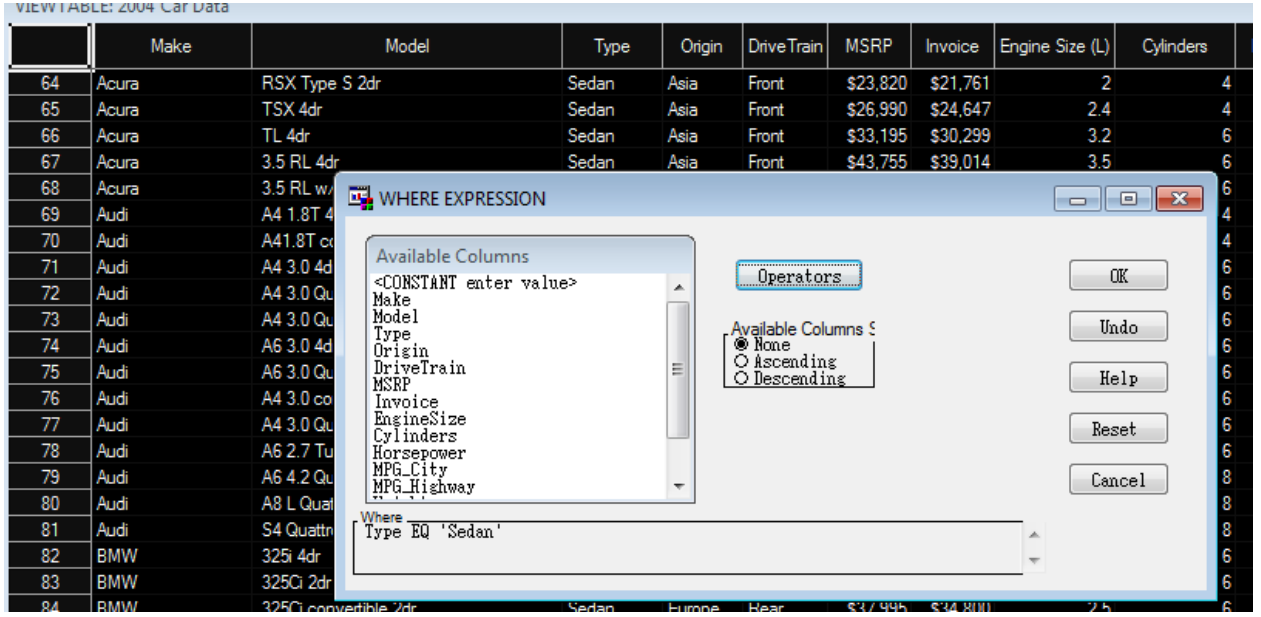



# Datové tabulky

- V rámci ViewTable lze provádět např. setřídění podle vybraného sloupce.
- Také lze data filtrovat pomocí Where filtru (vyvolá se stisknutím pravého tlačítka myši).

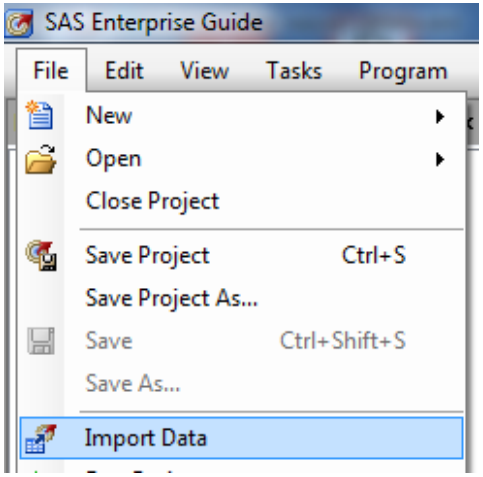


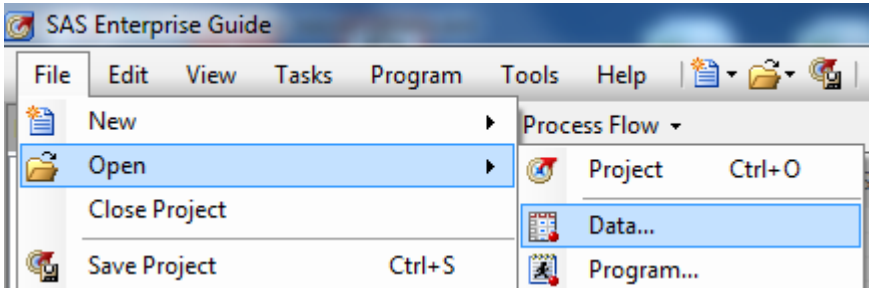



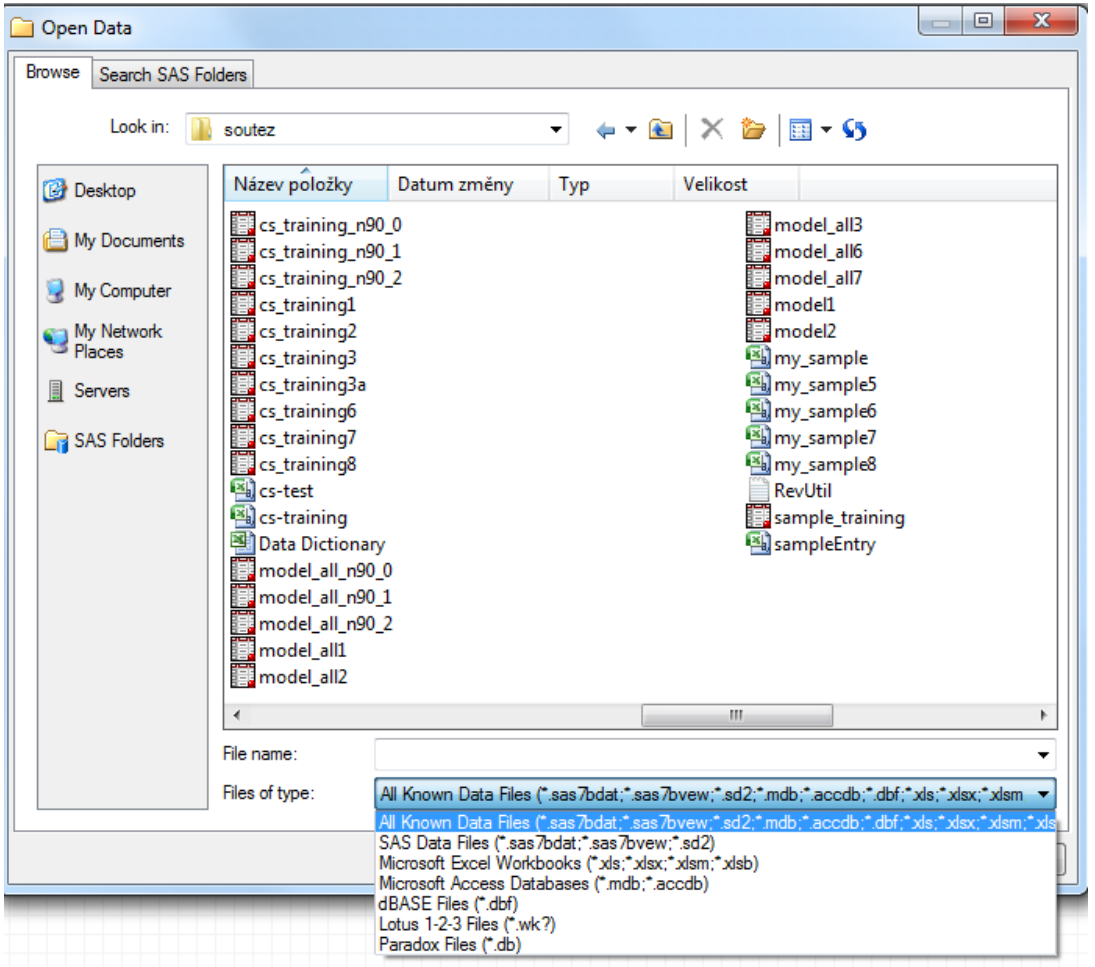



# Import v SAS EG

- Pomocí File - Open – Data
- nebo File – Import Data

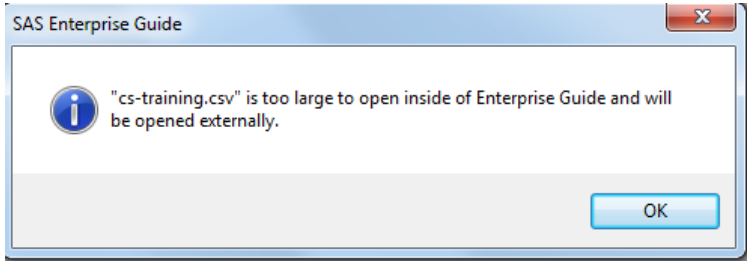


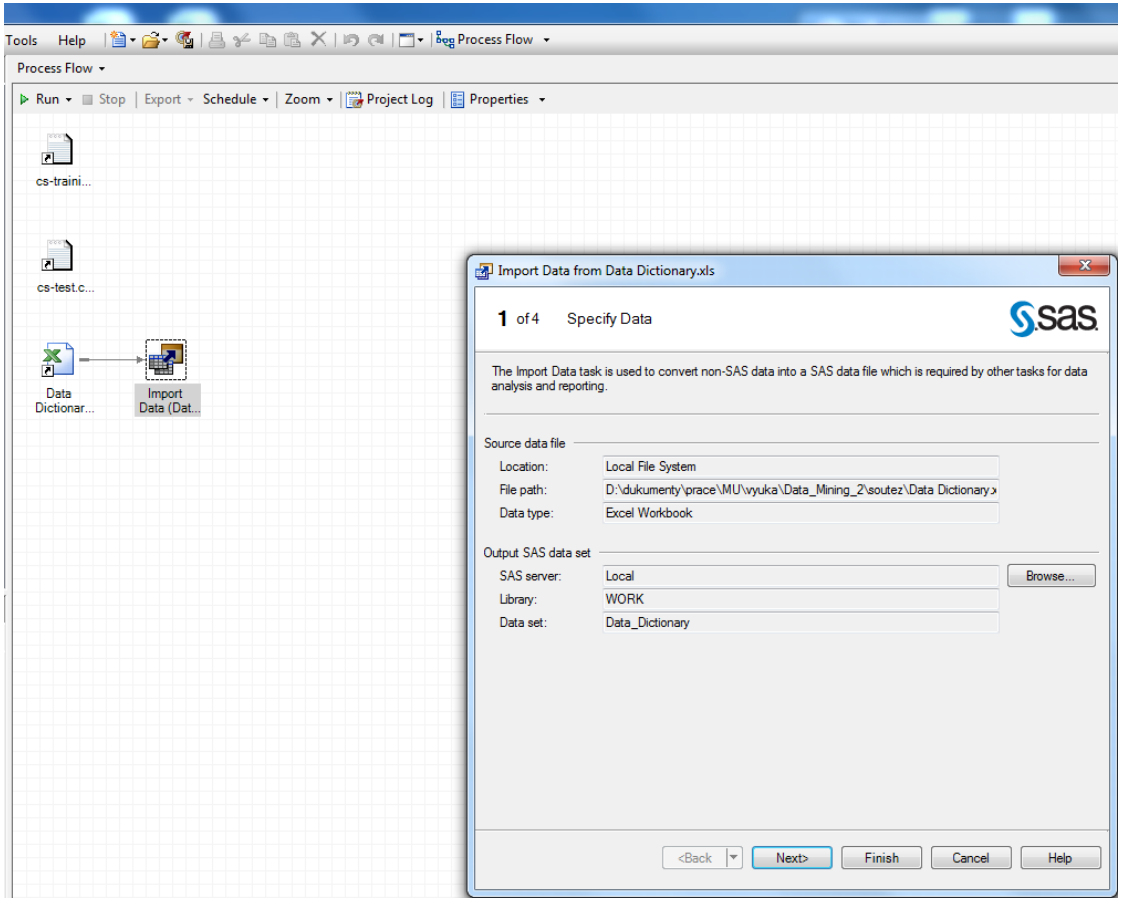



# Import v SAS EG

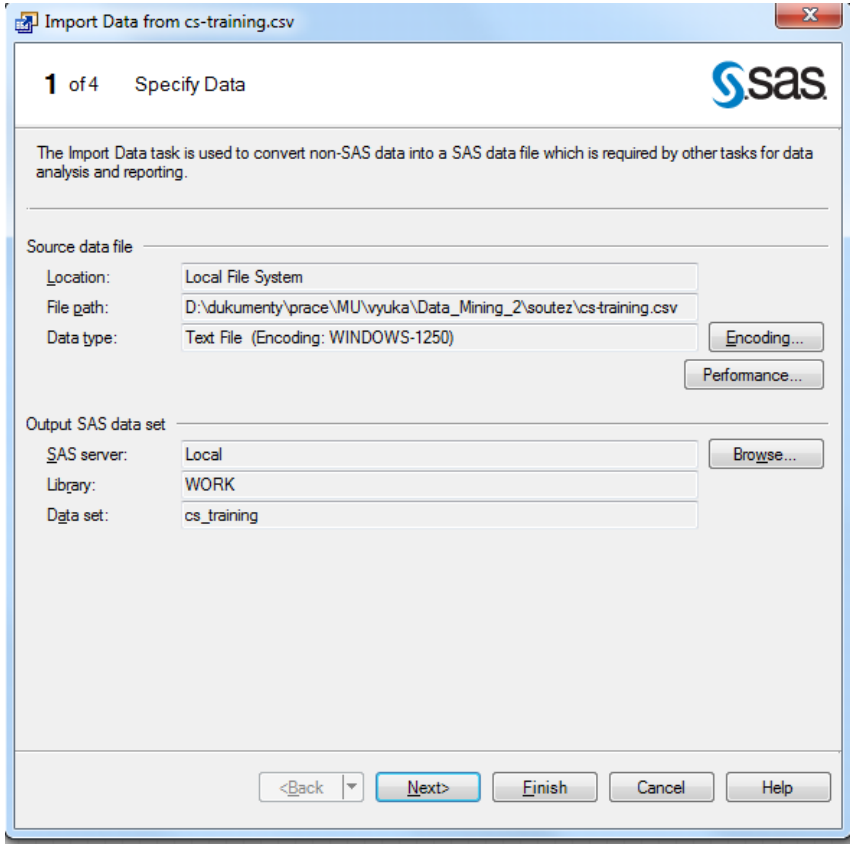
- „Open data“ neumí načíst příliš velká data (otevře pomocí poznámkového bloku). Pro „malá“ data se spustí „Import data“ wizard.
- File – Import data funguje i pro velká data.

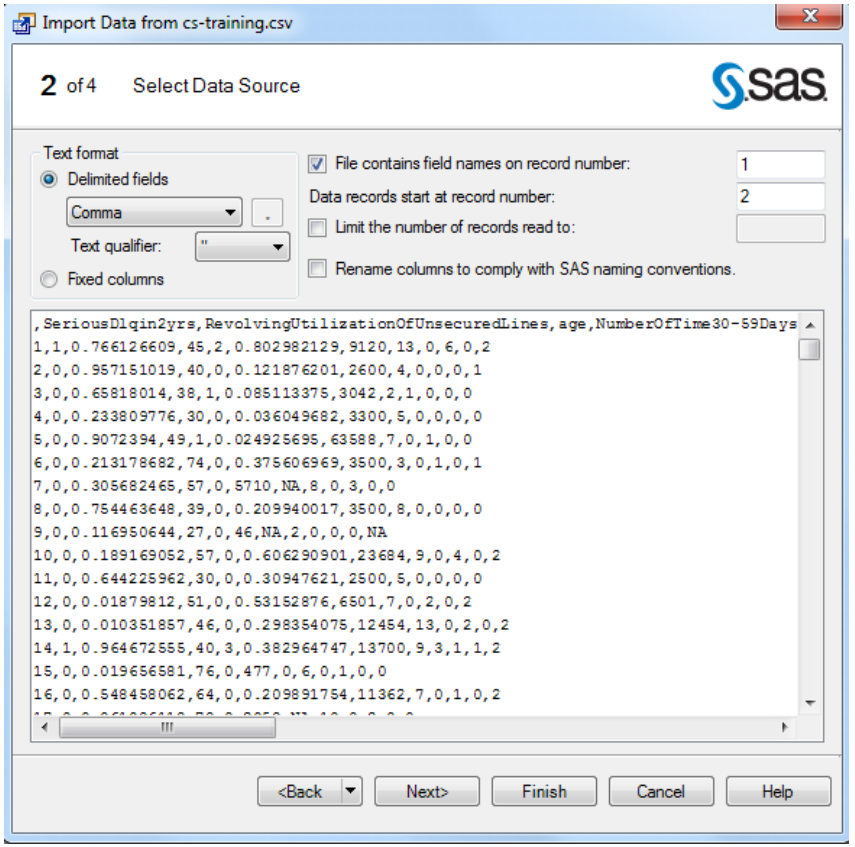


# Import v SAS EG

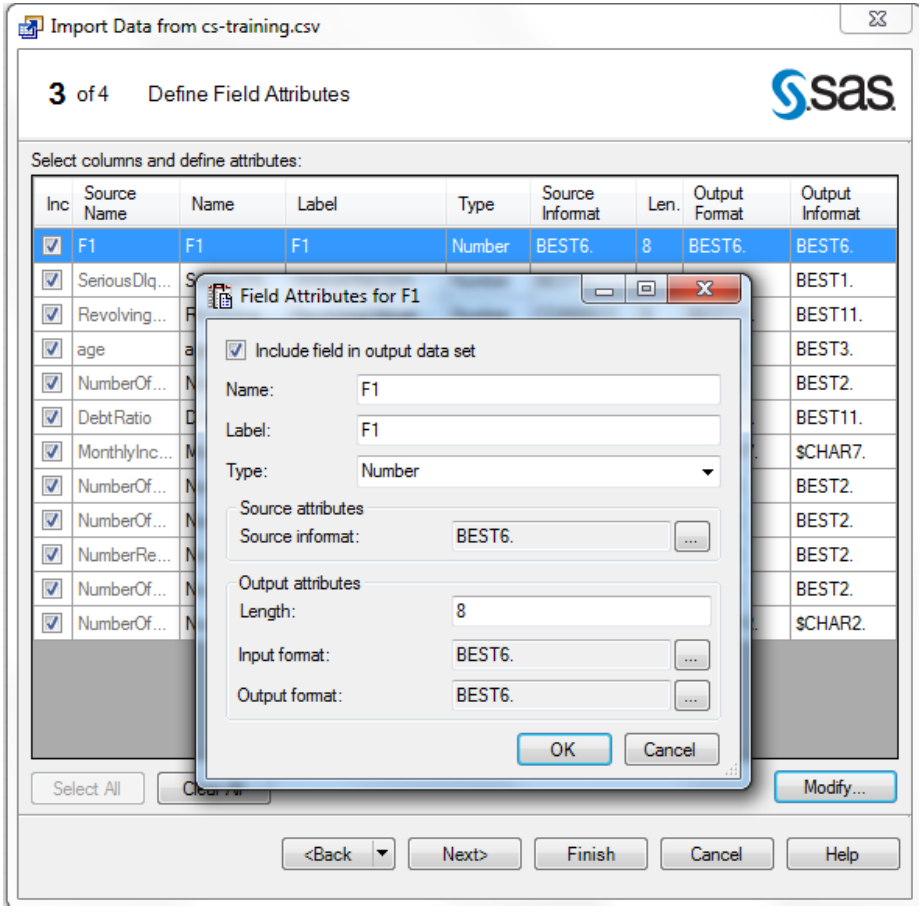
- Lze nastavit kódování, oddělovač sloupců (čárka, středník, tabulátor,...), info zda první řádek obsahuje názvy sloupců,...

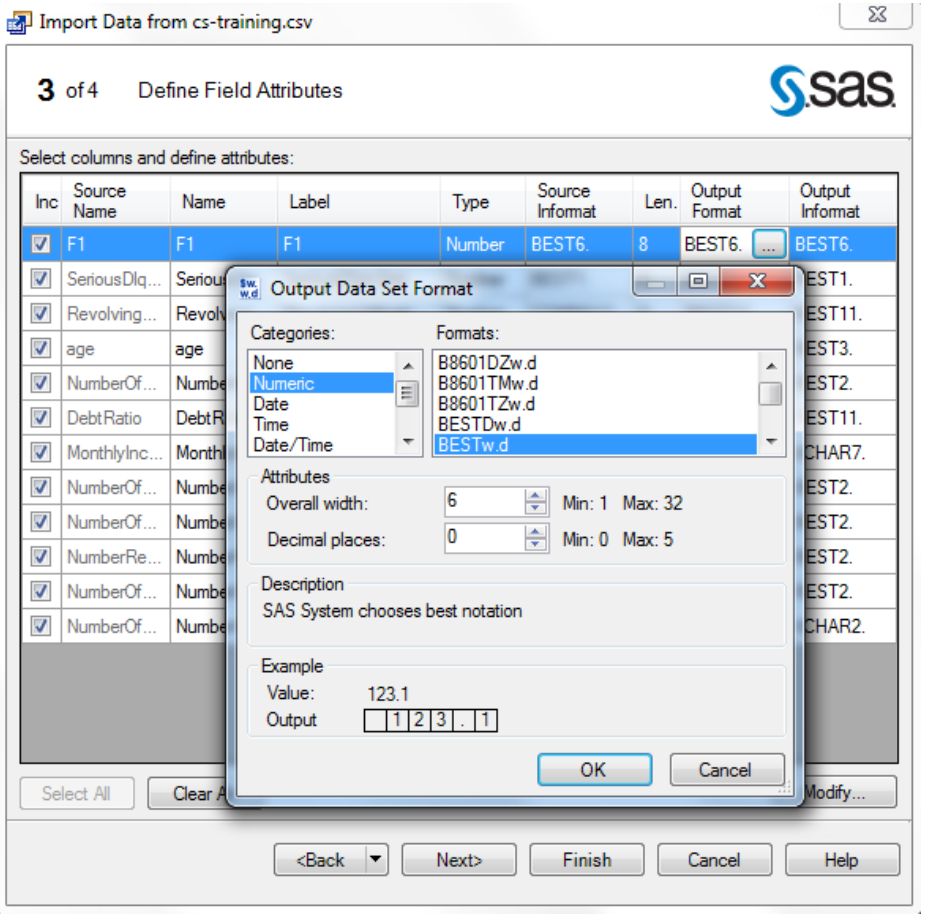


# Import v SAS EG

- Lze ručně nastavit názvy sloupců a jejich formáty.



# Úkoly

1. Vytvořte si svji knihovnu. Zkopírujte do ní tabulku Cars z knihovny Sashelp. Zjistěte jaké sloupce obsahuje, včetně formátů. Seřaďte tabulku podle sloupce Type (sestupně). Vyfiltrujte data jen na řádky s hodnotou „Truck“ ve sloupsi Type.

2. Importujte soubor cs-training.csv (pomocí SAS EG, Wizardu v programovacím prostředí i pomocí Data Stepu. Vytvořenou tabulku uložte (pomocí Data Stepu) v komprimované podobě a porovnejte velikosti tabulek na disku.

3. Pomocí ODS vytvořte html, rtf a pdf soubor obsahující značku, název modelu a výkon automobilů z tabulky sashelp.cars (where Type EQ ‘Truck’).

# Cvičení 3

# Proc SQL

- Mimo základní využití proc sql pro výběr definovaných podmnožin daných datových tabulek lze proc sql použít také pro:
  - vytváření nových tabulek
  - update existujících tabulek
  - úpravu existujících tabulek
  - mazání existujících tabulek
  - …

Více na:
http://support.sas.com/documentation/cdl/en/sqlproc/62086/HTML/default/viewer.htm#a001384710.htm

• With the SET clause, you assign values to columns by name. The columns can appear in any order in the SET clause. The following INSERT statement uses multiple SET clauses to add two rows to NEWCOUNTRIES:

```
proc sql;
insert into sql.newcountries
set    name='Bangladesh',
       capital='Dhaka',
       population=126391060
set    name='Japan',
       capital='Tokyo',
       population=126352003;
quit;
```

http://support.sas.com/documentation/cdl/en/sqlproc/62086/HTML/default/viewer.htm#a001387930.htm

- With the VALUES clause, you assign values to a column by position. The following INSERT statement uses multiple VALUES clauses to add rows to NEWCOUNTRIES.

```
proc sql;
 insert into sql.newcountries
       values ('Pakistan', 'Islamabad', 123060000, ., ' ', .)
       values ('Nigeria', 'Lagos', 99062000, ., ' ', .);
quit;
```

http://support.sas.com/documentation/cdl/en/sqlproc/62086/HTML/default/viewer.htm#a001387930.htm

- You can insert the rows from a query result into a table. The following query returns rows for large countries (over 130 million in population) from the COUNTRIES table. The INSERT statement adds the data to the empty table NEWCOUNTRIES, which was created earlier in "Creating Tables Like an Existing Table":

```
proc sql;
 create table sql.newcountries
like sql.countries;

proc sql;
insert into sql.newcountries
 select * from sql.countries
 where population ge 130000000;
quit;
```

http://support.sas.com/documentation/cdl/en/sqlproc/62086/HTML/default/viewer.htm#a001387930.htm

# Úkoly

1. Spojením (pomocí proc sql) tabulek customers a customerorders vytvořte tabulku obsahující typ zákazníka (customer type), celkový počet nákupů/kusů zboží, celkový objem prodeje, průměrnou prodejní cenu pro skupiny dané typem zákazníka a seřazené sestupně podle celkového objemu prodeje. Názvy nových sloupců opatřete vhodným labelem a formát posledních dvou sloupců nastavte na dollar12.2.

2. viz 1, ale skupiny dané pomocí CustomerGroup a jen ty, které mají celkový objem prodeje >= 10.000.

3. Zjistěte kolik zákazníků z tabulky customers má nějaký záznam v tabulce customorders, kolik jich tam nemá žádný záznam a zda tabulka customers neobsahuje duplicity.

# Úkoly

4. Vypište prvních 5 záznamů tabulky customers.

5. Vytvořte tabulku obsahující všechny sloupce tabulky customers a obsahující klienty (jen unikátní záznamy), jejichž příjmení začíná písmenem „M“ a kteří podle údajů v customerorders nakoupili zboží s jednotkovou cenou v intervalu 100 – 150. Do takto vytvořené tabulky přidejte řádky splňující předchozí podmínky s tím rozdílem, že příjmení začíná písmenem „H“.

6. Vytvořte tabulku obsahující všechny údaje tabulky customorders a navíc sloupec, jehož hodnoty jsou definované takto:
   „high unit price“ pokud UnitPrice > 120
   „mid unit price“ pokud 40 <UnitPrice< =120
   „low unit price“ pokud 20 <UnitPrice< =40
   „very low unit price“  jinak

# Cvičení 4

# SAS functions nad CALL routiens

Seznam všech funkcí podle kategorie:

http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000245860.htm

Funkce substr:

**<variable=>SUBSTR(string, position<,length>)**

- If length is zero, a negative value, or larger than the length of the expression that remains in string after position, SAS extracts the remainder of the expression. SAS also sets _ERROR_ to 1 and prints a note to the log indicating that the length argument is invalid.
- If you omit length, SAS extracts the remainder of the expression.
- Více na: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000212264.htm

# SAS functions nad CALL routiens

Funkce find:

**FIND(string,substring<,startpos><,modifiers>)**

The FIND function searches string for the first occurrence of the specifiedsubstring, and returns the position of that substring. If the substring is not found in string, FIND returns a value of 0.

**string**...specifies a character constant, variable, or expression that will be searched for substrings.

**substring**...is a character constant, variable, or expression that specifies the substring of characters to search for in string.

**startpos**...is a numeric constant, variable, or expression with an integer value that specifies the position at which the search should start and the direction of the search.

# SAS functions nad CALL routiens

**FIND(string,substring<,startpos><,modifiers>)**

If startpos is not specified, FIND starts the search at the beginning of the stringand searches the string from left to right. If startpos is specified, the absolute value of startpos determines the position at which to start the search. The sign of startpos determines the direction of the search.

| Value of startpos | Action |
|---|---|
| greater than 0 | starts the search at position **startpos** and the direction of the search is to the right. If **startpos** is greater than the length of **string**, FIND returns a value of 0. |
| less than 0 | starts the search at position -**startpos** and the direction of the search is to the left. If -**startpos** is greater than the length of **string**, the search starts at the end of **string**. |
| equal to 0 | returns a value of 0. |

Více na: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002267763.htm

# SAS functions nad CALL routiens

**WEEKDAY(date)**

The WEEKDAY function produces an integer that represents the day of the week, where 1=Sunday, 2=Monday, …, 7=Saturday.

**INTCK(interval,start-from, increment,< 'alignment'>)**

Returns the count of the number of interval boundaries between two dates, two times, or two datetime values.

**TODAY()**

Returns the current date as a numeric SAS date value.

**FLOOR(argument)**

Returns the largest integer that is less than or equal to the argument

Více na: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000212868.htm

K řešení úkolů je jinak dostačující učební text k přednášce. V případě hlubšího zájmu viz:

Proc Sort: http://support.sas.com/documentation/cdl/en/proc/61895/HTML/default/viewer.htm#a000057941.htm

Proc Format: http://support.sas.com/documentation/cdl/en/proc/61895/HTML/default/viewer.htm#a000063536.htm

Data step: http://support.sas.com/documentation/cdl/en/basess/58133/HTML/default/viewer.htm#a001302699.htm

# Úkoly

1. Vypište (pomocí proc sql) křestní jméno a příjmení zákazníků z tabulky Customers
    a) jejichž příjmení obsahuje „oo“ (pomocí like i contains)
    b) jejichž příjmení má druhé a třetí písmeno „o“

   a výsledky porovnejte.

2. Vytvořte tabulku z tabulky Customers (proc sql), kde vytvoříte nový sloupec s kódem státu klienta (z CustomerAdrress2 pomocí funkcí substr, find). Následně nastavte délku tohoto sloupce na 2 a zjistěte úsporu diskového prostoru.

# Úkoly

3. Vytvořte tabulku (pomocí proc sort), která bude obsahovat údaje z tabulky sales a bude seřazená podle pohlaví (Gender)… tak, že nejprve budou uvedeni muži… a současně seřazená vzestupně podle příjmení (Last_Name).

4. Vypište (do rtf/pdf) vytvořenou tabulku z bodu 1 se sloupci Employee_ID, Gender, Salary a Country s vhodnými formáty sloupců (u sloupců Gender, Salary a Country vlastní formát (viz přednáška). U všech sloupců použijte popisky (labely) i hodnoty ve sloupcích v češtině.

5. Pomocí data stepu vytvořte tabulku obsahující prvních pět sloupců a řádky tabulky sales splňující podmínky: Gender = „M“, Salary > 30 000.

# Úkoly

6. Vytvořte tabulku z tabulky sales, ve které vzniknou nové sloupce:
   - odchylka od průměrného příjmu
   - rok narození
   - měsíc narození
   - den v týdnu příslušný datu narození (s českým názvem dne)
   - rok nástupu do firmy
   - měsíc nástupu do firmy
   - věk v letech (k aktuálnímu datu)
   - věk v letech k datu nástupu do firmy

# Cvičení 5

# SAS formats

**MONNAMEw. format**

Writes date values as the name of the month

The example table uses the input value of 16500, which is the SAS date value that corresponds to March 5, 2005.

| SAS Statement | Results |
|---|---|
| put date monname1.; | M |
| put date monname3.; | Mar |
| put date monname5.; | March |

Více na:

http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000201049.htm

# PUT function

**PUT(source, format.)**

Returns a value using a specified format.

Např.:

```
put(OrderDate,monname.) as order_month
Value_after_30_years =  put(Retirement, dollar12.2);
```

Více na: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000199354.htm

# Úkoly

1. Vytvořte formát …1=‘leden’, 2=‘únor’, other=‘ostatní’. Pomocí něj v tabulce Customerorders transformujte sloupec OdrerDate a vypište celkový úhrn tržeb (format dollar12.2) pro skupiny nově vytvořeného sloupce. Výpis seřaďte podle vypočteného úhrnu tržeb sestupně.

2. Vypište celkový úhrn tržeb v tabulce Customerorders pro jednotlivé měsíce v roce. Výpis bude obsahovat číslo měsíce, jeho anglický název a úhrn tržeb…seřazeno podle čísla měsíce vzestupně.

# Úkoly

3. Pomocí jednoho data stepu vytvořte dvě nové tabulky z tabulky Customers tak, že v první nové tabulce budou zákazníci s CustomerType = “inactive”, ve druhé nové tabulce budou zákazníci s CustomerType různým od “inactive”. Ve druhé tabulce současně vytvořte sloupec Type, který nabývá hodnoty “Club Member” pro CustomerID <2000 a hodnoty “Gold Club Member” jinak.

4. Z tabulky Customerorders vytvořte pomocí data stepu tabulku, která obsahuje nový sloupec s názvem Level. Jeho hodnoty jsou v každém řádku podmíněny hodnotou UnitPrice takto: Level = ‘Level I’ pro UnitPrice <=30, Level = ‘Level II’ pro 30<UnitPrice <=60, Level = ‘Level III’ pro 60<UnitPrice <=120 a Level = ‘Level IV’ pro UnitPrice > 120.

# Úkoly

5. Z tabulky USemps pomocí data stepu vytvořte tabulku obsahující sloupce EmployeeID , Salary, Investment, Value_after_30_years, Value_after_40_years a Value_after_50_years. Poslední tři sloupce (ve formátu dollar12.2) představují částku naspořenou po 30-ti, 40-ti a 50-ti letech, za předpokladu, že daný zaměstnanec ročně uloží 3% svého ročního příjmu (salary), nejvýše však 10000, a roční úroková míra je 4%.

6. Vytvořte tabulku (pomocí data stepu a array) z tabulky z bodu 5, ve které budou poslední tři sloupce vyjadřovat potenciální měsíční výplatu penze po dobu pěti let po ukončení spoření.

7. Z tabulky Customerorders vypište CustomerID, datum prvního nákupu (příslušející k danému CustomerID), datum posledního nákupu (příslušející k danému CustomerID) a počet dnů mezi těmito daty (tabulku vhodně setřiďte, pak použijte first. a last.).

# Cvičení 6

# SAS functions nad CALL routiens

**LOG10(argument)**

Vrací dekadický logaritmus argumentu.

**CATS(string-1 <, …, string-n>)**

Odstraní mezery na začátcích a koncích zadaných řetězců a vrátí jejich spojení do jednoho řetězce (související funkce: CAT, CATT a CATX).

Více na:

http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000245910.htm
http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002256540.htm

# Změna pořadí sloupců

Any of the following statements may be used to change the order of variables in the program data vector:

ATTRIB, ARRAY, FORMAT, INFORMAT, LENGTH, and RETAIN.

Např.

```
data dm1.annual_orders1;
retain customer_ID mesic1-mesic12;
set dm1.annual_orders;
run;
```

Více na:

http://www.repole.com/dinosaur/
http://www.repole.com/dinosaur/reordervars.html
http://analytics.ncsu.edu/sesug/2002/PS12.pdf

# Úkoly

1. Předpokládejte, že je prosinec roku 2001. Máte za úkol určit roční bonus zaměstnanců (tabulka USemp). Za každý započatý rok přísluší zaměstnanci $50, nejvýše však $500. Služebně nejstarší zaměstnanci každého oddělení (EmployeeDepartment) dostanou navíc $100.

2. Vytvořte tabulku z tabulky Cars, ve které vzniknou nové sloupce obsahující počet cifer všech numerických sloupců tabulky Cars

   (pomocí data stepu s využitím „array“ a „do“ cyklu, počet cifer je spodní celá část dekadického logaritmu +1).

# Úkoly

3. Vytvořte tabulky, které vzniknou z tabulek UScustomers a USnewcustomers :
   a) Spojením (concatenation)
   b) Proložením (interleaving)
   c) Setříděním tab. z bodu a).

   Výsledky porovnejte.

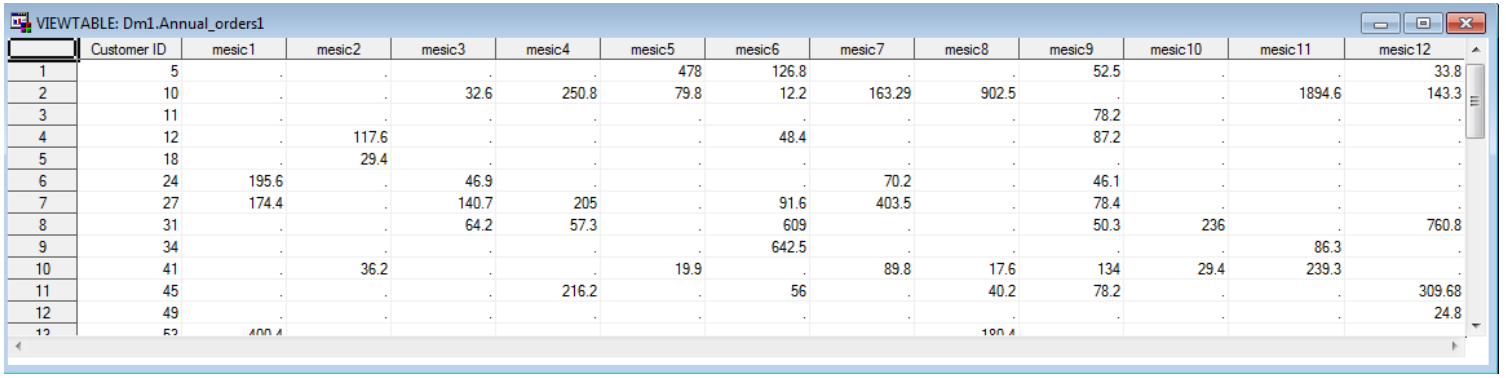
4. Vytvořte tabulky, které vzniknou z tabulek Customerorders a Customers:
   a) Sloučením (merge) přes CustomerID
   b) Sloučením (merge) přes CustomerID tak, aby výsledná tabulka obsahovala jen klienty, kteří učinili nějaký nákup.

   Výsledky porovnejte.

# Úkoly

5. Z tabulky Employee_donations vytvořte tabulku obsahující sloupec Employee_ID, sloupec obsahující kvartál darů a sloupec obsahující finanční výši darů (nejprve pomocí array, pak pomocí transpose).

6. Z tabulky Order_summary vytvořte tabulku obsahující sloupec Employee_ID a sloupce (s vhodnými názvy) obsahující výši nákupů/objednávek v jednotlivých měsících v roce (pomocí transpose).


VIEWTABLE: Dm1.Annual_orders1

| | Customer ID | mesic1 | mesic2 | mesic3 | mesic4 | mesic5 | mesic6 | mesic7 | mesic8 | mesic9 | mesic10 | mesic11 | mesic12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | . | . | . | . | 478 | 126.8 | . | . | 52.5 | . | . | 33.8 |
| 2 | 10 | . | . | 32.6 | 250.8 | 79.8 | 12.2 | 163.29 | 902.5 | . | . | 1894.6 | 143.3 |
| 3 | 11 | . | . | . | . | . | . | . | . | 78.2 | . | . | . |
| 4 | 12 | . | 117.6 | . | . | . | 48.4 | . | . | 87.2 | . | . | . |
| 5 | 18 | . | 29.4 | . | . | . | . | . | . | . | . | . | . |
| 6 | 24 | 195.6 | . | 46.9 | . | . | . | 70.2 | . | 46.1 | . | . | . |
| 7 | 27 | 174.4 | . | 140.7 | 205 | . | 91.6 | 403.5 | . | 78.4 | . | . | . |
| 8 | 31 | . | . | 64.2 | 57.3 | . | 609 | . | . | 50.3 | 236 | . | 760.8 |
| 9 | 34 | . | . | . | . | . | 642.5 | . | . | . | . | 86.3 | . |
| 10 | 41 | . | 36.2 | . | . | 19.9 | . | 89.8 | 17.6 | 134 | 29.4 | 239.3 | . |
| 11 | 45 | . | . | . | 216.2 | . | 56 | . | 40.2 | 78.2 | . | . | 309.68 |
| 12 | 49 | . | . | . | . | . | . | . | . | . | . | . | 24.8 |
| 13 | 53 | 400.4 | | | | | | | 180.4 | | | | |

# Úkoly

7. Z tabulky Customerorders vytvořte tabulku obsahující sloupec CustomerID a sloupce (s vhodnými názvy) vyjadřující měsíc nákupu (relativně vzhledem k datu prvního nákupu každého zákazníka, tj. den prvního nákupu a vše ve stejný kalendářní měsíc = OrderMonth01, následující kalendářní měsíc = OrderMonth02, a tak dále) a obsahující celkovou výši nákupů v jednotlivých měsících.

VIEWTABLE: Dm1.Orders_time

| | customerID | orderMonth01 | orderMonth02 | orderMonth03 | orderMonth04 | orderMonth05 | orderMonth06 | orderMonth07 | orderMonth08 | orderMonth09 | orderMonth10 | orderMonth11 | orderMonth12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 029858 | 520 | 677 | . | 181 | 158 | . | . | . | . | . | . | . |
| 2 | 030643 | 88 | . | . | . | . | . | . | . | . | . | . | . |
| 3 | 031116 | 432 | . | . | . | . | . | . | . | . | . | . | . |
| 4 | 032096 | 63 | . | 54 | 17 | 220 | . | 244 | 470 | . | 140 | 975 | 1348 |
| 5 | 033113 | 252 | . | . | . | . | . | . | . | . | . | . | . |
| 6 | 033312 | 31 | . | . | 721 | . | . | . | . | . | . | . | . |
| 7 | 033369 | 17 | 87 | . | . | . | . | . | . | . | . | . | . |
| 8 | 035668 | 143 | . | . | 230 | . | . | . | . | . | . | . | . |
| 9 | 035901 | 141 | . | 731 | . | . | . | 193 | . | . | . | 54 | 106 |
| 10 | 036324 | 223 | 532 | 120 | . | 28 | 72 | . | 82 | . | . | . | . |
| 11 | 036355 | 196 | . | . | . | . | . | . | . | . | . | . | . |
| 12 | 036732 | 84 | . | 620 | . | . | . | . | . | . | . | . | . |
| 13 | 037971 | 96 | 180 | | | | | | | | | | |

# Cvičení 7

# Úkoly

1. Z údajů v tabulce Sales, pro které název pozice(job_title) obsahuje řetězec „Rep“, vytvořte html/pdf/rtf obsahující kontingenční tabulku sloupců pohlaví(gender) a stát(country). Nastavte vhodný nadpis a potlačte výpis datumu. (PROC FREQ)

**Sales Rep Frequency Report**

The FREQ Procedure

| Frequency<br>Percent<br>Row Pct<br>Col Pct | Table of Gender by Country | | |
|---|---|---|---|
| | Country | | |
| Gender | AU | US | Total |
| F | 27<br>16.98<br>40.30<br>44.26 | 40<br>25.16<br>59.70<br>40.82 | 67<br>42.14 |
| M | 34<br>21.38<br>36.96<br>55.74 | 58<br>36.48<br>63.04<br>59.18 | 92<br>57.86 |
| Total | 61<br>38.36 | 98<br>61.64 | 159<br>100.00 |

2. Vytvořte tabulku Sales1 z tabulky Sales, ve které vznikne nový sloupec hire_age představující věk zaměstnance v okamžiku nástupu do zaměstnání. Vytvořte formát HireAge, který agreguje zadaný sloupec do kategorií low-<20, 20-<25 a 25-high. Následně vytvořte frekvenční tabulku pro sloupec hire_age formátovaný pomocí HireAge. (PROC FREQ)

The SAS System

The FREQ Procedure

| Hire_age | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| 1, pod 20 | 43 | 26.06 | 43 | 26.06 |
| 2, 20 - 25 | 68 | 41.21 | 111 | 67.27 |
| 3, nad 25 | 54 | 32.73 | 165 | 100.00 |

# Úkoly

3. Z tabulky Sales1 z úkolu 2
   a) vypište průměr (mean) a rozsah (range) příjmu (salary) pro všechny trojice hodnot sloupců pohlaví (gender), stát (country) a hire_age formátovaného pomocí HireAge z úkolu 2. (PROC MEANS)
   b) uložte výstup procedury (bez specifikace ukládaných údajů) do tabulky a porovnejte výstup bodu a) a b).

The SAS System

The MEANS Procedure

Analysis Variable : Salary

| Gender | Country | Hire_age | N Obs | Mean | Range |
|---|---|---|---|---|---|
| F | AU | 1. pod 20 | 10 | 27498.00 | 5600.00 |
| | | 2. 20 - 25 | 10 | 27849.00 | 5615.00 |
| | | 3. nad 25 | 7 | 27785.00 | 4695.00 |
| | US | 1. pod 20 | 7 | 28853.57 | 7055.00 |
| | | 2. 20 - 25 | 18 | 28285.83 | 6475.00 |
| | | 3. nad 25 | 16 | 31048.75 | 57825.00 |
| M | AU | 1. pod 20 | 9 | 35837.22 | 82510.00 |
| | | 2. 20 - 25 | 19 | 31463.95 | 61990.00 |
| | | 3. nad 25 | 8 | 28962.50 | 5975.00 |
| | US | 1. pod 20 | 17 | 40757.94 | 217905.00 |
| | | 2. 20 - 25 | 21 | 27750.95 | 7600.00 |
| | | 3. nad 25 | 23 | 32950.00 | 72380.00 |

4. Z tabulky Sales1 z úkolu 2 vytvořte kontingenční tabulku s absolutními četnostmi a řádkově a sloupcově podmíněnými relativními četnostmi. Řádková dimenze bude tvořena kartézským součinem hodnot sloupce hire_age formátovaného pomocí HireAge (včetně souhrnu (all)) a hodnot sloupce country. Sloupcová dimenze bude tvořena hodnotami sloupce gender. (PROC TABULATE)

The SAS System

| | | Gender | | | | | |
|---|---|---|---|---|---|---|---|
| | | F | | | M | | |
| | | N | RowPctN | ColPctN | N | RowPctN | ColPctN |
| Hire_age | Country | | | | | | |
| 1. pod 20 | AU | 10.00 | 52.63 | 14.71 | 9.00 | 47.37 | 9.28 |
| | US | 7.00 | 29.17 | 10.29 | 17.00 | 70.83 | 17.53 |
| 2. 20 - 25 | AU | 10.00 | 34.48 | 14.71 | 19.00 | 65.52 | 19.59 |
| | US | 18.00 | 46.15 | 26.47 | 21.00 | 53.85 | 21.65 |
| 3. nad 25 | AU | 7.00 | 46.67 | 10.29 | 8.00 | 53.33 | 8.25 |
| | US | 16.00 | 41.03 | 23.53 | 23.00 | 58.97 | 23.71 |
| All | AU | 27.00 | 42.86 | 39.71 | 36.00 | 57.14 | 37.11 |
| | US | 41.00 | 40.20 | 60.29 | 61.00 | 59.80 | 62.89 |

# Úkoly

5. Z tabulky Salesı z úkolu 2 vytvořte kontingenční tabulku, která bude obsahovat minimum, medián a maximum příjmu (salary). Řádková dimenze bude tvořena kartézským součinem hodnot sloupce hire_age formátovaného pomocí HireAge a hodnot sloupce country. Sloupcová dimenze bude tvořena hodnotami sloupce gender. U řádkové i sloupcové dimenze včetně všech souhrnů („all“). To vše ve formátu pdf se stylem sasweb. (PROC TABULATE)

| | | Gender | | | | | | All | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | | | M | | | | | |
| | | Salary | | | Salary | | | Salary | | |
| | | Min | P50 | Max | Min | P50 | Max | Min | P50 | Max |
| Hire_age | Country | | | | | | | | | |
| 1. pod 20 | AU | 25185.00 | 27362.50 | 30785.00 | 25745.00 | 26780.00 | 108255.00 | 25185.00 | 26970.00 | 108255.00 |
| | US | 25930.00 | 28325.00 | 32985.00 | 25285.00 | 27325.00 | 243190.00 | 25285.00 | 27400.00 | 243190.00 |
| | All | 25185.00 | 27465.00 | 32985.00 | 25285.00 | 27227.50 | 243190.00 | 25185.00 | 27260.00 | 243190.00 |
| 2. 20 - 25 | Country | | | | | | | | | |
| | AU | 25275.00 | 27445.00 | 30890.00 | 25985.00 | 27115.00 | 87975.00 | 25275.00 | 27440.00 | 87975.00 |
| | US | 25390.00 | 28132.50 | 31865.00 | 25125.00 | 27100.00 | 32725.00 | 25125.00 | 27425.00 | 32725.00 |
| | All | 25275.00 | 27742.50 | 31865.00 | 25125.00 | 27107.50 | 87975.00 | 25125.00 | 27432.50 | 87975.00 |
| 3. nad 25 | Country | | | | | | | | | |
| | AU | 25795.00 | 26850.00 | 30490.00 | 26515.00 | 28495.00 | 32490.00 | 25795.00 | 28480.00 | 32490.00 |
| | US | 25680.00 | 27510.00 | 83505.00 | 22710.00 | 27410.00 | 95090.00 | 22710.00 | 27460.00 | 95090.00 |
| | All | 25680.00 | 27460.00 | 83505.00 | 22710.00 | 27485.00 | 95090.00 | 22710.00 | 27472.50 | 95090.00 |
| All | Country | | | | | | | | | |
| | AU | 25185.00 | 27440.00 | 30890.00 | 25745.00 | 27165.00 | 108255.00 | 25185.00 | 27260.00 | 108255.00 |
| | US | 25390.00 | 28010.00 | 83505.00 | 22710.00 | 27260.00 | 243190.00 | 22710.00 | 27442.50 | 243190.00 |
| | All | 25185.00 | 27470.00 | 83505.00 | 22710.00 | 27240.00 | 243190.00 | 22710.00 | 27425.00 | 243190.00 |

# Úkoly

6. Analyzujte (zajímá nás základní sada popisných statistik, test pro charakteristiku polohy, kvantily, odlehlá pozorování) sloupec salary z tabulky Sales. Vytvořte výstup ve formátu rtf se stylem sasweb. (PROC UNIVARIATE)

| Moments | | | |
|---|---|---|---|
| N | 165 | Sum Weights | 165 |
| Mean | 31160.1212 | Sum Observations | 5141420 |
| Std Deviation | 20082.6671 | Variance | 403313519 |
| Skewness | 8.16761992 | Kurtosis | 78.5622611 |
| Uncorrected SS | 2.26351E11 | Corrected SS | 6.61434E10 |
| Coeff Variation | 64.4499006 | Std Error Mean | 1563.43352 |

| Basic Statistical Measures | | | |
|---|---|---|---|
| Location | | Variability | |
| Mean | 31160.12 | Std Deviation | 20083 |
| Median | 27425.00 | Variance | 403313519 |
| Mode | 26600.00 | Range | 220480 |
| | | Interquartile Range | 2825 |

| Tests for Location: Mu0=0 | | | | |
|---|---|---|---|---|
| Test | Statistic | | p Value | |
| Student's t | t | 19.93057 | Pr > \|t\| | <.0001 |
| Sign | M | 82.5 | Pr >= \|M\| | <.0001 |
| Signed Rank | S | 6847.5 | Pr >= \|S\| | <.0001 |

| Quantiles (Definition 5) | |
|---|---|
| Quantile | Estimate |
| 100% Max | 243190 |
| 99% | 108255 |
| 95% | 32985 |
| 90% | 31750 |
| 75% Q3 | 29385 |
| 50% Median | 27425 |
| 25% Q1 | 26560 |
| 10% | 25965 |
| 5% | 25680 |
| 1% | 25110 |
| 0% Min | 22710 |

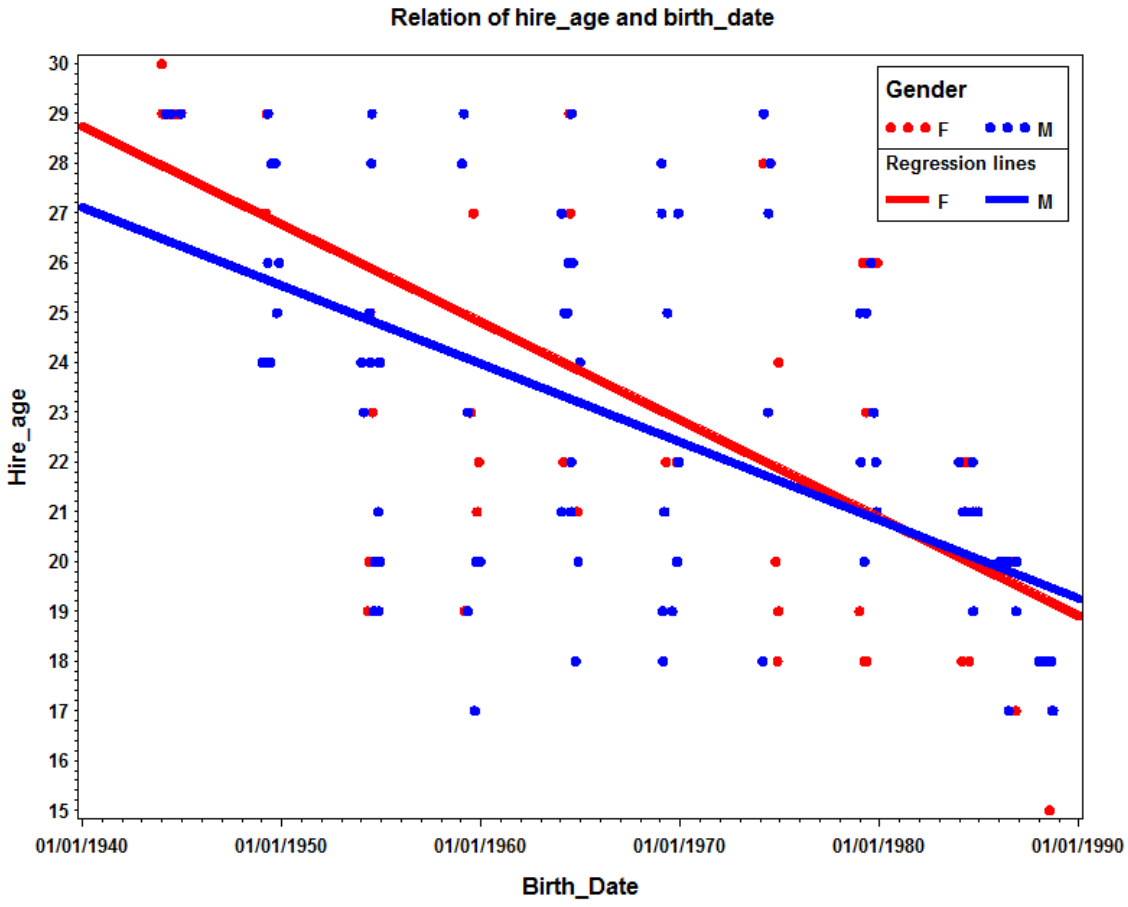
| Extreme Observations | | | |
|---|---|---|---|
| Lowest | | Highest | |
| Value | Obs | Value | Obs |
| 22710 | 131 | 84260 | 165 |
| 25110 | 111 | 87975 | 2 |
| 25125 | 104 | 95090 | 163 |
| 25185 | 49 | 108255 | 1 |
| 25275 | 50 | 243190 | 64 |

# Cvičení 8

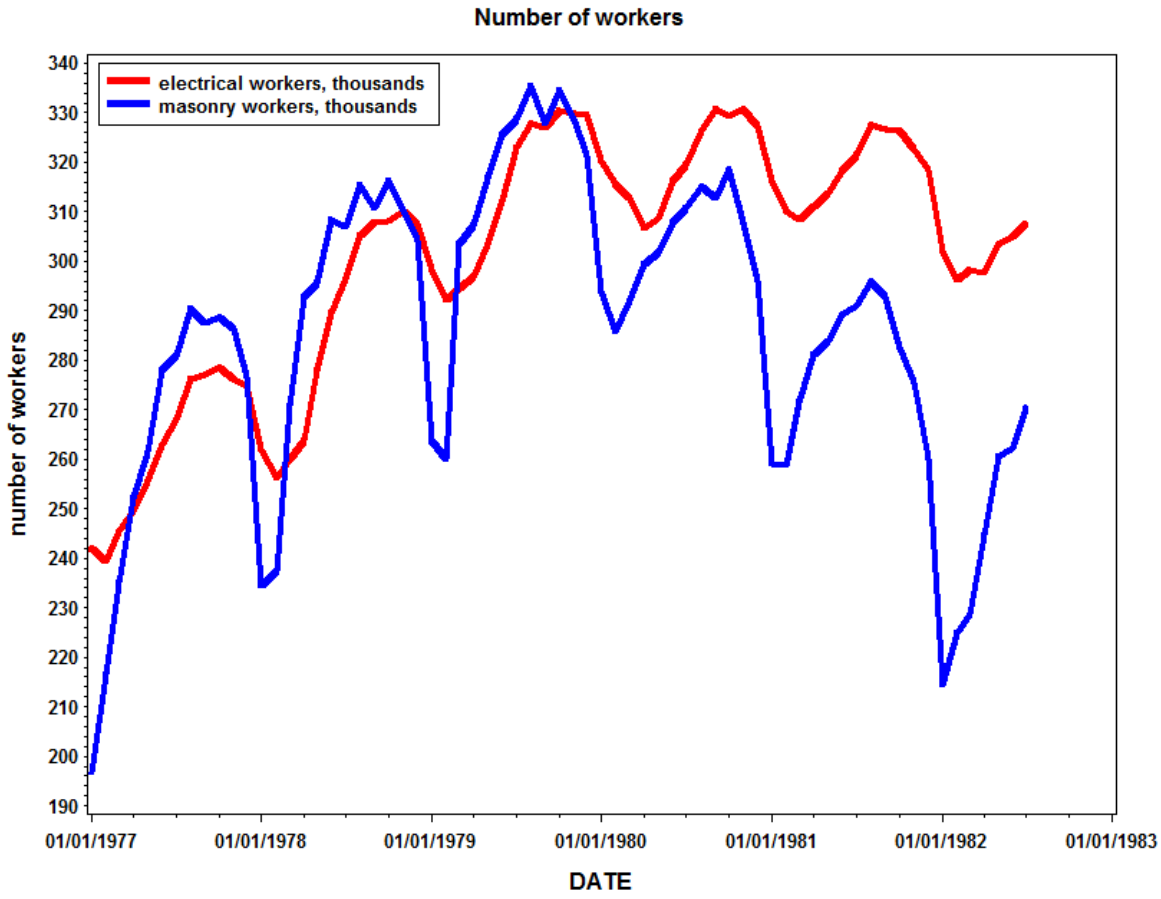
Detaily k řešení úkolů najdete v Helpu nebo např. na:

- http://www2.sas.com/proceedings/forum2007/163-2007.pdf
- http://support.sas.com/documentation/cdl/en/graphref/63022/HTML/default/viewer.htm#legendchap.htm
- http://www2.stat.unibo.it/manualisas/gref/c06.pdf
- http://www.nesug.org/proceedings/nesug08/np/np05.pdf
- http://support.sas.com/sassamples/graphgallery/PROC_GMAP.html
- http://support.sas.com/documentation/cdl/en/graphref/63022/HTML/default/viewer.htm#a0000729027.htm

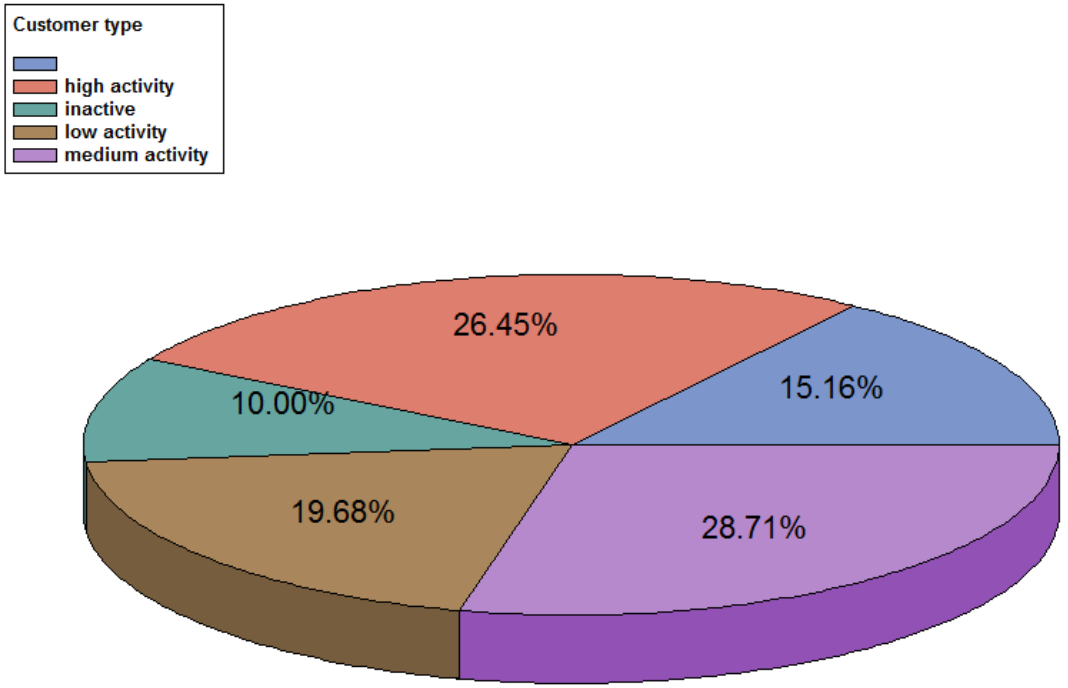
1. Z údajů v tabulce **Sales1** (cviční 7, úkol 2), vytvořte bodový graf závislosti **hire_age** na **birth_date** s rozlišením pohlaví (**gender**). Graf doplňte o regresní přímky a upravte vzhled podle vzoru (PROC GPLOT)… formát x-ové osy mmddyy10., tloušťka reg. přímek = 5, font popisu os i legendy = (arial bold, výška 12 bodů, resp. 10 bodů u „regression lines“), font hodnot na osách a hodnot v legendě= (arial bold, výška 10 bodů), výška nadpisu = 12 bodů.



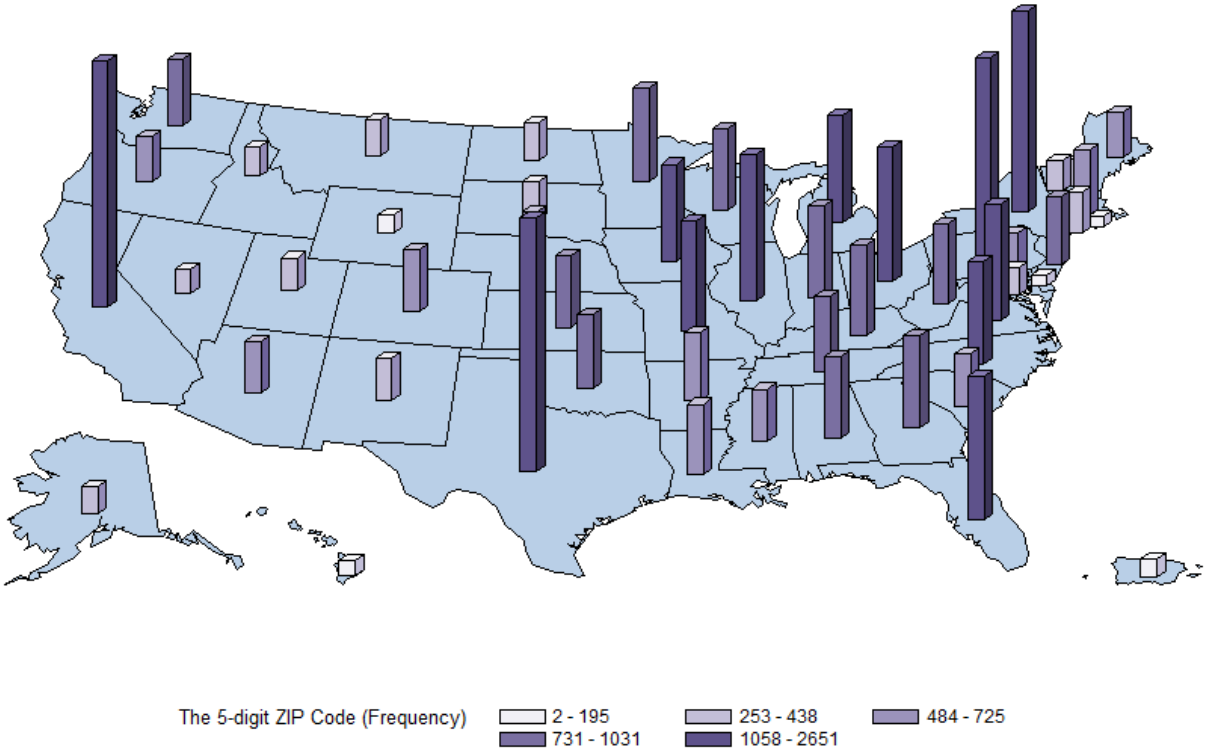
2. Z údajů v tabulce **Sashelp.workers** vytvořte graf počtu elektrikářů (**electric**) a počtu zedníků (**masonry**) v čase(**date**). Upravte vzhled podle vzoru (PROC GPLOT s overlay)… formát x-ové osy mmddyy10., tloušťka křivek = 5, font popisu os = (arial bold, výška 12 bodů), font hodnot na osách a hodnot v legendě= (arial bold, výška 10 bodů), výška nadpisu = 12 bodů, offset legendy = 1%.



3. Z údajů v tabulce **Customers** vytvořte koláčový 3D graf relativního zastoupení typů zákazníků (**customertype**). Upravte vzhled podle vzoru (PROC GCHART)… výška hodot v grafu =12 bodů, font hodnot v legendě= (arial bold, výška 10 bodů), font nadpisu v legendě= (arial bold, výška 10 bodů), offset legendy = 1%.

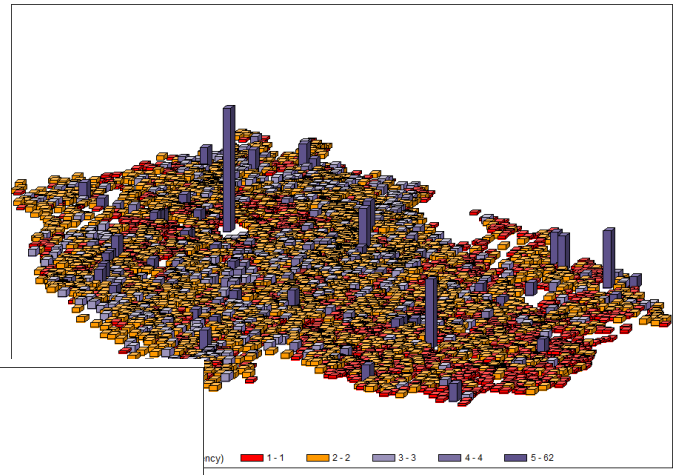


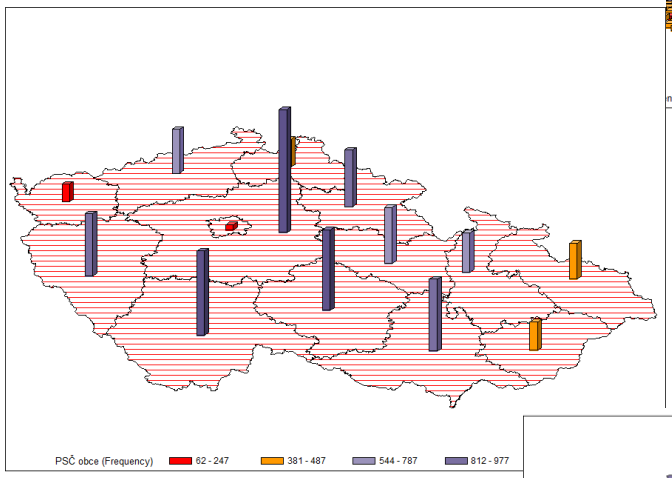
4. Z údajů v tabulce **sashelp.zipcode** a s využitím tabulky **maps.us**, vytvořte kartodiagram zobrazující počet zip kódů v jednotlivých státech USA. Barevnost sloupců uvažujte v 5-ti úrovních (levels=5) a výšku sloupce zobrazte relativně k nulovému počtu, ne k minimálnímu (relzero). (PROC GMAP).

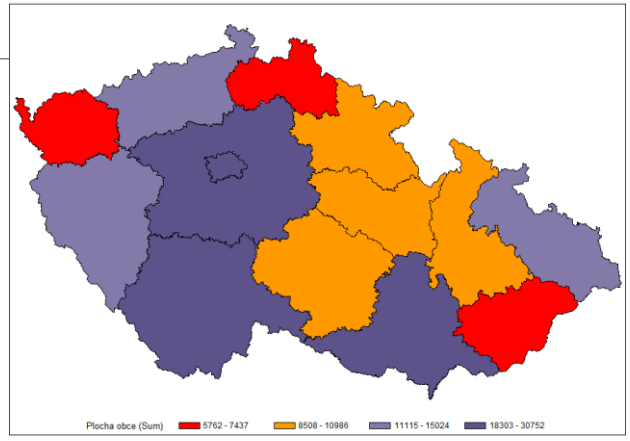


5. Z údajů v tabulce **czdata** a s využitím tabulky **czkraj_map**, vytvořte kartodiagram/kartogram zobrazující:
   a) počet psč kódů v jednotlivých obcích ČR
   b) počet psč kódů v jednotlivých krajích ČR
   c) součet ploch obcí v jednotlivých krajích ČR.

Barevnost sloupců uvažujte v 5-ti úrovních (levels=5) a výšku sloupce zobrazte relativně k nulovému počtu, ne k minimálnímu (relzero). (PROC GMAP).



# Cvičení 9

1. Vygenerujte data pro cvičení pomocí **gen_data_reg.sas**. Následně pro tabulku **fitness** vytvořte pdf report (použijte style=journal) obsahující, mimo jiné, korelační koeficienty sloupce **Oxygen_Consumption** se všemi ostatními číselnými sloupci seřazené v absolutní hodnotě od největšího po nejmenší. Současně vytvořte bodové grafy závislosti **Oxygen_Consumption** na všech ostatních číselných proměnných. Nadpis (title) nastavte např. na „ Correlations and Scatter Plots with Oxygen_Consumption“ (PROC CORR).

***Correlations and Scatter Plots with Oxygen_Consumption***

***The CORR Procedure***

| | | |
|---|---|---|
| *1 With Variables:* | Oxygen_Consumption | |
| *7 Variables:* | RunTime Performance | Age Weight Run_Pulse Rest_Pulse Maximum_Pulse |

*Simple Statistics*

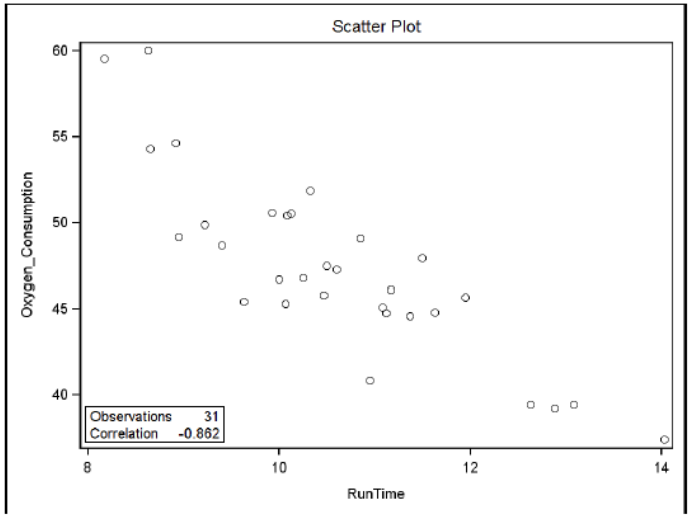
| *Variable* | *N* | *Mean* | *Std Dev* | *Sum* | *Minimum* | *Maximum* |
|---|---|---|---|---|---|---|
| *Oxygen_Consumption* | 31 | 47.37581 | 5.32777 | 1469 | 37.39000 | 60.06000 |
| *RunTime* | 31 | 10.58613 | 1.38741 | 328.17000 | 8.17000 | 14.03000 |
| *Age* | 31 | 47.67742 | 5.26236 | 1478 | 38.00000 | 57.00000 |
| *Weight* | 31 | 77.44452 | 8.32857 | 2401 | 59.08000 | 91.63000 |
| *Run_Pulse* | 31 | 169.64516 | 10.25199 | 5259 | 146.00000 | 186.00000 |
| *Rest_Pulse* | 31 | 53.45161 | 7.61944 | 1657 | 40.00000 | 70.00000 |
| *Maximum_Pulse* | 31 | 173.77419 | 9.16410 | 5387 | 155.00000 | 192.00000 |
| *Performance* | 31 | 56.64516 | 18.32584 | 1756 | 20.00000 | 94.00000 |

*Pearson Correlation Coefficients, N = 31*
*Prob > |r| under H0: Rho=0*

| | RunTime | Performance | Rest_Pulse | Run_Pulse | Age | Maximum_Pulse | Weight |
|---|---|---|---|---|---|---|---|
| *Oxygen_Consumption* | -0.86219 | 0.77890 | -0.39935 | -0.39808 | -0.31162 | -0.23677 | -0.16289 |
| | <.0001 | <.0001 | 0.0260 | 0.0266 | 0.0879 | 0.1997 | 0.3813 |

***Correlations and Scatter Plots with Oxygen_Consumption***

***The CORR Procedure***



2. Vytvořte html report (style=statistical) obsahující korelační matici všech číselných proměnných tabulky **fitness**, mimo **Oxygen_Consumption** , (včetně p-hodnot testu nulovosti korelačních koeficientů) a matici bodových grafů s histogramy na diagonále (PROC CORR).

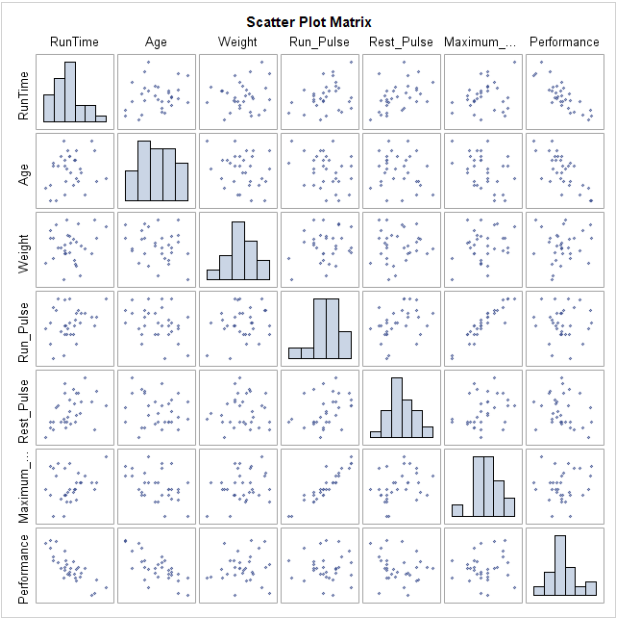
**Correlations and Scatter Plot Matrix of Fitness Predictors**

**The CORR Procedure**

**7 Variables:** RunTime Age Weight Run_Pulse Rest_Pulse Maximum_Pulse Performance

**Pearson Correlation Coefficients, N = 31**
**Prob > |r| under H0: Rho=0**

| | RunTime | Age | Weight | Run_Pulse | Rest_Pulse | Maximum_Pulse | Performance |
|---|---|---|---|---|---|---|---|
| **RunTime** | 1.00000 | 0.19523 | 0.14351 | 0.31365 | 0.45038 | 0.22610 | -0.82049 |
| | | 0.2926 | 0.4412 | 0.0858 | 0.0110 | 0.2213 | <.0001 |
| **Age** | 0.19523 | 1.00000 | -0.24050 | -0.31607 | -0.15087 | -0.41490 | -0.71257 |
| | 0.2926 | | 0.1925 | 0.0832 | 0.4178 | 0.0203 | <.0001 |
| **Weight** | 0.14351 | -0.24050 | 1.00000 | 0.18152 | 0.04397 | 0.24938 | 0.08974 |
| | 0.4412 | 0.1925 | | 0.3284 | 0.8143 | 0.1761 | 0.6312 |
| **Run_Pulse** | 0.31365 | -0.31607 | 0.18152 | 1.00000 | 0.35246 | 0.92975 | -0.02943 |
| | 0.0858 | 0.0832 | 0.3284 | | 0.0518 | <.0001 | 0.8751 |
| **Rest_Pulse** | 0.45038 | -0.15087 | 0.04397 | 0.35246 | 1.00000 | 0.30512 | -0.22560 |
| | 0.0110 | 0.4178 | 0.8143 | 0.0518 | | 0.0951 | 0.2224 |
| **Maximum_Pulse** | 0.22610 | -0.41490 | 0.24938 | 0.92975 | 0.30512 | 1.00000 | 0.09002 |
| | 0.2213 | 0.0203 | 0.1761 | <.0001 | 0.0951 | | 0.6301 |
| **Performance** | -0.82049 | -0.71257 | 0.08974 | -0.02943 | -0.22560 | 0.09002 | 1.00000 |
| | <.0001 | <.0001 | 0.6312 | 0.8751 | 0.2224 | 0.6301 | |



3. Vytvořte regresní model popisující závislost **Oxygen_Consumption** na **RunTime** v tabulce **finess**. Vykreslete všechny grafy poskytující prostředí ods graphics (PROC REG).

3b. Vypište 100(1-α)% konfidenční limity pro jednotlivé predikované hodnoty a pro očekávané hodnoty závisle proměnné.

Predicting Oxygen_Consumption from RunTime

The REG Procedure
Model: MODEL1
Dependent Variable: Oxygen_Consumption

| | |
|---|---|
| Number of Observations Read | 31 |
| Number of Observations Used | 31 |

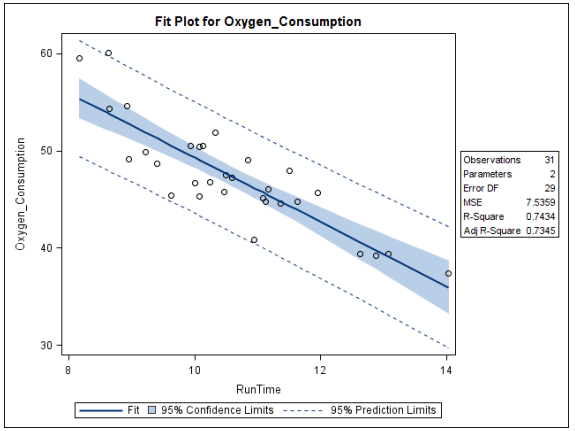
Analysis of Variance

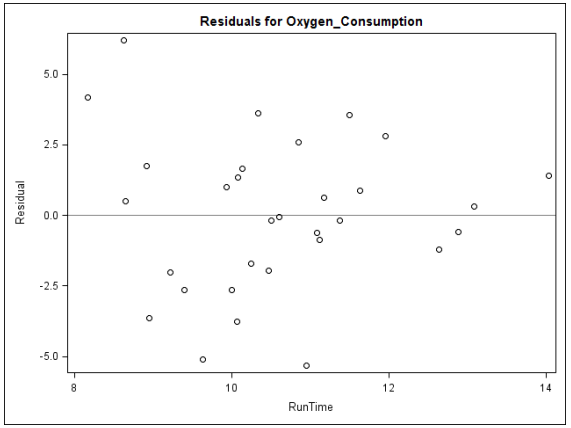
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 1 | 633,01458 | 633,01458 | 84,00 | <,0001 |
| Error | 29 | 218,53997 | 7,53586 | | |
| Corrected Total | 30 | 851,55455 | | | |

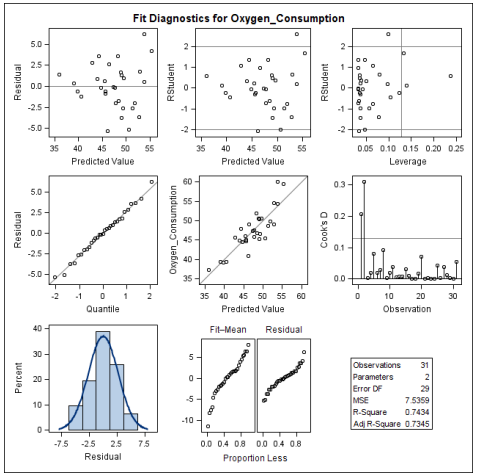
| | | | |
|---|---|---|---|
| Root MSE | 2,74515 | R-Square | 0,7434 |
| Dependent Mean | 47,37581 | Adj R-Sq | 0,7345 |
| Coeff Var | 5,79442 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 82,42494 | 3,85582 | 21,38 | <,0001 |
| RunTime | 1 | -3,31085 | 0,36124 | -9,17 | <,0001 |



4. Vytvořte tabulku **Need_Predictions** obsahující hodnoty 9 až 13. Spojte tuto tabulku s tabulkou **fitness**. Nad takto vzniklou tabulkou vytvořte regresní model popisující závislost **Oxygen_Consumption** na **RunTime**. Výstup má obsahovat, mimo jiné, predikovanou hodnotu a proměnnou **RunTime** (PROC REG).

4b. Vytvořte stejný model nad tabulkou **fitness** s tím, že regresní koeficienty uložíte do tabulky **Betas**. Následně, pomocí procedury SCORE, proveďte predikci pro hodnoty tabulky Need_Predictions a výsledek vypište (PROC SCORE).

Oxygen_Consumption=RunTime with Predicted Values

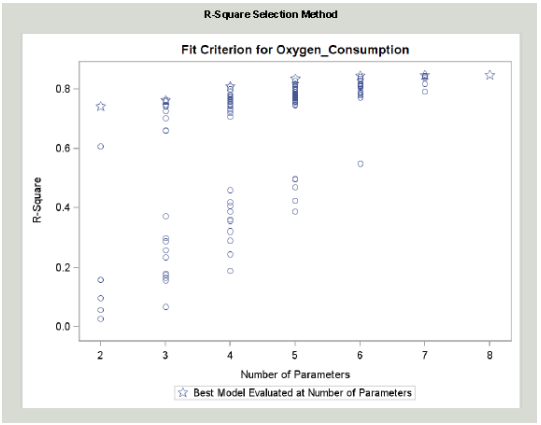
The REG Procedure
Model: MODEL1
Dependent Variable: Oxygen_Consumption

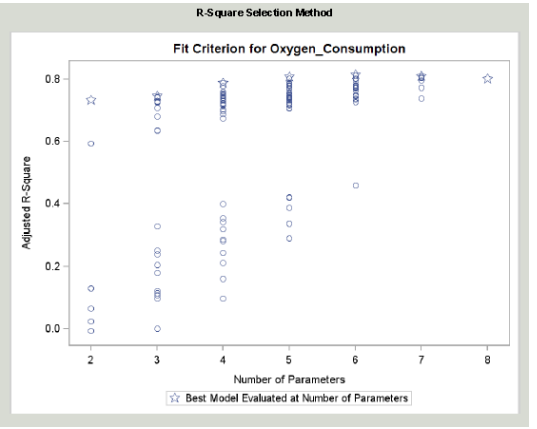
Output Statistics

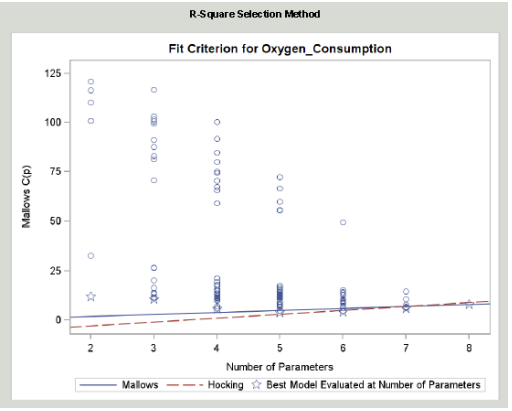
| Obs | Run Time | Dependent Variable | Predicted Value | Residual |
|---|---|---|---|---|
| 1 | 9,00 | . | 52,6272 | . |
| 2 | 10,00 | . | 49,3164 | . |
| 3 | 11,00 | . | 46,0055 | . |
| 4 | 12,00 | . | 42,6947 | . |
| 5 | 13,00 | . | 39,3838 | . |
| 6 | 8,17 | 59,5700 | 55,3753 | 4,1947 |
| 7 | 8,63 | 60,0600 | 53,8523 | 6,2077 |
| 8 | 8,65 | 54,3000 | 53,7860 | 0,5140 |
| 9 | 8,92 | 54,6300 | 52,8921 | 1,7379 |
| 10 | 8,95 | 49,1600 | 52,7928 | -3,6328 |
| 11 | 9,22 | 49,8700 | 51,8989 | -2,0289 |
| 12 | 9,40 | 48,6700 | 51,3029 | -2,6329 |
| 13 | 9,63 | 45,4400 | 50,5414 | -5,1014 |
| 14 | 9,93 | 50,5500 | 49,5482 | 1,0018 |
| 15 | 10,00 | 46,6700 | 49,3164 | -2,6464 |
| 16 | 10,07 | 45,3100 | 49,0846 | -3,7746 |
| 17 | 10,08 | 50,3900 | 49,0515 | 1,3385 |
| 18 | 10,13 | 50,5400 | 48,8860 | 1,6540 |
| 19 | 10,25 | 46,7700 | 48,4887 | -1,7187 |
| 20 | 10,33 | 51,8500 | 48,2238 | 3,6262 |
| 21 | 10,47 | 45,7900 | 47,7603 | -1,9703 |
| 22 | 10,50 | 47,4700 | 47,6610 | -0,1910 |
| 23 | 10,60 | 47,2700 | 47,3299 | -0,0599 |
| 24 | 10,85 | 49,0900 | 46,5022 | 2,5878 |
| 25 | 10,95 | 40,8400 | 46,1711 | -5,3311 |
| 26 | 11,08 | 45,1200 | 45,7407 | -0,6207 |
| 27 | 11,12 | 44,7500 | 45,6082 | -0,8582 |
| 28 | 11,17 | 46,0800 | 45,4427 | 0,6373 |
| 29 | 11,37 | 44,6100 | 44,7805 | -0,1705 |
| 30 | 11,50 | 47,9200 | 44,3501 | 3,5699 |
| 31 | 11,63 | 44,8100 | 43,9197 | 0,8903 |
| 32 | 11,95 | 45,6800 | 42,8602 | 2,8198 |
| 33 | 12,63 | 39,4100 | 40,6088 | -1,1988 |
| 34 | 12,88 | 39,2000 | 39,7811 | -0,5811 |
| 35 | 13,08 | 39,4400 | 39,1190 | 0,3210 |
| 36 | 14,03 | 37,3900 | 35,9736 | 1,4164 |

5. Vytvořte regresní model nad tabulkou **fitness** popisující závislost proměnné **oxygen_consumption** na proměnných **Performance**, **RunTime**, **Age**, **Weight**, **Run_Pulse**, **Rest_Pulse** a **Maximum_Pulse** tak, že uvážíte všechny možné kombinace vysvětlovaných proměnných.

| Model Index | Number in Model | R-Square | Adjusted R-Square | C(p) | Variables in Model |
|---|---|---|---|---|---|
| 92 | 4 | 0.7471 | 0.7082 | 17.4252 | Performance Weight Rest_Pulse Maximum_Pulse |
| 93 | 4 | 0.7462 | 0.7071 | 17.5698 | RunTime Weight Rest_Pulse Maximum_Pulse |
| 94 | 4 | 0.4979 | 0.4206 | 55.2965 | Age Weight Run_Pulse Rest_Pulse |
| 95 | 4 | 0.4960 | 0.4185 | 55.5812 | Age Run_Pulse Rest_Pulse Maximum_Pulse |
| 96 | 4 | 0.4686 | 0.3868 | 59.7486 | Age Weight Run_Pulse Maximum_Pulse |
| 97 | 4 | 0.4234 | 0.3347 | 66.6128 | Age Weight Rest_Pulse Maximum_Pulse |
| 98 | 4 | 0.3860 | 0.2916 | 72.2918 | Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 99 | 5 | 0.8469 | 0.8163 | 4.2598 | RunTime Age Weight Run_Pulse Maximum_Pulse |
| 100 | 5 | 0.8439 | 0.8127 | 4.7158 | Performance RunTime Weight Run_Pulse Maximum_Pulse |
| 101 | 5 | 0.8439 | 0.8127 | 4.7168 | Performance RunTime Age Run_Pulse Maximum_Pulse |
| 102 | 5 | 0.8356 | 0.8027 | 5.9783 | RunTime Age Run_Pulse Rest_Pulse Maximum_Pulse |
| 103 | 5 | 0.8356 | 0.8027 | 5.9856 | Performance Age Weight Run_Pulse Maximum_Pulse |
| 104 | 5 | 0.8293 | 0.7951 | 6.9446 | Performance RunTime Run_Pulse Rest_Pulse Maximum_Pulse |
| 105 | 5 | 0.8176 | 0.7811 | 8.7135 | Performance RunTime Age Weight Run_Pulse |
| 106 | 5 | 0.8167 | 0.7801 | 8.8473 | Performance RunTime Age Run_Pulse Rest_Pulse |
| 107 | 5 | 0.8162 | 0.7795 | 8.9266 | RunTime Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 108 | 5 | 0.8161 | 0.7794 | 8.9389 | RunTime Age Weight Run_Pulse Rest_Pulse |
| 109 | 5 | 0.8124 | 0.7748 | 9.5120 | Performance Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 110 | 5 | 0.8113 | 0.7736 | 9.6700 | Performance RunTime Weight Run_Pulse Rest_Pulse |
| 111 | 5 | 0.8096 | 0.7715 | 9.9341 | Performance Age Run_Pulse Rest_Pulse Maximum_Pulse |
| 112 | 5 | 0.8039 | 0.7646 | 10.8054 | Performance Age Weight Run_Pulse Rest_Pulse |
| 113 | 5 | 0.7911 | 0.7493 | 12.7457 | Performance RunTime Age Rest_Pulse Maximum_Pulse |
| 114 | 5 | 0.7904 | 0.7485 | 12.8462 | Performance RunTime Age Weight Maximum_Pulse |
| 115 | 5 | 0.7885 | 0.7462 | 13.1434 | RunTime Age Weight Rest_Pulse Maximum_Pulse |
| 116 | 5 | 0.7833 | 0.7400 | 13.9271 | Performance RunTime Weight Rest_Pulse Maximum_Pulse |
| 117 | 5 | 0.7801 | 0.7361 | 14.4150 | Performance RunTime Age Weight Rest_Pulse |
| 118 | 5 | 0.7730 | 0.7276 | 15.4964 | Performance Age Weight Rest_Pulse Maximum_Pulse |
| 119 | 5 | 0.5492 | 0.4590 | 49.5048 | Age Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 120 | 6 | 0.8483 | 0.8104 | 6.0492 | Performance RunTime Age Weight Run_Pulse Maximum_Pulse |
| 121 | 6 | 0.8475 | 0.8094 | 6.1758 | RunTime Age Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 122 | 6 | 0.8446 | 0.8057 | 6.6171 | Performance RunTime Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 123 | 6 | 0.8440 | 0.8049 | 6.7111 | Performance RunTime Age Run_Pulse Rest_Pulse Maximum_Pulse |
| 124 | 6 | 0.8373 | 0.7966 | 7.7279 | Performance Age Weight Run_Pulse Rest_Pulse Maximum_Pulse |
| 125 | 6 | 0.8181 | 0.7727 | 10.6357 | Performance RunTime Age Weight Run_Pulse Rest_Pulse |
| 126 | 6 | 0.7918 | 0.7398 | 14.6319 | Performance RunTime Age Weight Rest_Pulse Maximum_Pulse |
| 127 | 7 | 0.8486 | 0.8026 | 8.0000 | Performance RunTime Age Weight Run_Pulse Rest_Pulse Maximum_Pulse |



6. Vytvořte regresní model nad tabulkou **fitness** popisující závislost proměnné **oxygen_consumption** na proměnných **Performance**, **RunTime**, **Age**, **Weight**, **Run_Pulse**, **Rest_Pulse** a **Maximum_Pulse** tak, že postupně použijete metodu forward, backward a stepwise. Výsledky porovnejte.

# Cvičení 10

Detaily k řešení úkolů najdete v Helpu nebo např. na:

- http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_logistic_sect004.htm
- http://www.math.wpi.edu/saspdf/stat/chap39.pdf
- http://www.ats.ucla.edu/stat/sas/seminars/sas_logistic/logistic1.htm

1. Vygenerujte data pro cvičení pomocí **gen_data_reg.sas** (stačí tabulka **sales**). Pomocí data stepu vytvořte z tabulky **sales** tabulku **sales_inc**, ve které vznikne nový sloupec **IncLevel** překódováním hodnot sloupce **Income** (Low=1, Medium=2, High=3). Následně z hodnot tabulky **sales_inc** vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnné **Age**. (PROC LOGISTIC).
   - Pravděpodobnost jaké hodnoty proměnné Purchase jste modelovali?
   - Bylo splněno konvergenční kriterium pro odhad koeficientů?
   - Jaká je hodnota koeficientů?
   - Jaká je jejich statistická významnost?
   - Jaká je kvalita modelu (Somers'D)?

| | |
|---|---|
| Number of Observations Read | 431 |
| Number of Observations Used | 431 |

Response Profile

| Ordered Value | Purchase | Total Frequency |
|---|---|---|
| 1 | 0 | 269 |
| 2 | 1 | 162 |

Probability modeled is Purchase=0.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| AIC | 572,649 | 566,313 |
| SC | 576,715 | 574,445 |
| -2 Log L | 570,649 | 562,313 |

Testing Global Null Hypothesis: BETA=0

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 8,3365 | 1 | 0,0039 |
| Score | 8,2831 | 1 | 0,0040 |
| Wald | 8,1129 | 1 | 0,0044 |

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | 2,4682 | 0,6990 | 12,4671 | 0,0004 |
| Age | 1 | -0,0509 | 0,0179 | 8,1129 | 0,0044 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 54,3 | Somers' D | 0,136 |
| Percent Discordant | 40,8 | Gamma | 0,143 |
| Percent Tied | 4,9 | Tau-a | 0,064 |
| Pairs | 43578 | c | 0,568 |

2. Tentokrát vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnné **Age**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘1’. Navíc vykreslete ROC křivku. (PROC LOGISTIC).
   - Jak se změnily koeficienty?
   - Jak se změnily ostatní údaje popisující model?

| | |
|---|---|
| Number of Observations Read | 431 |
| Number of Observations Used | 431 |

**Response Profile**

| Ordered Value | Purchase | Total Frequency |
|---|---|---|
| 1 | 0 | 269 |
| 2 | 1 | 162 |

Probability modeled is Purchase=1.

**Model Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Model Fit Statistics**

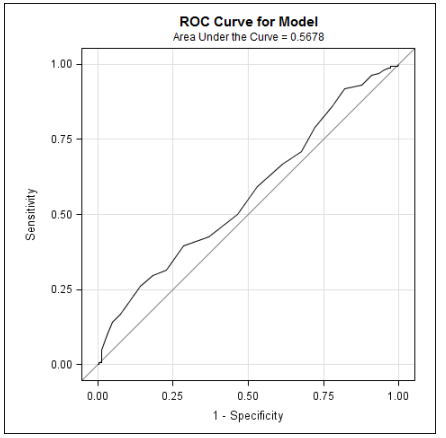
| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| AIC | 572.649 | 566.313 |
| SC | 576.715 | 574.445 |
| -2 Log L | 570.649 | 562.313 |

**Testing Global Null Hypothesis: BETA=0**

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| Likelihood Ratio | 8.3365 | 1 | 0.0039 |
| Score | 8.2831 | 1 | 0.0040 |
| Wald | 8.1129 | 1 | 0.0044 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Intercept | 1 | -2.4682 | 0.6990 | 12.4671 | 0.0004 |
| Age | 1 | 0.0509 | 0.0179 | 8.1129 | 0.0044 |



3. Vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnné **Gender**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘1’. Navíc vykreslete ROC křivku a přidejte výpis konfidenčního intervalu pro poměr šancí (PROC LOGISTIC).

- Jaké jsou koeficienty modelu?

**Class Level Information**

| Class | Value | Design Variables |
|---|---|---|
| Gender | Female | 1 |
| | Male | -1 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|
| Intercept | | 1 | -0,5380 | 0,1015 | 28,1144 | <,0001 |
| Gender | Female | 1 | 0,2186 | 0,1015 | 4,6436 | 0,0312 |

**Odds Ratio Estimates**

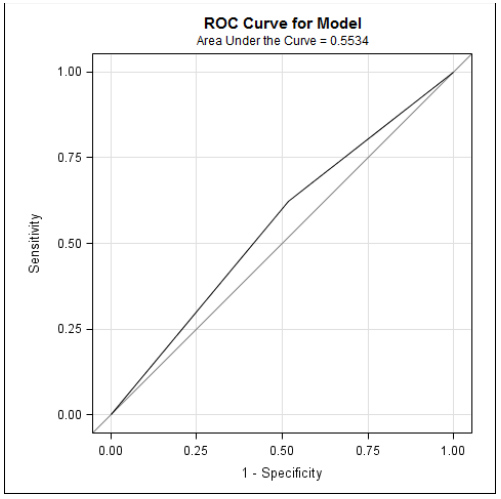
| Effect | Point Estimate | 95% Wald Confidence Limits | |
|---|---|---|---|
| Gender Female vs Male | 1,549 | 1,040 | 2,305 |

**Association of Predicted Probabilities and Observed Responses**

| | | | |
|---|---|---|---|
| Percent Concordant | 30,1 | Somers' D | 0,107 |
| Percent Discordant | 19,5 | Gamma | 0,215 |
| Percent Tied | 50,4 | Tau-a | 0,050 |
| Pairs | 43578 | c | 0,553 |

**Profile Likelihood Confidence Interval for Odds Ratios**

| Effect | Unit | Estimate | 95% Confidence Limits | |
|---|---|---|---|---|
| Gender Female vs Male | 1,0000 | 1,549 | 1,043 | 2,312 |



4. Vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnné **Gender**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘1’. Ponechte kódování typu effect, ale za referenční hodnotu nastavte ‘Female’. Navíc vykreslete ROC křivku a přidejte výpis konfidenčního intervalu pro poměr šancí (PROC LOGISTIC).
   - Co se změnilo oproti př. 3 (designová matice, koeficienty, Somers’D, ROC,…)?

5. Vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnných **Gender**, **Income** a **Age**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘1’. Změnte kódování klasifikačních proměnných na typ reference a za referenční hodnoty nastavte ‘Male’ a ‘Low’. Navíc vykreslete ROC křivku a ‘EffectPlot’. Model vytvořte pomocí backward metody. Vypište korelační matici (PROC LOGISTIC).

Type 3 Analysis of Effects

| Effect | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Gender | 1 | 6,0563 | 0,0139 |
| Age | 1 | 9,5102 | 0,0020 |
| Income | 2 | 13,0023 | 0,0015 |

Analysis of Maximum Likelihood Estimates

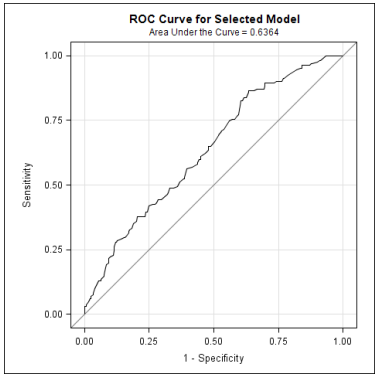
| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|
| Intercept | | 1 | -3,3071 | 0,7589 | 18,9930 | <,0001 |
| Gender | Female | 1 | 0,5204 | 0,2115 | 6,0563 | 0,0139 |
| Age | | 1 | 0,0560 | 0,0182 | 9,5102 | 0,0020 |
| Income | High | 1 | 0,8186 | 0,2556 | 10,2523 | 0,0014 |
| Income | Medium | 1 | 0,1064 | 0,2656 | 0,1605 | 0,6887 |

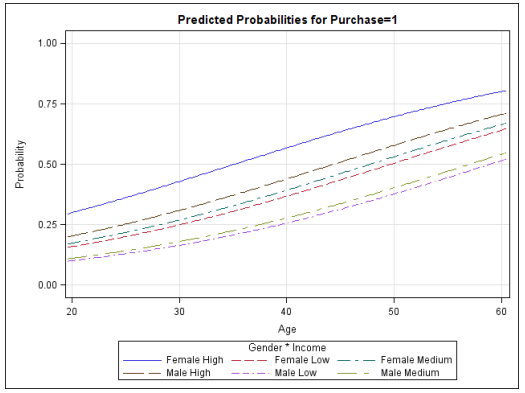
Odds Ratio Estimates

| Effect | Point Estimate | 95% Wald Confidence Limits | |
|---|---|---|---|
| Gender Female vs Male | 1,683 | 1,112 | 2,547 |
| Age | 1,058 | 1,021 | 1,096 |
| Income High vs Low | 2,267 | 1,374 | 3,742 |
| Income Medium vs Low | 1,112 | 0,661 | 1,872 |

Association of Predicted Probabilities and Observed Responses

| | | | |
|---|---|---|---|
| Percent Concordant | 63,2 | Somers' D | 0,273 |
| Percent Discordant | 35,9 | Gamma | 0,275 |
| Percent Tied | 0,8 | Tau-a | 0,128 |
| Pairs | 43578 | c | 0,636 |



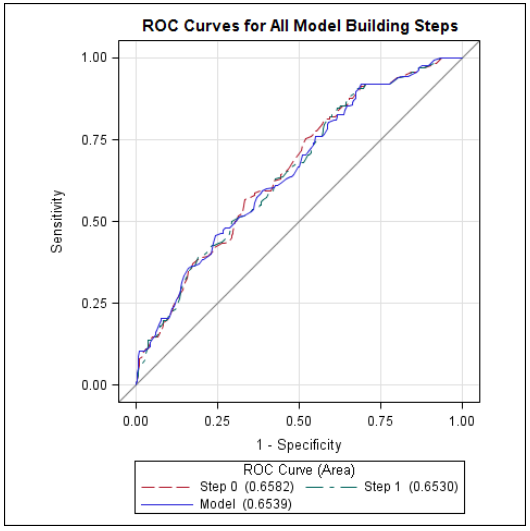

Estimated Correlation Matrix

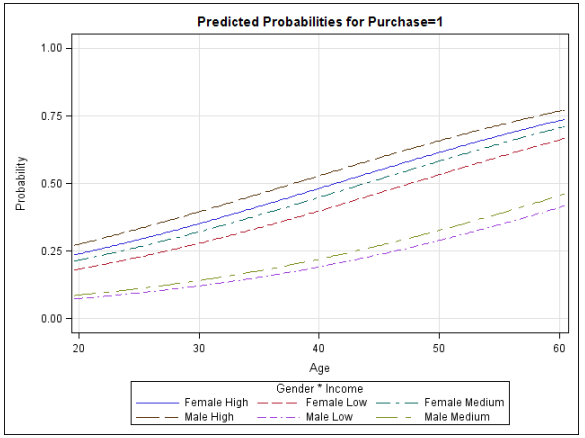
| Parameter | Intercept | Gender Female | Age | Income High | Income Medium |
|---|---|---|---|---|---|
| Intercept | 1,0000 | -0,2388 | -0,9474 | -0,3068 | -0,2209 |
| GenderFemale | -0,2388 | 1,0000 | 0,0420 | 0,1660 | 0,1365 |
| Age | -0,9474 | 0,0420 | 1,0000 | 0,0931 | 0,0153 |
| IncomeHigh | -0,3068 | 0,1660 | 0,0931 | 1,0000 | 0,5557 |
| IncomeMedium | -0,2209 | 0,1365 | 0,0153 | 0,5557 | 1,0000 |

6. Vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnných **Gender**, **Income** a **Age**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘1’. Změnte kódování klasifikačních proměnných na typ reference a za referenční hodnoty nastavte ‘Male’ a ‘Low’. Do modelu zahrňte také všechny interakce proměnných do druhého řádu. Navíc vykreslete ROC křivku a ‘EffectPlot’. Model vytvořte pomocí backward metody. Vypište detaily týkající se všech kroků výpočtu (opt. details) (PROC LOGISTIC).

| Analysis of Maximum Likelihood Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | | | 1 | -3.6026 | 0.8331 | 18.6985 | <.0001 |
| Gender | Female | | 1 | 1.0286 | 0.4528 | 5.1612 | 0.0231 |
| Age | | | 1 | 0.0540 | 0.0184 | 8.6169 | 0.0033 |
| Income | High | | 1 | 1.5547 | 0.4595 | 11.4449 | 0.0007 |
| Income | Medium | | 1 | 0.1756 | 0.4913 | 0.1278 | 0.7208 |
| Gender*Income | Female | High | 1 | -1.2133 | 0.5579 | 4.7298 | 0.0296 |
| Gender*Income | Female | Medium | 1 | 0.0295 | 0.5904 | 0.0025 | 0.9602 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 65.0 | Somers' D | 0.308 |
| Percent Discordant | 34.2 | Gamma | 0.310 |
| Percent Tied | 0.8 | Tau-a | 0.145 |
| Pairs | 43578 | c | 0.654 |



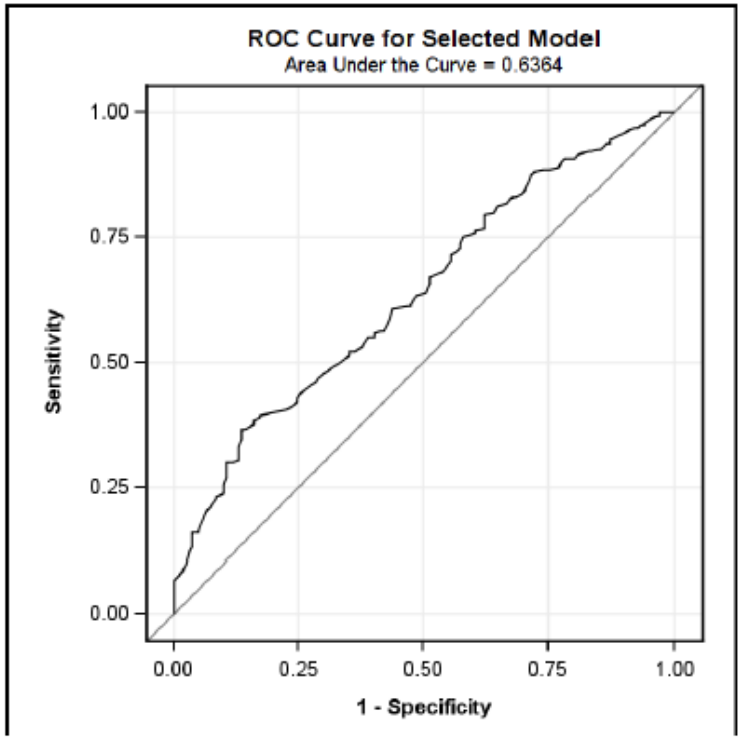
# Cvičení 11

1. Použijte tabulku **sales_inc** z minulého cvičení. Vytvořte logistický model vysvětlující proměnnou **Purchase** pomocí proměnných **Gender**, **Income** a **Age**, s tím že modelovaná bude pravděpodobnost pro hodnotu ‘0’. Změnte kódování klasifikačních proměnných na typ reference a za referenční hodnoty nastavte ‘Male’ a ‘Low’. Vykreslete ROC křivku a ‘EffectPlot’. Model vytvořte pomocí backward metody. Vypište korelační matici (PROC LOGISTIC). Dále zjistěte hodnotu KS statistiky a vykreslete empirické distribuční funkce získaného skóre pro hodnoty proměnné **Purchase** (PROC NPAR1WAY). Nakonec vypište tabulku s hodnotami absolutního a kumulativního Liftu pro decily skóre a vykreslete příslušný graf.

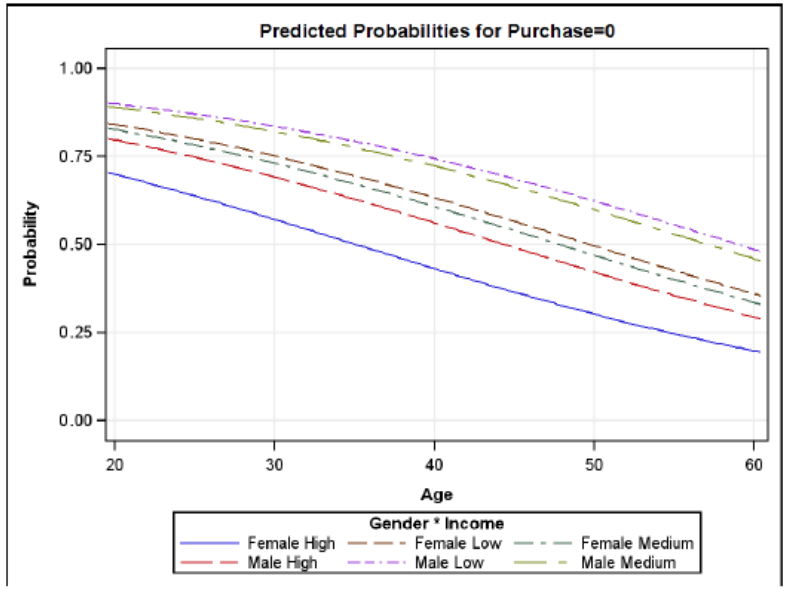
| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | | 1 | 3.3071 | 0.7589 | 18.9930 | <.0001 |
| Gender | Female | 1 | -0.5204 | 0.2115 | 6.0563 | 0.0139 |
| Age | | 1 | -0.0560 | 0.0182 | 9.5102 | 0.0020 |
| Income | High | 1 | -0.8186 | 0.2556 | 10.2523 | 0.0014 |
| Income | Medium | 1 | -0.1064 | 0.2656 | 0.1605 | 0.6887 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 63.2 | Somers' D | 0.273 |
| Percent Discordant | 35.9 | Gamma | 0.275 |
| Percent Tied | 0.8 | Tau-a | 0.128 |
| Pairs | 43578 | c | 0.636 |

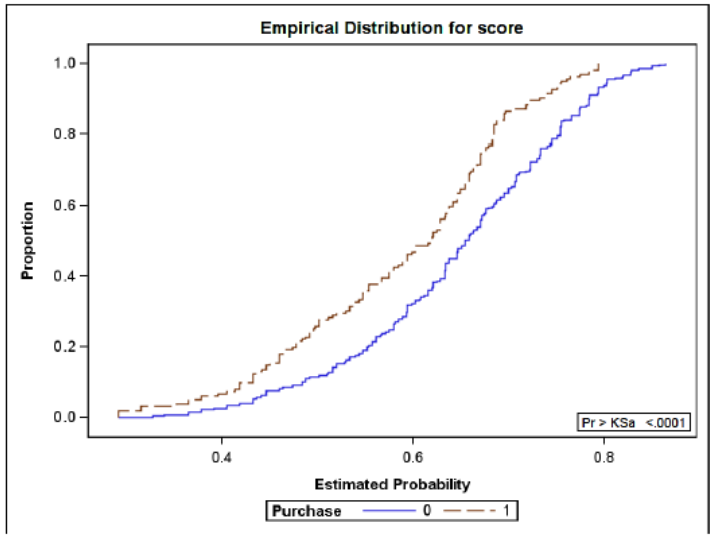
The LOGISTIC Procedure




| Estimated Correlation Matrix | | | | | |
|---|---|---|---|---|---|
| Parameter | Intercept | GenderFemale | Age | IncomeHigh | IncomeMedium |
| Intercept | 1.0000 | -0.2388 | -0.9474 | -0.3068 | -0.2209 |
| GenderFemale | -0.2388 | 1.0000 | 0.0420 | 0.1660 | 0.1365 |
| Age | -0.9474 | 0.0420 | 1.0000 | 0.0931 | 0.0153 |
| IncomeHigh | -0.3068 | 0.1660 | 0.0931 | 1.0000 | 0.5557 |
| IncomeMedium | -0.2209 | 0.1365 | 0.0153 | 0.5557 | 1.0000 |



| Kolmogorov-Smirnov Two-Sample Test (Asymptotic) | | | |
|---|---|---|---|
| KS | 0.110678 | D | 0.228510 |
| KSa | 2.297734 | Pr > KSa | <.0001 |




## Lift

| decile | N | N_of_bad | bad_rate | abs_lift | cum_lift |
|---|---|---|---|---|---|
| 1 | 43 | 23 | 53.5 | 1.423 | 1.423 |
| 2 | 43 | 24 | 55.8 | 1.485 | 1.454 |
| 3 | 43 | 17 | 39.5 | 1.052 | 1.320 |
| 4 | 43 | 15 | 34.9 | 0.928 | 1.222 |
| 5 | 43 | 18 | 41.9 | 1.114 | 1.200 |
| 6 | 44 | 18 | 40.9 | 1.088 | 1.181 |
| 7 | 43 | 21 | 48.8 | 1.299 | 1.198 |
| 8 | 43 | 10 | 23.3 | 0.619 | 1.126 |
| 9 | 43 | 10 | 23.3 | 0.619 | 1.070 |
| 10 | 43 | 6 | 14.0 | 0.371 | 1.000 |

**Absolutni a kumulativni lift**