Learning outcomes:	**SAS components learnt:**

Learning outcomes:

1. Identify SAS statements for reporting procedures
2. SAS procedures for creating reports
3. Global statements enhancing report
4. Output delivery system

SAS components learnt:

1. OPTIONS statement
2. ODS statement
3. TITLE statement
4. FOOTNOTE statement
5. PROC PRINT
6. PROC FREQ
7. PROC MEANS
8. PROC SUMMARY
9. PROC TABULATE

7.1 Creating reports using procedures

- Manipulation on data sets might be required so that the report contains all the required variables.
- Some global statements on the page size, line size, etc might be required to enhance the format of a report.
- Some local statements, like LABEL or FORMAT statement are useful for formatting a report. Note that the LABEL or FORMAT statement can be temporary or permanent in nature.
- A primary method is to use a procedure step, including
  - PROC PRINT
  - PROC FREQ
  - PROC MEANS
  - PROC SUMMARY
  - PROC TABULATE
- PROC PRINT
  - Syntax

```
PROC PRINT DATA = SAS-data-set <LABEL SPLIT=option>;
  <VAR variable(s) </option(s)>>;
  <WHERE = Where-expression>;
  <BY <DESCENDING> variable(s)>;
  <LABEL variable-1= 'Label'
                ;
        Variable-n= 'Label' >;
  <FORMAT variable(s) format>;
RUN;
```

- Use WHERE statement to display a report by selecting observations which fulfill the *where-expression* (a condition with operands including constants and variables and operators which are symbol for comparison, arithmetic calculation or logical operation).
- Use VAR statement to select variables for display (in the variable list order).
- Use BY statement to set the BY-variable for displaying a report by groups. But the data is required to be sorted according to the BY-variable using PROC SORT.
- LABEL option or SPLIT option for displaying variable labels listed in the LABEL statement.
- SAS applies the format to the corresponding variables listed in FORMAT statement.

- An example on PolyU overseas staff personal particulars
  - To display a report of staff ID, first name, last name, gender and origin country of staff from Australia (AU):

```
*p7_print.sas;
libname xlsdata 'd:\SAS_datasets\PolyU_overseas_staff.xlsx';
options ls=70;
proc print data=xlsdata.'PolyU_overseas_staff$'n;
*select variables;
var Staff_ID First_Name Last_Name Gender Origin_Country;
*select observations;
where origin_country='AU';
run;
```

Partial output:

```
                                                           Origin_
       Obs    Staff_ID    First_Name    Last_Name    Gender    Country

        1    A740103    Wilson       Dawes           M         AU
        2    A740121    Irenie       Elvish          F         AU
        3    A740166    Fadi         Nowd            M         AU
        4    A740172    Edwin        Comber          M         AU
        5    A741035    James        Blackley        M         AU
        ┋       ┋          ┋                                    ┋
       61    A061028    William      Smades          M         AU
       62    A061052    Richard      Fay             M         AU
       63    A061108    Libby        Levi            F         AU
       64    A060198    Meera        Body            F         AU
       65    A071088    Momolu       Kernitzki       M         AU
```

  - To display a report of staff by their country origins(BY-variable = origin_country):

```
*p7_print.sas-print observations by groups;
proc sort data=xlsdata.'PolyU_overseas_staff$'n
out=polyu_staff;
by origin_country;
run;

proc print data=polyu_staff;
var Staff_ID First_Name Last_Name origin_country;
by origin_country;
run;
```

The partial output:

```
                         The SAS System      11:55 Wednesday, August 15, 2012   27

--------------------------- Origin Country=AU ----------------------------------------

                                                          Origin_
        Obs    Staff_ID    First_Name    Last_Name        Country

         1    A741138    Hershell     Tolley              AU
         2    A971143    Louis        Favaron             AU
         3    A751044    Ray          Abbott              AU
         4    A740160    Chuck        Segrave             AU
         5    A741135    Tammy        Ruta                AU
                          ┋
--------------------------- Origin Country=AU ----------------------------------------
                          (continued)
                                                          Origin_
        Obs    Staff_ID    First_Name    Last_Name        Country

        49    A740166    Fadi         Nowd                AU
        50    A061098    Hal          Heatwole            AU
        51    A061108    Libby        Levi                AU
                          ┋
```

```
-------------------------- Origin Country=DE -----------------------------------

                                                             Origin_
          Obs     Staff_ID     First_Name     Last_Name      Country

           66     A051063      Regi           Kinol             DE
           67     A061032      Nasim          Smith             DE
           68     A061062      Debra          Armant            DE
           69     A061101      Burnetta       Buckner           DE
           70     A060140      Michael        Minas             DE
           71     A780148      Michael        Zubak             DE
                                 ┋
-------------------------- Origin Country=UK -----------------------------------
                                                             Origin_
          Obs     Staff_ID     First_Name     Last_Name      Country

           95     A860159      Lynelle        Phoumirath        UK
           96     A790124      Lucian         Daymond           UK
           97     A801090      Betty          Klibbe            UK
           98     A861104      Leoma          Johnson           UK
                                 ┋
-------------------------- Origin Country=US -----------------------------------
                                                             Origin_
          Obs     Staff_ID     First_Name     Last_Name      Country

          126     A740151      Julianna       Phaiyakounh       US
          127     A741053      Tywanna        Mcdade            US
          128     A761100      Tzue-Ing       Cormell           US
                                 ┋
```

- To obtain the staff details of female staff from Austrialia:

```
*p7_print.sas - female Australian staff info;
proc print data=polyu_staff label;
var Staff_ID First_Name Last_Name gender salary
origin_country;
where origin_country='AU' and gender='F';
label Staff_ID ='Staff no.'
      First_Name='First Name'
        Last_Name='Last Name'
        gender='Gender'
        origin_country='Country of Origin'
        salary='Monthly Salary';
Format salary DOLLAR10.;
run;
```

From the partial output, note that format is applied to salary and label option is used to label all variables listed:

```
                                                                 Country
            Staff                                       Monthly     of
      Obs    no.       First Name     Last Name   Gender  Salary    Origin

        5   A741135    Tammy          Ruta          F    $27,010     AU
        6   A991087    Virtina        O'Suilleabhain F   $28,325     AU
        9   A780132    Fancine        Kaiser        F    $28,525     AU


                         ┋

       59   A061047    Karen          Grzebien      F    $25,820     AU
       60   A740121    Irenie         Elvish        F    $26,600     AU
       63   A061065    Corneille      Malta         F    $28,040     AU
       64   A741041    Jaime          Wetherington  F    $26,120     AU
```

Ex. 7.1: Submit the following SAS program with multiple WHERE statements.

```
*p7ex1.sas;
libname SAS_data 'd:\SAS_datasets';

proc print data=SAS_data.electricity;
where scan(Time_period,2,' ')='1:';
where scan(Time_period,2,' ')='2:';
run;
```

(a) What is the output of the multiple WHERE statements?

(b) Please generalize the effect of multiple WHERE statements to DATA step and other PROCEDURE step.

(c) Rewrite the WHERE statement so that both Year 1 and Year 2 data are listed:

🔸 PROC FREQ
- Produce one-way to *n*-way table(s) and crosstabulation (contingency) tables
- Compute chi-square tests for one-way to *n*-way tables and measures of association and agreement for contingency tables
- Automatically display the output in a report and save the output in a SAS data set
- Syntax

```
PROC FREQ DATA = SAS-data-set <option(s)>;
  <TABLES variable(s) </option(s)>>;
  <OUTPUT OUT= SAS-data-set <options>>;
RUN;
```
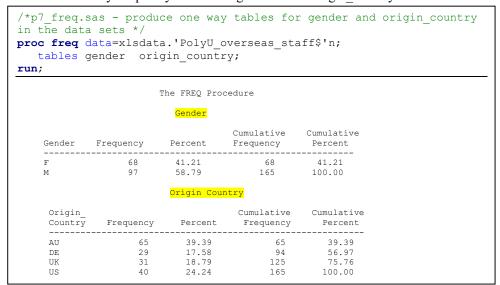
TABLES statement is used to make one way up to *n*-way tables. For one way table, put the variables in the list of the statement. Note that an asterisk '*' is required for 2-way or higher dimensional table. For two way table, *row-variable*column-variable* is used. So, for *n*-way table, *variable-1*variable-2*…*variable-n* is listed in the statement.
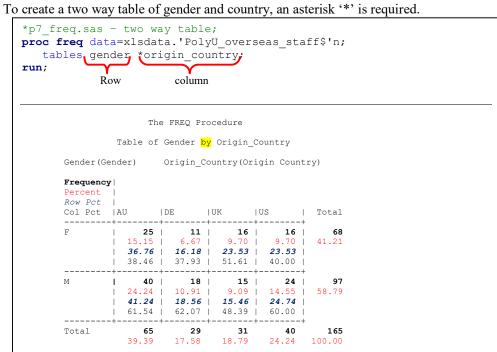
- An example on PolyU overseas staff personal particulars: The worksheet 'PolyU_overseas_staff' of the EXCEL file 'd:\SAS_datasets\ polyu_overseas_staff.xlsx' contains the overseas staff variables: Staff_ID, First_Name, Last_Name, Gender, Salary, Rank, Origin_Country, DOB, Hire_Date.
  - To create one way frequency tables for all variables in the data set:

```
*p7_freq.sas;
libname xlsdata 'd:\SAS_datasets\polyu_overseas_staff.xlsx';
/* produce one way tables for all variables in the data sets
*/
proc freq data=xlsdata.'PolyU_overseas_staff$'n;

Tables _ALL_;       Omitting this statement can create the
                    one way tables of ALL variables.

run;
```

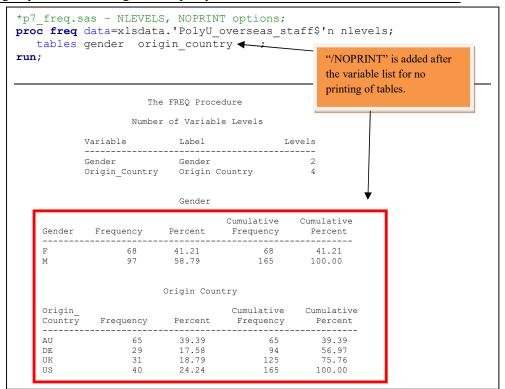 - To create one way frequency tables for "gender" and "origin_country":

```
/*p7_freq.sas - produce one way tables for gender and origin_country
in the data sets */
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
  tables gender  origin_country;
run;
```

```
                        The FREQ Procedure

                           Gender

                                   Cumulative    Cumulative
      Gender    Frequency    Percent   Frequency      Percent
      ----------------------------------------------------------
      F              68       41.21          68        41.21
      M              97       58.79         165       100.00

                        Origin Country

      Origin_                         Cumulative    Cumulative
      Country   Frequency    Percent   Frequency      Percent
      ----------------------------------------------------------
      AU             65       39.39          65        39.39
      DE             29       17.58          94        56.97
      UK             31       18.79         125        75.76
      US             40       24.24         165       100.00
```

- To create a two way table of gender and country, an asterisk '*' is required.

```
*p7_freq.sas – two way table;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender *origin_country;
run;
             Row            column
```

```
                        The FREQ Procedure

                  Table of Gender by Origin_Country

        Gender(Gender)     Origin_Country(Origin Country)

        Frequency|
        Percent  |
        Row Pct  |
        Col Pct  |AU      |DE      |UK      |US      |  Total
        ---------+--------+--------+--------+--------+
        F        |     25 |     11 |     16 |     16 |     68
                 |  15.15 |   6.67 |   9.70 |   9.70 |  41.21
                 |  36.76 |  16.18 |  23.53 |  23.53 |
                 |  38.46 |  37.93 |  51.61 |  40.00 |
        ---------+--------+--------+--------+--------+
        M        |     40 |     18 |     15 |     24 |     97
                 |  24.24 |  10.91 |   9.09 |  14.55 |  58.79
                 |  41.24 |  18.56 |  15.46 |  24.74 |
                 |  61.54 |  62.07 |  48.39 |  60.00 |
        ---------+--------+--------+--------+--------+
        Total          65       29       31       40      165
                    39.39    17.58    18.79    24.24   100.00
```

- Options in the PROC FREQ statement:

Option	Description
NLEVELS	Display a table which provides the number of levels for each variable named in the TABLES statement
PAGE	Display only one table per page
COMPRESS	Begin the display of the next one-way frequency table on the same page as the preceding one-way table if there is enough space to begin the table

PolyU overseas staff data revisit:
(a) NLEVELS option is used:

```
*p7_freq.sas – NLEVELS, NOPRINT options;
proc freq data=xlsdata.'PolyU_overseas_staff$'n nlevels;
    tables gender  origin_country        ;
run;
```
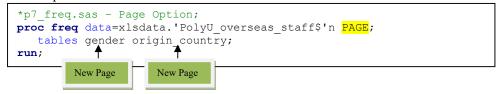
"/NOPRINT" is added after the variable list for no printing of tables.

```
                        The FREQ Procedure

                    Number of Variable Levels

        Variable            Label               Levels
        -------------------------------------------
        Gender              Gender                  2
        Origin_Country      Origin Country          4


                            Gender

                                    Cumulative    Cumulative
        Gender    Frequency    Percent    Frequency     Percent
        ------------------------------------------------------------
        F              68       41.21          68        41.21
        M              97       58.79         165       100.00


                        Origin Country

        Origin_                         Cumulative    Cumulative
        Country    Frequency    Percent    Frequency     Percent
        ------------------------------------------------------------
        AU             65       39.39          65        39.39
        DE             29       17.58          94        56.97
        UK             31       18.79         125        75.76
        US             40       24.24         165       100.00
```

(b) PAGE option is used:

```
*p7_freq.sas – Page Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n PAGE;
    tables gender origin_country;
run;
```

New Page   New Page

Table gender on page 1 and table origin_country on page 2, i.e. start a new page for each table.
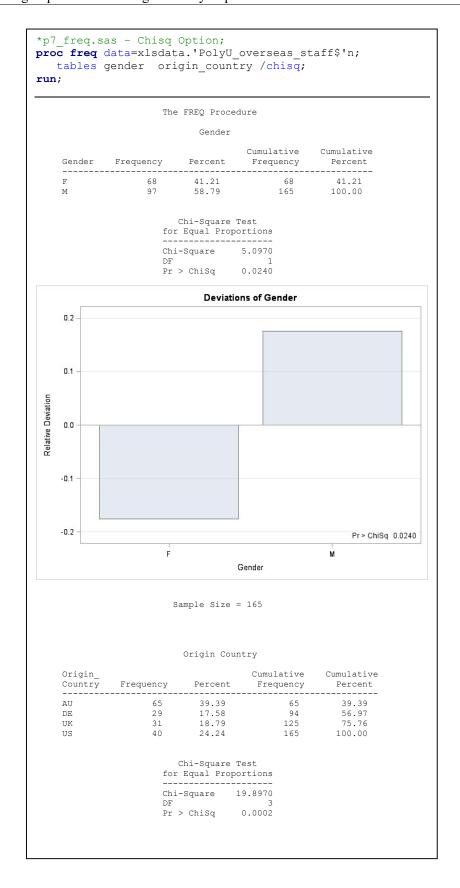
(c) COMPRESS option is used:
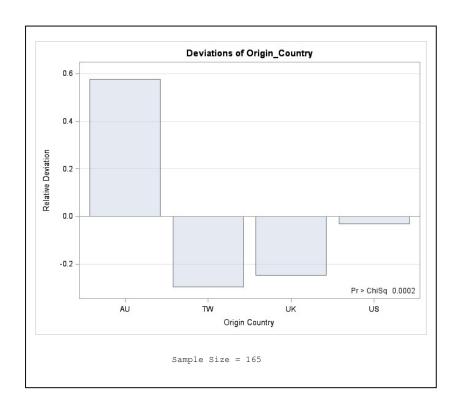
```
*p7_freq.sas – COMPRESS Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n COMPRESS;
    tables gender  origin_country;
run;
```

Table gender and table origin_country are on page 1 so that **all space of one page is utilized.** If one more table is added, SAS will check whether there is enough remaining space on the page for this additional table. If so, the additional table will be added to page 1. Otherwise, it will printed onto a new page.
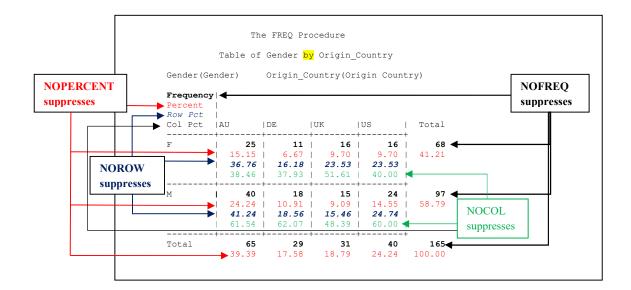
- Options in TABLES statement
  (a) When the TABLES statement is omitted, SAS create one-way tables for all variables of the data set. i.e. same results as "TABLES _ALL_;"
  (b) NOPRINT option suppresses the display of all output.
  (c) CHISQ option: SAS performs chi-square test with degree of freedom and p-value output.

```
*p7_freq.sas – Chisq Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender  origin_country /chisq;
run;
```

```
                        The FREQ Procedure

                             Gender

                                   Cumulative    Cumulative
     Gender    Frequency    Percent   Frequency     Percent
     -------------------------------------------------------------
     F            68         41.21        68         41.21
     M            97         58.79       165        100.00


                      Chi-Square Test
                    for Equal Proportions
                    --------------------
                    Chi-Square      5.0970
                    DF                   1
                    Pr > ChiSq      0.0240
```


Deviations of Gender

```
                    Sample Size = 165




                         Origin Country

     Origin_                          Cumulative    Cumulative
     Country   Frequency    Percent   Frequency     Percent
     -------------------------------------------------------------
     AU           65         39.39        65         39.39
     DE           29         17.58        94         56.97
     UK           31         18.79       125         75.76
     US           40         24.24       165        100.00


                      Chi-Square Test
                    for Equal Proportions
                    --------------------
                    Chi-Square     19.8970
                    DF                   3
                    Pr > ChiSq      0.0002
```

Deviations of Origin_Country

Sample Size = 165

(d) suppress statistics:

Option	Description
NOCUM	Suppress the display of cumulative frequency and cumulative percentage (only for ONE-way table)
NOPERCENT	Suppress the display of percentage, cumulative percentage and total percentage (only for ONE-way table)
NOFREQ	Suppress the display of the cell frequency and total frequency (NOT for one-way table)
NOROW	Suppress the display of the row percentage (NOT for ONE-WAY table)
NOCOL	Suppress the display of the column percentage (NOT for ONE-WAY table)

```
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender *origin_country    ;
run;
```

/NOCUM NOPERCENT NOFREQ

```
                          The FREQ Procedure

                   Table of Gender by Origin_Country

           Gender(Gender)      Origin_Country(Origin Country)

           Frequency|
           Percent  |
           Row Pct  |
           Col Pct  |AU      |DE      |UK      |US      |   Total
           ---------+--------+--------+--------+--------+
           F        |     25 |     11 |     16 |     16 |     68
                    |  15.15 |   6.67 |   9.70 |   9.70 |  41.21
                    |  36.76 |  16.18 |  23.53 |  23.53 |
                    |  38.46 |  37.93 |  51.61 |  40.00 |
           ---------+--------+--------+--------+--------+
           M        |     40 |     18 |     15 |     24 |     97
                    |  24.24 |  10.91 |   9.09 |  14.55 |  58.79
                    |  41.24 |  18.56 |  15.46 |  24.74 |
                    |  61.54 |  62.07 |  48.39 |  60.00 |
           ---------+--------+--------+--------+--------+
           Total          65       29       31       40      165
                       39.39    17.58    18.79    24.24   100.00
```

NOPERCENT suppresses

NOFREQ suppresses

NOROW suppresses

NOCOL suppresses

Examples:

```
*p7_freq.sas – NOCUM Option (1-way table);
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender  origin_country /nocum;
run;

*p7_freq.sas – NOROW, NOCOL Options (2-way table);
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender *origin_country /norow nocol;
run;
```

(e) control the format:

Option	Description
LIST	Display *n*-way tables in list format
CROSSLIST	Display *n*-way tables in column format (almost same as LIST but with sub-totals)
FORMAT=	Format the **frequencies** in *n*-way tables This option is applied to **crosstabulation** table displayed in default format. It does **not apply** to the crosstabulation table produced with the **LIST or CROSSLIST option**. It is the SAS numeric formats and the user's numeric format defined with PROC FORMAT. By default, the option is BEST6. or BEST7.

PolyU overseas staff data revisit:

```
*p7_freq.sas – LIST Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender *origin_country / LIST;
run;
```

```
                        The FREQ Procedure

                                       Cumulative  Cumulative
Gender  Origin_Country  Frequency    Percent    Frequency    Percent
---------------------------------------------------------------------
F       AU                25       15.15         25       15.15
F       DE                11        6.67         36       21.82
F       UK                16        9.70         52       31.52
F       US                16        9.70         68       41.21
M       AU                40       24.24        108       65.45
M       DE                18       10.91        126       76.36
M       UK                15        9.09        141       85.45
M       US                24       14.55        165      100.00
```

```
*p7_freq.sas – CROSSLIST Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender *origin_country / CROSSLIST;
run;
```

```
                        The FREQ Procedure

                 Table of Gender by Origin_Country

              Origin_                          Row      Column
   Gender     Country   Frequency    Percent    Percent    Percent
   ------------------------------------------------------------------
   F          AU            25       15.15      36.76      38.46
              DE            11        6.67      16.18      37.93
              UK            16        9.70      23.53      51.61
              US            16        9.70      23.53      40.00

              Total         68       41.21     100.00
   ------------------------------------------------------------------
   M          AU            40       24.24      41.24      61.54
              DE            18       10.91      18.56      62.07
              UK            15        9.09      15.46      48.39
              US            24       14.55      24.74      60.00

              Total         97       58.79     100.00
   ------------------------------------------------------------------
   Total      AU            65       39.39                100.00
              DE            29       17.58                100.00
              UK            31       18.79                100.00
              US            40       24.24                100.00

              Total        165      100.00
   ------------------------------------------------------------------
```
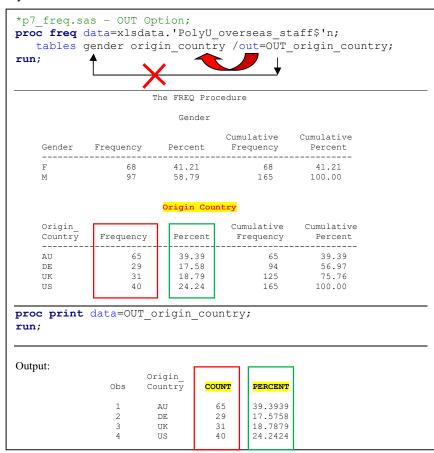
```
*p7_freq.sas – FORMAT= Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender *origin_country / Format=10.3;
run;
```

```
                        The FREQ Procedure

                 Table of Gender by Origin_Country

   Gender(Gender)      Origin_Country(Origin Country)

   Frequency|
   Percent  |
   Row Pct  |
   Col Pct  |AU        |DE        |UK        |US        |     Total
   ---------+----------+----------+----------+----------+
   F        |   25.000 |   11.000 |   16.000 |   16.000 |    68.000
            |    15.15 |     6.67 |     9.70 |     9.70 |     41.21
            |    36.76 |    16.18 |    23.53 |    23.53 |
            |    38.46 |    37.93 |    51.61 |    40.00 |
   ---------+----------+----------+----------+----------+
   M        |   40.000 |   18.000 |   15.000 |   24.000 |    97.000
            |    24.24 |    10.91 |     9.09 |    14.55 |     58.79
            |    41.24 |    18.56 |    15.46 |    24.74 |
            |    61.54 |    62.07 |    48.39 |    60.00 |
   ---------+----------+----------+----------+----------+
   Total        65.000     29.000     31.000     40.000    165.000
```
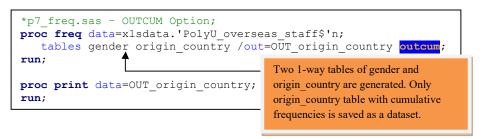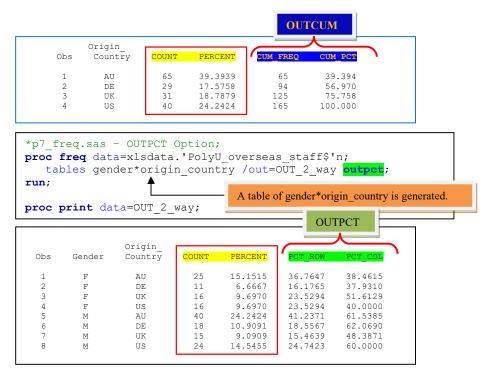
(f) OUT= option:

It is used to create a data set containing frequencies and percentages of the table of the **LAST variable** in the list of TABLE statement. Two automatic variables COUNT and PERCENT is created in the output SAS data set and other frequency and percentage variables are required to be requested with options in the TABLES statement.

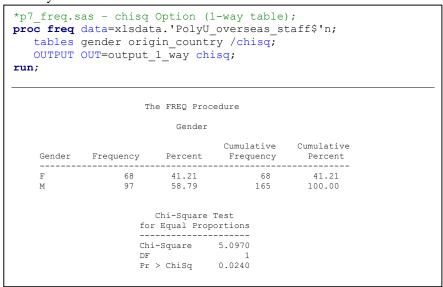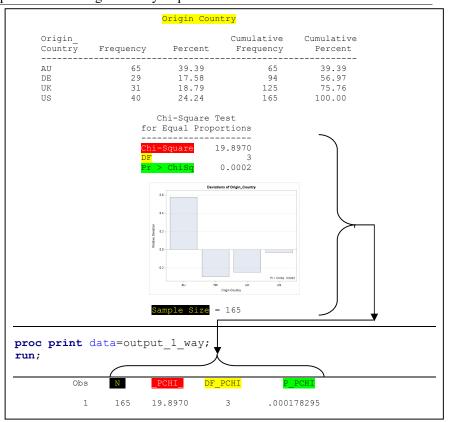| Option	Description
OUTCUM	Include the cumulative frequency and cumulative percentage in the output data set for ONE-WAY frequency tables
OUTPCT	Include the percentage of column frequency and row frequency in the output data set for *n*-way (*n*>1) frequency tables

PolyU overseas staff data revisit:

```
*p7_freq.sas – OUT Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender origin_country /out=OUT_origin_country;
run;
```

```
                       The FREQ Procedure

                          Gender

                                    Cumulative    Cumulative
   Gender    Frequency     Percent   Frequency      Percent
   -----------------------------------------------------------
   F              68        41.21          68        41.21
   M              97        58.79         165       100.00

                       Origin Country

   Origin_                          Cumulative    Cumulative
   Country   Frequency     Percent   Frequency      Percent
   -----------------------------------------------------------
   AU             65        39.39          65        39.39
   DE             29        17.58          94        56.97
   UK             31        18.79         125        75.76
   US             40        24.24         165       100.00
```

```
proc print data=OUT_origin_country;
run;
```

Output:
```
                    Origin_
          Obs      Country     COUNT     PERCENT

           1         AU          65      39.3939
           2         DE          29      17.5758
           3         UK          31      18.7879
           4         US          40      24.2424
```

```
*p7_freq.sas – OUTCUM Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender origin_country /out=OUT_origin_country outcum;
run;

proc print data=OUT_origin_country;
run;
```

Two 1-way tables of gender and origin_country are generated. Only origin_country table with cumulative frequencies is saved as a dataset.

OUTCUM

```
              Origin_
 Obs         Country      COUNT      PERCENT      CUM_FREQ      CUM_PCT

  1             AU          65       39.3939          65        39.394
  2             DE          29       17.5758          94        56.970
  3             UK          31       18.7879         125        75.758
  4             US          40       24.2424         165       100.000
```

```
*p7_freq.sas – OUTPCT Option;
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender*origin_country /out=OUT_2_way outpct;
run;

                              A table of gender*origin_country is generated.
proc print data=OUT_2_way;
```

OUTPCT

```
                  Origin_
 Obs    Gender   Country    COUNT    PERCENT    PCT_ROW    PCT_COL

  1       F        AU         25     15.1515    36.7647    38.4615
  2       F        DE         11      6.6667    16.1765    37.9310
  3       F        UK         16      9.6970    23.5294    51.6129
  4       F        US         16      9.6970    23.5294    40.0000
  5       M        AU         40     24.2424    41.2371    61.5385
  6       M        DE         18     10.9091    18.5567    62.0690
  7       M        UK         15      9.0909    15.4639    48.3871
  8       M        US         24     14.5455    24.7423    60.0000
```
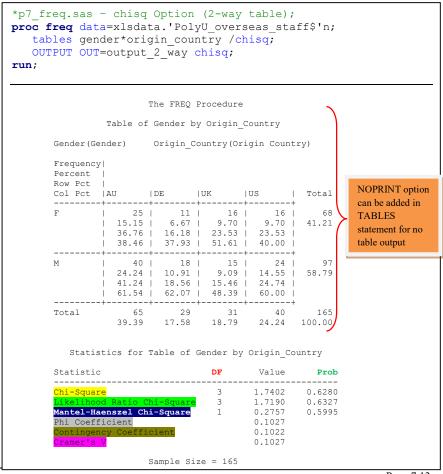
- Options in OUTPUT statement
  (a) OUT= option specifies the name of output SAS data set
  (b) Create a data set with specified statistics, e.g chi-square statistics, when CHISQ option is used in the TABLES statement. The statistics are specified in the option after the data set name
  (c) When more than one table are requested in the TABLE statement, ONLY the content of the LAST table will be output
  (d) PolyU overseas staff data revisit:
    (i) One-way table

```
*p7_freq.sas – chisq Option (1-way table);
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
    tables gender origin_country /chisq;
    OUTPUT OUT=output_1_way chisq;
run;
```

```
                          The FREQ Procedure

                                Gender

                                    Cumulative    Cumulative
      Gender    Frequency    Percent    Frequency      Percent
      -----------------------------------------------------------
      F             68        41.21          68        41.21
      M             97        58.79         165       100.00


                        Chi-Square Test
                       for Equal Proportions
                       ---------------------
                       Chi-Square      5.0970
                       DF                   1
                       Pr > ChiSq      0.0240
```

```
                            Origin Country

Origin_                                 Cumulative    Cumulative
Country     Frequency     Percent       Frequency     Percent
-----------------------------------------------------------------
AU              65          39.39            65          39.39
DE              29          17.58            94          56.97
UK              31          18.79           125          75.76
US              40          24.24           165         100.00


                        Chi-Square Test
                      for Equal Proportions
                      ---------------------
                      Chi-Square      19.8970
                      DF                    3
                      Pr > ChiSq       0.0002
```


Deviations of Origin_Country

```
                      Sample Size = 165
```

```sas
proc print data=output_1_way;
run;
```

```
        Obs       N      PCHI     DF_PCHI        P_PCHI

         1       165    19.8970      3        .000178295
```

(ii) Two-way table

```sas
*p7_freq.sas – chisq Option (2-way table);
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
   tables gender*origin_country /chisq;
   OUTPUT OUT=output_2_way chisq;
run;
```

```
                    The FREQ Procedure

            Table of Gender by Origin_Country

     Gender(Gender)      Origin_Country(Origin Country)

     Frequency|
     Percent  |
     Row Pct  |
     Col Pct  |AU      |DE      |UK      |US      |  Total
     ---------+--------+--------+--------+--------+
     F        |     25 |     11 |     16 |     16 |     68
              |  15.15 |   6.67 |   9.70 |   9.70 |  41.21
              |  36.76 |  16.18 |  23.53 |  23.53 |
              |  38.46 |  37.93 |  51.61 |  40.00 |
     ---------+--------+--------+--------+--------+
     M        |     40 |     18 |     15 |     24 |     97
              |  24.24 |  10.91 |   9.09 |  14.55 |  58.79
              |  41.24 |  18.56 |  15.46 |  24.74 |
              |  61.54 |  62.07 |  48.39 |  60.00 |
     ---------+--------+--------+--------+--------+
     Total          65       29       31       40      165
                 39.39    17.58    18.79    24.24   100.00
```

NOPRINT option can be added in TABLES statement for no table output

```
     Statistics for Table of Gender by Origin_Country

     Statistic                    DF       Value      Prob
     ------------------------------------------------------
     Chi-Square                    3       1.7402    0.6280
     Likelihood Ratio Chi-Square   3       1.7190    0.6327
     Mantel-Haenszel Chi-Square    1       0.2757    0.5995
     Phi Coefficient                       0.1027
     Contingency Coefficient               0.1022
     Cramer's V                            0.1027

                 Sample Size = 165
```

```
proc print data=output_2_way;
run;

Obs    N     PCHI    DF_PCHI    P_PCHI    LRCHI    DF_LRCHI    P_LRCHI

 1    165   1.74024     3       0.62803  1.71896      3       0.63273

Obs    MHCHI    DF_MHCHI    P_MHCHI    PHI       CONTGY     CRAMV

 1    0.27568      1        0.59954   0.10270   0.10216    0.10270
```

⚡ **Ex. 7.2:** Submit the following SAS program.

```
*p7ex2.sas;
libname xlsdata 'd:\SAS_datasets\polyu_overseas_staff.xlsx';
proc freq data=xlsdata.'PolyU_overseas_staff$'n;
tables gender*origin_country /nofreq;
run;
```

Print the output and compare the output with the previous example:

```
          Table of Gender by Origin_Country

   Gender(Gender)      Origin_Country(Origin Country)

   Percent   |
   Row Pct   |
   Col Pct   |AU      |DE      |UK      |US      |   Total
   ---------+--------+--------+--------+--------+
   F        |  15.15 |   6.67 |   9.70 |        |
            |  36.76 |  16.18 |  23.53 |        |
            |  38.46 |  37.93 |  51.61 |        |
   ---------+--------+--------+--------+--------+
   M        |  24.24 |        |   9.09 |  14.55 |
            |  41.24 |  18.56 |  15.46 |  24.74 |
            |  61.54 |  62.07 |  48.39 |  60.00 |
   ---------+--------+--------+--------+--------+
   Total
```

How to get a table without row percentage and column percentage?

🔸 PROC MEANS
- Calculate summary statistics and **multilevel** summaries
- Syntax

```
PROC MEANS DATA = SAS-data-set <statistic(s)><option(s)>;
  <VAR analysis-variable(s)>;
  <CLASS classification-variable(s)>;
  <OUTPUT OUT= SAS-data-set <options>>;
RUN;
```

- Report the number of non-missing observations, the mean, the standard deviation, the minimum value and the maximum value of all **numeric** variables of the input-SAS-data-set, by default.
- Analysis-variables are target variables for summary
- PolyU overseas staff data revisit:
  - Standard MEANS procedure

```
*p7_means.sas – standard means procedure;
libname xlsdata 'd:\SAS_datasets\Polyu_overseas_staff.xlsx';

proc means data=xlsdata.'Polyu_overseas_staff$'n;
run;
```
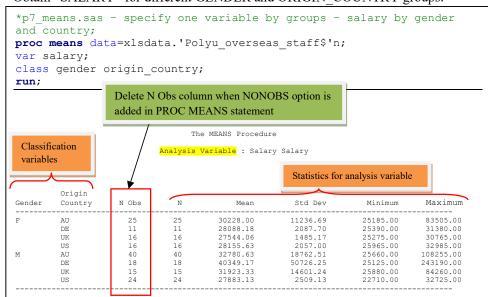
```
                              The MEANS Procedure

Variable     Label               N          Mean       Std Dev        Minimum        Maximum
--------------------------------------------------------------------------------------------
Salary       Salary            165      31160.12      20082.67       22710.00      243190.00
DOB*         DOB               165       3622.58       5456.29       -5842.00       10490.00
Hire_Date*   Hire_Date         165      12054.28       4619.94        5114.00       17167.00
--------------------------------------------------------------------------------------------
```

*DOB and HIRE_DATE are date variables. In SAS date are numeric. Therefore, statistics of DOB and HIRE_DATE are computed.

- Specify "SALARY" as analysis-variable in VAR statement
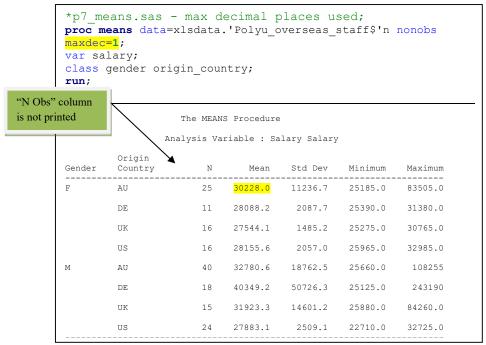
```
*p7_means.sas – specify variable(s) for analysis - salary;
proc means data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
run;
```

```
                          The MEANS Procedure

                   Analysis Variable : Salary Salary

        N          Mean       Std Dev        Minimum        Maximum
   ----------------------------------------------------------------
      165      31160.12      20082.67       22710.00      243190.00
   ----------------------------------------------------------------
```

- Obtain "SALARY" for different GENDER and ORIGIN_COUNTRY groups:

```
*p7_means.sas – specify one variable by groups – salary by gender
and country;
proc means data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
class gender origin_country;
run;
```

Delete N Obs column when NONOBS option is added in PROC MEANS statement

```
                          The MEANS Procedure

                   Analysis Variable : Salary Salary
```

Classification variables

Statistics for analysis variable

```
             Origin
Gender       Country    N Obs       N          Mean       Std Dev        Minimum        Maximum
-----------------------------------------------------------------------------------------------
F            AU            25       25      30228.00      11236.69       25185.00       83505.00
             DE            11       11      28088.18       2087.70       25390.00       31380.00
             UK            16       16      27544.06       1485.17       25275.00       30765.00
             US            16       16      28155.63       2057.00       25965.00       32985.00
M            AU            40       40      32780.63      18762.51       25660.00      108255.00
             DE            18       18      40349.17      50726.25       25125.00      243190.00
             UK            15       15      31923.33      14601.24       25880.00       84260.00
             US            24       24      27883.13       2509.13       22710.00       32725.00
-----------------------------------------------------------------------------------------------
```

- Options in PROC MEANS statement:

Option	Description
MAXDEC=	Specify the number of decimal places to use in printing the statistics
FW=	Specify the field width to use in displaying the statistics Default value is 12.
NONOBS	Suppress reporting the total number of observations (i.e. "N Obs" column) for each **unique** combination of the class variables

```
*p7_means.sas - max decimal places used;
proc means data=xlsdata.'Polyu_overseas_staff$'n nonobs
maxdec=1;
var salary;
class gender origin_country;
run;
```

"N Obs" column
is not printed

```
                          The MEANS Procedure

                      Analysis Variable : Salary Salary

            Origin
Gender      Country       N       Mean    Std Dev    Minimum    Maximum
------------------------------------------------------------------------
F           AU           25    30228.0    11236.7    25185.0    83505.0

            DE           11    28088.2     2087.7    25390.0    31380.0

            UK           16    27544.1     1485.2    25275.0    30765.0

            US           16    28155.6     2057.0    25965.0    32985.0

M           AU           40    32780.6    18762.5    25660.0     108255

            DE           18    40349.2    50726.3    25125.0     243190

            UK           15    31923.3    14601.2    25880.0    84260.0

            US           24    27883.1     2509.1    22710.0    32725.0
------------------------------------------------------------------------
```

```
*p7_means.sas – field width displayed;
proc means data=xlsdata.'Polyu_overseas_staff$'n nonobs fw=14 ;
var salary;
class gender origin_country;
run;
```

```
                               The MEANS Procedure

                         Analysis Variable : Salary Salary

            Origin
Gender      Country          N          Mean        Std Dev       Minimum       Maximum
----------------------------------------------------------------------------------------
F           AU              25       30228.0        11236.7       25185.0       83505.0

            DE              11       28088.2    2087.70241261     25390.0       31380.0
                                                Field width=14
            UK              16       27544.1    1485.17475824     25275.0       30765.0

            US              16       28155.6    2056.99771107     25965.0       32985.0

M           AU              40       32780.6        18762.5       25660.0      108255.0

            DE              18       40349.2        50726.3       25125.0      243190.0

            UK              15       31923.3        14601.2       25880.0       84260.0

            US              24       27883.1    2509.13453488     22710.0       32725.0
----------------------------------------------------------------------------------------
```

- Statistics keywords in PROC MEANS statement:
  - (a) Descriptive:

CLM	CSS	CV	LCLM	MAX
MEAN	MIN	MODE	N	NMISS
KURTOSIS	RANGE	SKEWESS	STDDEV	STDERR
SUM	SUMWGT	UCLM	USS	VAR

  - (b) Quantile:

MEDIAN (P50)	P1	P5	P10	QRANGE
Q1 (P25)	Q3 (P75)	P90	P95	P99

(c) Hypothesis: PROBT, T

```
*p7_means.sas – statistics options;
proc means data=xlsdata.'Polyu_overseas_staff$'n CV Median t ;
var salary;
class gender origin_country;
run;
```

```
                    The MEANS Procedure

              Analysis Variable : Salary Salary

          Origin                        Coeff of
Gender    Country          N Obs        Variation        Median    t Value
-----------------------------------------------------------------------------
F         AU                  25        37.1731182       28040.00     13.45

          DE                  11         7.4326720       27475.00     44.62

          UK                  16         5.3919960       27345.00     74.18

          US                  16         7.3058144       27582.50     54.75

M         AU                  40        57.2365856       27165.00     11.05

          DE                  18       125.7182206       27080.00      3.37

          UK                  15        45.7384418       28015.00      8.47

          US                  24         8.9987565       27200.00     54.44
-----------------------------------------------------------------------------
```

- The optional OUTPUT statement specify the output SAS data set name by OUT= option. The output data set contains the following variables:
  (a) BY variables
  (b) Class variables
  (c) The automatic variables _TYPE_ and _FREQ_
  (d) The variable requested in the OUTPUT statement
  (e) By default, _STAT_ refers to N, MIN, MAX, MEAN, STD
  (f) The statistics specified to be output can be renamed using this syntax:

  ```
  Statistic-1 = variable-name-1
          ..............
  Statistic-n = variable-name-n;
  ```

(g) Options:

Option	Description
NWAY	Specify that output data set contains only statistics for the observations with the **highest** _TYPE_ vlaue
DESCENDTYPES	Order the output data set by descending _TYPE_ value
CHARTYPE	Specify that the _TYPE_ variable in the output data set is a character representation of the binary value of _TYPE_. <table><tr><td>TYPE</td><td>Binary representation</td></tr><tr><td>0</td><td>00</td></tr><tr><td>1</td><td>01</td></tr><tr><td>2</td><td>10</td></tr><tr><td>3</td><td>11</td></tr></table>

PolyU overseas staff data revisit:

```
*p7_means.sas – output to file option;
proc means data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
class gender origin_country;
output out=out_mean;
run;
proc print data=out_mean;
run;
```

The output:

Obs	Gender	Origin_Country	_TYPE_*	_FREQ_	_STAT_	Salary
1			0	165	N	165.00
2			0	165	MIN	22710.00
3			0	165	MAX	243190.00
4			0	165	MEAN	31160.12
5			0	165	STD	20082.67
6		AU	1	65	N	65.00
7		AU	1	65	MIN	25185.00
8		AU	1	65	MAX	108255.00
9		AU	1	65	MEAN	31798.85
10		AU	1	65	STD	16230.66
11		DE	1	29	N	29.00
12		DE	1	29	MIN	25125.00
13		DE	1	29	MAX	243190.00
14		DE	1	29	MEAN	35698.45
15		DE	1	29	STD	40006.07
16		UK	1	31	N	31.00
17		UK	1	31	MIN	25275.00
18		UK	1	31	MAX	84260.00
19		UK	1	31	MEAN	29663.06
20		UK	1	31	STD	10273.44
21		US	1	40	N	40.00
22		US	1	40	MIN	22710.00
23		US	1	40	MAX	32985.00
24		US	1	40	MEAN	27992.13
25		US	1	40	STD	2314.86
26	F		2	68	N	68.00
27	F		2	68	MIN	25185.00
28	F		2	68	MAX	83505.00
29	F		2	68	MEAN	28762.72
30	F		2	68	STD	6974.15
31	M		2	97	N	97.00
32	M		2	97	MIN	22710.00
33	M		2	97	MAX	243190.00
34	M		2	97	MEAN	32840.77
35	M		2	97	STD	25458.26
36	F	AU	3	25	N	25.00
37	F	AU	3	25	MIN	25185.00
38	F	AU	3	25	MAX	83505.00
39	F	AU	3	25	MEAN	30228.00
40	F	AU	3	25	STD	11236.69
41	F	DE	3	11	N	11.00
42	F	DE	3	11	MIN	25390.00
43	F	DE	3	11	MAX	31380.00
44	F	DE	3	11	MEAN	28088.18
45	F	DE	3	11	STD	2087.70
46	F	UK	3	16	N	16.00
47	F	UK	3	16	MIN	25275.00
48	F	UK	3	16	MAX	30765.00
49	F	UK	3	16	MEAN	27544.06
50	F	UK	3	16	STD	1485.17
51	F	US	3	16	N	16.00
52	F	US	3	16	MIN	25965.00
53	F	US	3	16	MAX	32985.00
54	F	US	3	16	MEAN	28155.63
55	F	US	3	16	STD	2057.00

*_TYPE_=0 overall summary
TYPE=1 summary by ORIGIN_COUNTRY only
TYPE=2 summary by GENDER only
TYPE=3 summary by ORIGIN_COUNTRY and GENDER

Renaming statistics in the option of OUTPUT statement:

```
*p7_means.sas – output to file and renaming statistics;
proc means data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
class gender origin_country;
output out=out_mean1 min=minsalary sum=totalsalary;
run;

proc print data=out_mean1;
run;
```

```
                  Origin_
Obs     Gender    Country    _TYPE_     _FREQ_      minsalary    totalsalary

  1                             0         165        22710        5141420
  2                AU           1          65        25185        2066925
  3                DE           1          29        25125        1035255
  4                UK           1          31        25275         919555
  5                US           1          40        22710        1119685
  6        F                    2          68        25185        1955865
  7        M                    2          97        22710        3185555
  8        F       AU           3          25        25185         755700
  9        F       DE           3          11        25390         308970
 10        F       UK           3          16        25275         440705
 11        F       US           3          16        25965         450490
 12        M       AU           3          40        25660        1311225
 13        M       DE           3          18        25125         726285
 14        M       UK           3          15        25880         478850
 15        M       US           3          24        22710         669195
```

- WHERE, FORMAT, TITLE statement can be added in the MEANS procedure

Proc Format : Refer to Section 2.7!

```
*p7_means.sas - user defined format used in PROC MEANS
printout;
proc format;
  value $countrycode 'AU'='Australia'
                     'US'='United State'
                     'UK'='United Kingdom'
                     'DE'='Germany';
run;

options nocenter ls=130;
proc means data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
class gender origin_country;
where rank='Professor';
format origin_country $countrycode.;
title 'Professor gender and country of origin summary report';
run;
```

```
------------------------------------------------------------------------------------------------

Professor gender and country of origin summary report
20:12 Thursday, August 16, 2012  28

The MEANS Procedure
                        Analysis Variable : Salary Salary

Gender   Origin Country    N Obs        N       Mean     Std Dev     Minimum     Maximum
------------------------------------------------------------------------------------------------
F        Australia            1         1      83505.0        .       83505.0     83505.0

M        Australia            1         1      87975.0        .       87975.0     87975.0

         United Kingdom       1         1      84260.0        .       84260.0     84260.0
------------------------------------------------------------------------------------------------
```

- PROC SUMMARY
  - Syntax

```
PROC SUMMARY DATA = SAS-data-set <statistic(s)><option(s)>;
  VAR analysis-variable(s);
  CLASS classification-variable(s);
  OUTPUT OUT= SAS-data-set <options>;
RUN;
```

  - The SUMMARY procedure uses the same syntax as the MEANS procedure
  - There are only two difference to the two procedures:

	PROC MEANS	PROC SUMMARY
PRINT Option	By default, this option is set to display output.	By default, this option is set to display NO output.
VAR statement	Omitting this statement, SAS analyses all the numeric variables.	Omitting this statement, SAS produces a simple count of observations.

  - PolyU overseas staff data revisit:

```
*p7_summary.sas;
proc summary data=xlsdata.'Polyu_overseas_staff$'n;
var salary;
class gender origin_country;
output out=output_summary;
run;

proc print data=output_summary;
run;
```

  Same output as SAS data set: WORK.OUT_MEAN on page 7.17.

- PROC TABULATE
  - Create one-, two-, and three-dimensional tabular reports
  - Computes many of the same statistics which are computed by other descriptive statistical procedures such as PROC MEANS and PROC FREQ.
  - Syntax

```
PROC TABULATE DATA=SAS-data-set <OUT=SAS-data-set option(s)>;
   CLASS classification-variable(s);
   VAR analysis-variable(s);
   TABLE page-expression,
         row expression,
         column-expression </option(s)>;
RUN;
```

  - CLASS statement
    (a) Identify variable to be used as classification, or group, variables.
    (b) N, the number of non-missing values, is the default statistic for classification variables
    (c) Class variables can be numeric for discrete categories representation or character for identifying class or categories on which calculations are done.
    (d) Examples:
      (i) One-dimensional table:

```sas
*p7_tabulate.sas – one way table;
title ' ';
libname xlsdata 'd:\SAS_datasets\Polyu_overseas_staff.xlsx';
proc tabulate data=xlsdata.'Polyu_overseas_staff$'n;
class origin_country;
table origin_country;
run;
```

TABLE statement must be included in the program.

```
------------------------------------------------------
|                   Origin Country                   |
|----------------------------------------------------|
|    AU      |    DE      |    UK      |    US      |
|------------+------------+------------+------------|
|    N       |    N       |    N       |    N       |
|------------+------------+------------+------------|
|      65.00|      29.00|      31.00|      40.00|
------------------------------------------------------
```
*N - frequency

(ii) Two-dimensional table:

```sas
*p7_tabulate.sas – two-way table;
proc tabulate data=xlsdata.'Polyu_overseas_staff$'n;
class gender origin_country;
table gender, origin_country;
run;
```

```
------------------------------------------------------------------------------
|                              |                Origin Country              |
|                              |--------------------------------------------|
|                              |    AU     |    DE     |    UK     |    US     |
|                              |-----------+-----------+-----------+-----------|
|                              |    N      |    N      |    N      |    N      |
|------------------------------+-----------+-----------+-----------+-----------|
|Gender                        |           |           |           |           |
|------------------------------|           |           |           |           |
|F                             |      25.00|      11.00|      16.00|      16.00|
|------------------------------+-----------+-----------+-----------+-----------|
|M                             |      40.00|      18.00|      15.00|      24.00|
------------------------------------------------------------------------------
```

(iii) Three-dimensional table:

```sas
*p7_tabulate.sas – three-way table;
proc   tabulate data=xlsdata.'Polyu_overseas_staff$'n;
class rank gender origin_country;
table rank, gender, origin_country;
run;
```

Rank Instructor I
```
------------------------------------------------------------------------------
|                              |                Origin Country              |
|                              |--------------------------------------------|
|                              |    AU     |    DE     |    UK     |    US     |
|                              |-----------+-----------+-----------+-----------|
|                              |    N      |    N      |    N      |    N      |
|------------------------------+-----------+-----------+-----------+-----------|
|Gender                        |           |           |           |           |
|------------------------------|           |           |           |           |
|F                             |       6.00|       5.00|       6.00|       4.00|
|------------------------------+-----------+-----------+-----------+-----------|
|M                             |      19.00|       6.00|       5.00|      11.00|
------------------------------------------------------------------------------
```

Rank Assistant Professor
```
------------------------------------------------------------------------------
|                              |                Origin Country              |
|                              |--------------------------------------------|
|                              |    AU     |    DE     |    UK     |    US     |
|                              |-----------+-----------+-----------+-----------|
|                              |    N      |    N      |    N      |    N      |
|------------------------------+-----------+-----------+-----------+-----------|
|Gender                        |           |           |           |           |
|------------------------------|           |           |           |           |
|M                             |       2.00|       2.00|       2.00|       2.00|
------------------------------------------------------------------------------
```

There are many pages. Each new page for one rank.

- VAR statement
  - (a) Identify the numeric variable for which statistics are calculated.
  - (b) SUM is the default statistic for analysis variables
  - (c) Analysis variables are always numeric, tend to be continuous and are appropriate for calculating averages, sums or other statistics
- A CLASS statement or a VAR statement must be specified but both statements together are not required.
- TABLE statement
  - (a) Dimension expressions: page expression, row expression and column expression
  - (b) Element which can be used in a dimension expression: classification variables, analysis variables, the universal class variable ALL, keywords for statistics
  - (c) Operator that can be used in a dimension expression:
    - (i) Blank, which concatenates table information
    - (ii) Asterisk *, which crosses table information
    - (iii) Parentheses (), which group elements
    - (iv) Brackets < >, which name the denominator for row or column percentage
    - (v) Equal sign =, which changes the label for a variable or a statistic
    - (vi) Every variable which is a part of a dimension expression must be specified as a classification variable (CLASS statement) or an analysis variable (VAR statement)
    - (vii) Dimensional tables:

	Page dimension	Row dimension	Column dimension
One-dimensional			√
Two-dimensional		√	√
Three-dimensional	√	√	√

- PolyU overseas staff data revisit:

```
*p7_tabulate.sas – Var statement, default statistic: sum;
proc  tabulate data=xlsdata.'Polyu_overseas_staff$'n;
class gender origin_country;
var salary;
table gender all, origin_country*salary;
run;
```

```
-------------------------------------------------------------------------------------------
|                                   |                   Origin Country                    |
|                                   |-----------------------------------------------------|
|                                   |    AU     |    DE     |    UK     |    US     |
|                                   |-----------+-----------+-----------+-----------|
|                                   |  Salary   |  Salary   |  Salary   |  Salary   |
|                                   |-----------+-----------+-----------+-----------|
|                                   |    Sum    |    Sum    |    Sum    |    Sum    |
|-----------------------------------+-----------+-----------+-----------+-----------|
|Gender                             |           |           |           |           |
|-----------------------------------|           |           |           |           |
|F                                  |  755700.00|  308970.00|  440705.00|  450490.00|
|-----------------------------------+-----------+-----------+-----------+-----------|
|M                                  | 1311225.00|  726285.00|  478850.00|  669195.00|
|-----------------------------------+-----------+-----------+-----------+-----------|
|All                                | 2066925.00| 1035255.00|  919555.00| 1119685.00|
-------------------------------------------------------------------------------------------
```

(a) PROC TABULATE statistics
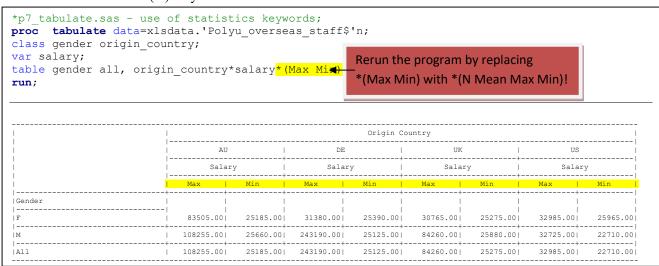
(i) Descriptive statistic keywords:

CSS	CV	LCLM	MEAN	
MAX	MIN	MODE	N	NMISS
KURTOSIS	RANGE	SKEWNESS	STDDEV	STDERR
SUM	SUMWGT	UCLM	USS	VAR
PCTN	REPPCTN	PAGEPCTN	ROWPCTN	COLPCT
PCTSUM	REPPCTSUM	PAGEPCTSUM	ROWPCTSUM	COLPCTSUM

(ii) Quantile statistic keywords:

MEDIAN (P50)	P1	P5	P10	QRANGE
Q1 (P25)	Q3 (P75)	P90	P95	P99

(iii) Hypothesis testing keywords: PROBT, T

(iv) PolyU overseas staff data revisit

```
*p7_tabulate.sas – use of statistics keywords;
proc  tabulate data=xlsdata.'Polyu_overseas_staff$'n;
class gender origin_country;
var salary;
table gender all, origin_country*salary*(Max Min);
run;
```

Rerun the program by replacing *(Max Min) with *(N Mean Max Min)!

```
-------------------------------------------------------------------------------------------------------------
|                               |                                   Origin Country                          |
|                               |-----------------------------------------------------------------------------|
|                               |        AU         |        DE         |        UK         |        US       |
|                               |-------------------+-------------------+-------------------+-----------------|
|                               |      Salary       |      Salary       |      Salary       |     Salary      |
|                               |-------------------+-------------------+-------------------+-----------------|
|                               |   Max   |   Min   |   Max   |   Min   |   Max   |   Min   |   Max  |   Min  |
|-------------------------------+---------+---------+---------+---------+---------+---------+--------+--------|
|Gender                         |         |         |         |         |         |         |        |        |
|-------------------------------|         |         |         |         |         |         |        |        |
|F                              | 83505.00| 25185.00| 31380.00| 25390.00| 30765.00| 25275.00|32985.00|25965.00|
|-------------------------------+---------+---------+---------+---------+---------+---------+--------+--------|
|M                              |108255.00| 25660.00|243190.00| 25125.00| 84260.00| 25880.00|32725.00|22710.00|
|-------------------------------+---------+---------+---------+---------+---------+---------+--------+--------|
|All                            |108255.00| 25185.00|243190.00| 25125.00| 84260.00| 25275.00|32985.00|22710.00|
-------------------------------------------------------------------------------------------------------------
```

(b) If there are only class variables in the TABLE statement, the default statistic is N, or number of non-missing values. For numeric variables, the default statistic is SUM. See the example on page 7.20.

(c) The OUT= option specify the name of output SAS data set which contains the following variables:

(i) BY variables

(ii) Class variables

(iii) The automatic variable _TYPE_, _PAGE_, and _TABLE_

(iv) Calculated statistics

(v) PolyU overseas staff data revisit:

```
*p7_tabulate.sas – output SAS data set for 3 tables;
proc  tabulate data=xlsdata.'Polyu_overseas_staff$'n out=work.tabulate;
class rank gender origin_country;
table gender;
table origin_country;
table rank, gender, origin_country;
run;
proc print data=work.tabulate;
run;
```

SAS will produce tables in the output window and at the same time, results are output to Work.Tabulate.

In Work.tabulate:

```
                                 Origin_
Obs    Rank              Gender   Country    _TYPE_    _PAGE_    _TABLE_    N

 1                          F                 010        1         1       68
 2                          M                 010        1         1       97
 3                                    AU      001        1         2       65
 4                                    DE      001        1         2       29
 5                                    UK      001        1         2       31
 6                                    US      001        1         2       40
 7     Assistant Professor   M       AU       111        1         3        2
 8     Assistant Professor   M       DE       111        1         3        2
 9     Assistant Professor   M       UK       111        1         3        2
10     Assistant Professor   M       US       111        1         3        2
11     Chair Professor       M       AU       111        2         3        2
12     Instructor I          F       AU       111        3         3        6
13     Instructor I          F       DE       111        3         3        5
14     Instructor I          F       UK       111        3         3        6
15     Instructor I          F       US       111        3         3        4
16     Instructor I          M       AU       111        3         3       19
17     Instructor I          M       DE       111        3         3        6
18     Instructor I          M       UK       111        3         3        5
19     Instructor I          M       US       111        3         3       11
20     Instructor II         F       AU       111        4         3        9
21     Instructor II         F       DE       111        4         3        2
22     Instructor II         F       UK       111        4         3        7
23     Instructor II         F       US       111        4         3        6
24     Instructor II         M       AU       111        4         3        7
25     Instructor II         M       DE       111        4         3        5
26     Instructor II         M       UK       111        4         3        4
27     Instructor II         M       US       111        4         3        5
28     Instructor III        F       AU       111        5         3        5
29     Instructor III        F       DE       111        5         3        1
30     Instructor III        F       UK       111        5         3        2
31     Instructor III        F       US       111        5         3        4
32     Instructor III        M       AU       111        5         3        8
33     Instructor III        M       DE       111        5         3        2
34     Instructor III        M       UK       111        5         3        3
35     Instructor III        M       US       111        5         3        2
36     Professor             F       AU       111        6         3        1
37     Professor             M       AU       111        6         3        1
38     Professor             M       UK       111        6         3        1
39     Teaching Fellow       F       AU       111        7         3        4
40     Teaching Fellow       F       DE       111        7         3        3
41     Teaching Fellow       F       UK       111        7         3        1
42     Teaching Fellow       F       US       111        7         3        2
43     Teaching Fellow       M       AU       111        7         3        1
44     Teaching Fellow       M       DE       111        7         3        2
45     Teaching Fellow       M       US       111        7         3        4
46     Vice President        M       DE       111        8         3        1
```

(d) WHERE, LABEL, FORMAT, TITLE statements can be added to enhance the report

7.2 Global statements enhancing report

- Global statements can be added **anywhere** in SAS program and they remain in effect until cancelled, changed or the SAS session ends.
- Global statements for enhancing report:
  - OPTIONS option
    - Syntax

    ```
    OPTIONS option(s);
    ```

    - Selected SAS system options:

Options	SAS action
DATE (default)*	Display the date and time that the SAS session began at the top of each page of SAS output
NODATE#	Display NO date and time that the SAS session began at the top of each page of SAS output
NUMBER(default)*	Print page numbers on the first line of each page of SAS output
NONUMBER#	Print NO page numbers on the first line of each page of SAS output
PAGENO = n#	Define a beginning page number (n) for the next page of SAS output.

CENTER*[#]	Center SAS output
NOCENTER	Left-align SAS output
PAGESIZE = n PS = n	Define the number of lines (n) that can be printed per page of SAS output
LINESIZE = width LS = width	Define the line size (width) for the SAS log and SAS output.
DTRESET	Update date and time at the top of each page of SAS output
NODTRSET*[#]	No update on date and time at the top of each page of SAS output
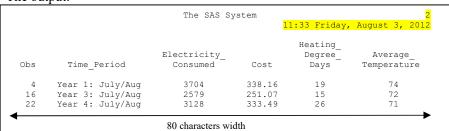
[#] used by default (HTML environment),

 *used in listing environment (See Section 7.4).

- An example, date, page number.

```
*p7_options.sas;
libname SASDATA 'd:\SAS_datasets';
options ls=80 date number;

proc print data=SASDATA.electricity;
where average_temperature >70;
run;
```

The output:

```
                          The SAS System                                2
                                              11:33 Friday, August 3, 2012


                                                 Heating_
                             Electricity_         Degree_       Average_
      Obs     Time_Period     Consumed    Cost     Days       Temperature

       4    Year 1: July/Aug    3704     338.16     19             74
      16    Year 3: July/Aug    2579     251.07     15             72
      22    Year 4: July/Aug    3128     333.49     26             71

                    80 characters width
```

- TITLE option
  - Syntax

```
TITLE<n> 'text';
```

  (a) Title appear at the top of the page
  (b) The default title is 'The SAS System'
  (c) Range of n is from 1 to 10
  (d) An unnumbered TITLE is equivalent to TITLE1.
  (e) Titles are printed according to the ascending order of the value n.
  (f) Replace a previous TITLE with the same number
  (g) Cancel ALL Titles with higher numbers (than n).
  - Title remain in effect until they are changed, cancelled or SAS session is ended.
  - To cancel ALL titles:

```
TITLE;
```

- An example

SAS Code	Resultant Title(s)	Syntax Remarks
```\n*p7_titles.sas;\nlibname SAS_data 'd:\SAS_datasets';\n```		
```\nproc print data=SAS_data.electricity;\nwhere scan(Time_period,2,' ')='1:';\ntitle 'The 1st title line';\ntitle2 'The 2nd title line';\nrun; /*print two title lines */\n```	The 1st title line The 2nd title line	(d), (e)
```\nproc print data=SAS_data.electricity;\nwhere scan(Time_period,2,' ')='1:';\ntitle2 'The NEXT title line';\nrun; /*only title2 option */\n```	The 1st title line The NEXT title line	(f)
```\nproc print data=SAS_data.electricity;\nwhere scan(Time_period,2,' ')='1:';\ntitle 'The TOP title line';\nrun; /*only title option */\n```	The TOP title line	(g)
```\nproc print data=SAS_data.electricity;\nwhere scan(Time_period,2,' ')='1:';\ntitle3 'The 3rd title line';\nrun; /*only title3 option */\n```	The TOP title line The 3rd title line	(e)
```\nproc print data=SAS_data.electricity;\nwhere scan(Time_period,2,' ')='1:';\ntitle;\nrun; /* delete all options */\n```		

- Title with date and time macro variables
  - (a) Double quotation marks must be used when a macro variable is used.
  - (b) The macro variable of date can be defined in the following way:

    ```
    %LET macro-variable =%SYSFUNC(today(),data-format);
    ```

  - (c) The macro variable of time can be defined in the following way:

    ```
    %LET macro-variable =%SYSFUNC(time(),data-format);
    ```

  - (d) An example

    ```
    *p7_titles.sas;
    /* testing macro */
    options ls=65;
    %let currentdate=%sysfunc(today(),worddate.);
    %let currenttime=%sysfunc(time(),timeampm.);
    proc print data=SAS_data.electricity;
    where scan(Time_period,2,' ')='1:';
    title1 ' Electricity Consumption Data for Year 1';
    title2 "created on &currentdate";
    title3 "at &currenttime";
    run;
    ```

```
            Electricity Consumption Data for Year 1          13
                    created on August 13, 2012
                        at 12:15:42 PM
                                  11:02 Monday, August 13, 2012


                                    Heating_
                       Electricity_          Degree_   Average_
   Obs    Time_Period    Consumed    Cost     Days   Temperature

    1 Year 1: Jan/Feb       3637    295.33    2226       29
    2 Year 1: March/Apr     2888    230.08    1616       37
    3 Year 1: May/June      2359    213.43     479       57
    4 Year 1: July/Aug      3704    338.16      19       74
    5 Year 1: Sept/Oct      3432    299.76     184       66
    6 Year 1: Nov/Dec       2446    214.44    1105       47
```

- FOOTNOTE option
  - Syntax (Very similar to syntax of Title Statement)

  ```
  FOOTNOTE<n> 'text';;
  ```

  (a) Footnotes appear at the bottom of the page
  (b) Range of *n* is from 1 to 10
  (c) An unnumbered FOOTNOTE is equivalent to FOOTNOTE1.
  (d) Footnotes are printed according to the ascending order of the value *n*.

  ```
                              Title1
                              Title 2
                              Title 3
                               ....



                               ....

                             Footnote1
                             Footnote2
                             Footnote3
  ```

  (e) Replace a previous FOOTNOTE with the same number
  (f) Cancel ALL Footnote(s) with higher numbers (than *n*).
  - Footnotes remain in effect until they are changed, cancelled or SAS session is ended.
  - To cancel ALL footnotes:

  ```
  FOOTNOTE;
  ```

7.3 Adding labels and formats
- Labels
  - Syntax

  ```
  LABEL variable = 'label'
        variable = 'label'
                 ⋮
        variable = 'label';
  ```

- Permanent: LABEL statement is added in DATA step
- Temporary: LABEL statement is used in procedures, like PROC PRINT with LABEL option is used, or SPLIT= option is specified:

  ```
  SPLIT='split-character';
  ```

- Details of LABEL statement, refer to Chapter 2:

⬥ Formats
- Syntax

```
FORMAT variable(s) format;
```

- SAS Format, the following are commonly used:

Variable type	SAS Formats	Stored (in bold) and displayed value
Character		**Wednesday**
$w.	$3.	Wed
Numeric		**12345678.91**
zw.d	z12.2	012345678.91
COMMA*w.d*	Comma13.2	12,345,678.91
DOLLAR*w.d*	Dollar14.2	$12,345,678.91
Date		**18991**
mmddyy*w.* *	mmddyy8.	12/30/11
Date*w.*	Date7.	30DEC12

- User defined Format, using PROC FORMAT to define the format.
- Details are given in Chapter 2.

Ex. 7.3: Fill in the displayed value for the given format.

Format	Stored value	Displayed value
$3.	Thursday	
6.1	4234.145	
COMMAX5.	4234.145	
DOLLAR9.2	4234.145	
DDMMYY8.	1	
DATE9.	1	
YEAR4.	1	

7.4 Reports output to external files

⬥ Output to SAS output window or SAS/GRAPH window
- Syntax

```
ODS LISTING;
```

⬥ ODS option
- ODS stands for Output Delivery System.
- ODS is a command for SAS to deliver the output to the following destinations:

Destination	Type of File	Viewed in
LISTING*		SAS Output Window or SAS/GRAPH Window
HTML#	Hypertext Markup Language (.html)	Web Browsers such as Internet Explorer
PDF	Portable Document Format (.pdf)	Adobe Products such as Acrobat Reader
RTF	Rich Text Format (.rtf)	Word Processors such as Microsoft Word

CSVALL	Comma Separated Value (.csv)	Editor or Microsoft Excel
EXCELXP/ TAGSETS.EXCELXP	Extensible Markup Language (.xml)	Microsoft Excel
MSOFFICE2K	Hypertext Markup Language (.html)	Web Brower or Microsoft Word or Microsoft Excel

*LISTING is used by default in SAS 9.2.
#HTML is used by default in SAS 9.3 and SAS 9.4.

- Syntax for external file destinations:

```
ODS destination FILE ='filname.ext' <options>;

<SAS code to generate a report(s)>

ODS destination CLOSE;
```

- RS= option is an alias for the RECORD_SEPARATOR= option. It is not needed in PDF destination but needed in CSVALL, MSOFFICE2K and EXCELXP destinations.
- STYLE= option is used to specify style definition which describes how to display the presentation aspects such as colors and fonts of SAS output. It cannot be used with the LISTING destination. Here are the SAS supplied style definitions:

Analysis	Astronomy	Banker	BarrettsBlue
Beige	blockPrint	Brick	Brown
Curve	D3d	Default	Education
EGDefault	Electronics	fancyPrinter	Festival
FestivalPrinter	Gears	Journal	Magnify
Meadow	MeadowPrinter	Minimal	Money
NoFontDefault	Normal	NormalPrinter	Printer
Rsvp	Rtf	sansPrinter	sasdocPrinter
Sasweb	Science	Seaside	SeasidePrinter
serifPrinter	Sketch	Statdoc	Statistical
Theme	Torn	Watercolor	
grayscalePrinter	Harvest	HighContrast	Journal2
Journal3	Listing	monochromePrinter	Ocean
Solutions			

- "ODS _ALL_ CLOSE;" is used to close all destinations including the LISTING and HTML destinations.
- Single destination

```
*p7_ods_1dest.sas;
ods html close;
Libname SASDATA 'd:\SAS_datasets';
ods html file='d:\SAS_datasets\p7_ods_1dest.html';
proc print data=SASDATA.electricity;
run;
ods html close;
ods html;
```

It is a good habit to open the HTML destination at the end of a program to guarantee an open destination for the next program submission.

- Results viewer shows the html file contents with the file name in the top menu.
- The file is called 'p7_ods_1dest.html' is created in the folder d:\SAS_datasets.

- Multiple destinations

```
*p7_ods_mdest.sas;
Libname SASDATA 'd:\SAS_datasets';

ods listing;
ods pdf file='d:\SAS_datasets\p7_ods_mdest.pdf';
ods rtf file='d:\SAS_datasets\p7_ods_mdest.rtf';

proc print data=SASDATA.electricity;
run;

ods pdf close;
ods rtf close;
```

- Results viewer shows the pdf file contents with the file name in the top menu.
- Microsoft Word is asked to open the rtf file.
- PDF and RTF destination is closed before running next program.

Ex. 7.4: Submit the following SAS program.

```
* p7ex4.sas;
Libname SASDATA 'd:\SAS_datasets';
ods html close;
ods html file='d:\SAS_datasets\p7_ods.html';
ods listing;
proc print data=SASDATA.electricity;
run;
ods html close;
ods listing close;
ods html;
```

Where will SAS direct the output to?

Ex. 7.5: Write a SAS program to print only electricity data of 'Year 1' and 'Year 2' and direct the output to d:\SAS_Datasets\Smith_Elect.html, d:\SAS_Datasets\Smith_Elect.csv and the output window. The output listed in the output window is as follows.

```
                          Electricity Consumption                              11
                             of Mr Smith's Family
                               (Year 1 to Year 2)

                      Electricity              No of days      Average
                      Consumption      Cost       with       Temperature
          Period       (in KWh)       (US$)     heating          (F)

      Year 1: Jan/Feb     3637        $295.3      2226           29
      Year 1: March/Apr   2888        $230.1      1616           37
      Year 1: May/June    2359        $213.4       479           57
      Year 1: July/Aug    3704        $338.2        19           74
      Year 1: Sept/Oct    3432        $299.8       184           66
      Year 1: Nov/Dec     2446        $214.4      1105           47
      Year 2: Jan/Feb     4463        $384.1      2351           28
      Year 2: March/Apr   2482        $295. 8     1508           41
      Year 2: May/June    2762        $255.9       657           54
      Year 2: July/Aug    2288        $219.7        35           68
      Year 2: Sept/Oct    2423        $256.6       308           62
      Year 2: Nov/Dec     2483        $276.1      1257           42




                          Survey conducted by MIT
```

25 lines per page

Answers:

```
*p7ex5.sas;
libname SASDATA 'D:\SAS_Datasets';
options ps= 25 nodate;
ods
ods html
ods html
ods         file='d:\SAS_Datasets\Smith_Elect.csv' ;


proc print data=SASDATA.       ;
label Time_Period ='Period'
        Electricity_Consumed =' Electricity Consumption (in KWh)'




title2 'of Mr Smith''s Family';
title3 '(Year 1 to Year 2)';
footnote  'Survey conducted by MIT';
run;

ods csvall
ods html
```