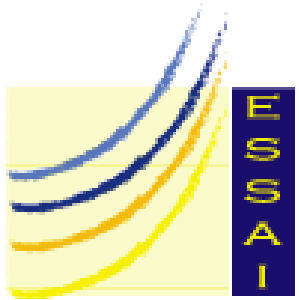
SAS Studio for beginners

Pr. Mokhtar KOUKI

Université de Carthage (ESSAI)

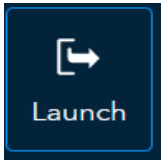
mokhtar.kouki@essai.ucar.tn

January 2023

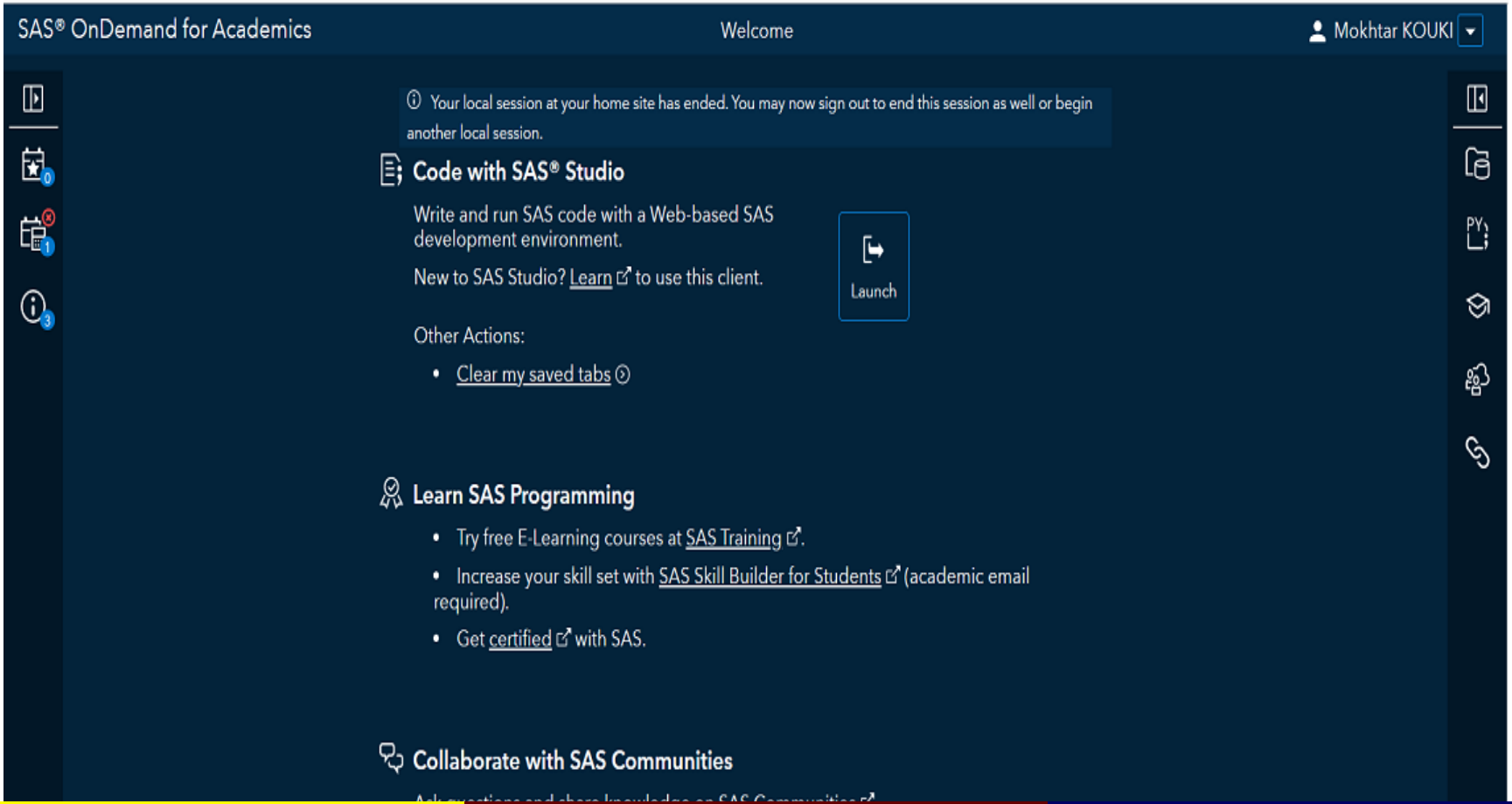


Contents

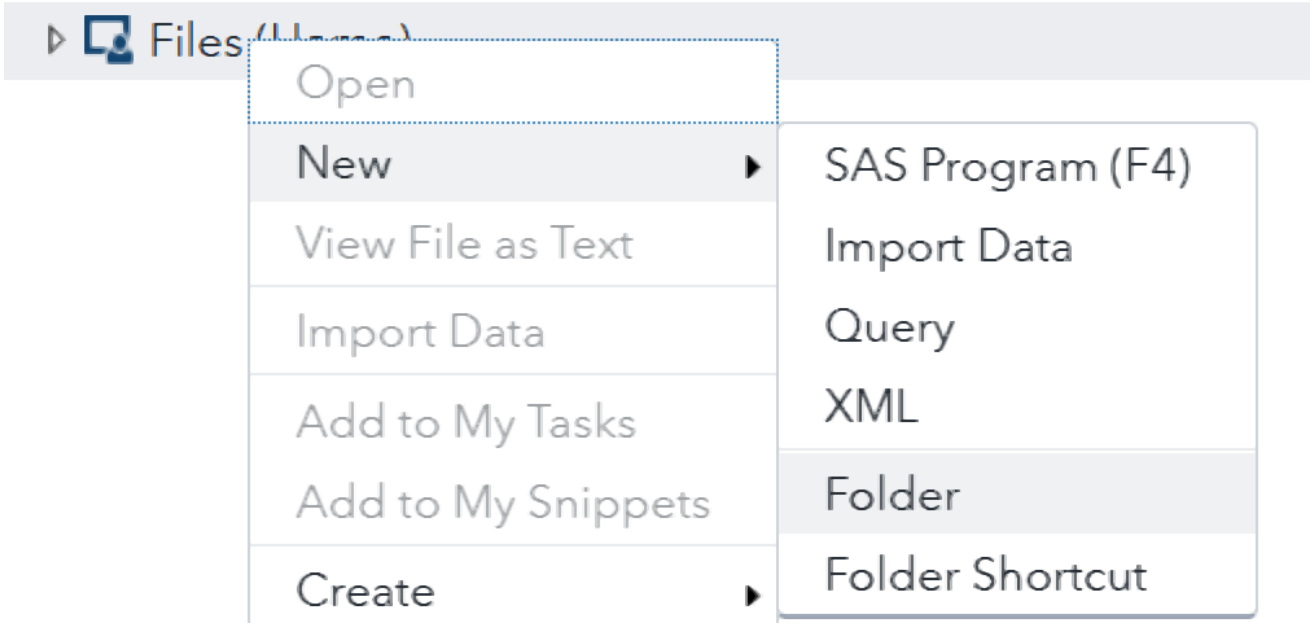
1. Get started : create a sas studio account
2. Create a library and import data
3. Assign labels and Formats
4. Proc Freq Procedure
5. Proc Means vs Proc Tabulate
6. Proc Boxplot
7. Proc sgplot and Proc sgscatter
8. Proc CORR
9. Proc Princomp
10. Proc SQL
    - Create a table from sql query
    - Joinning two data frame

- In your Browser write the following **SAS on Demand for academics** https://welcome.oda.sas.com/
- Log on or create an account
- When you log on, click on 

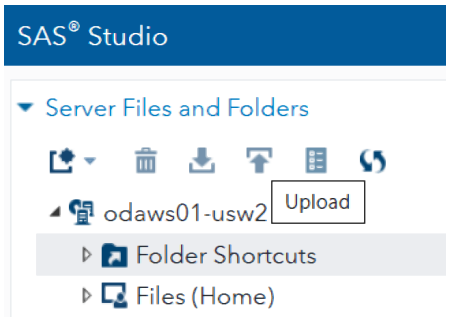
botton



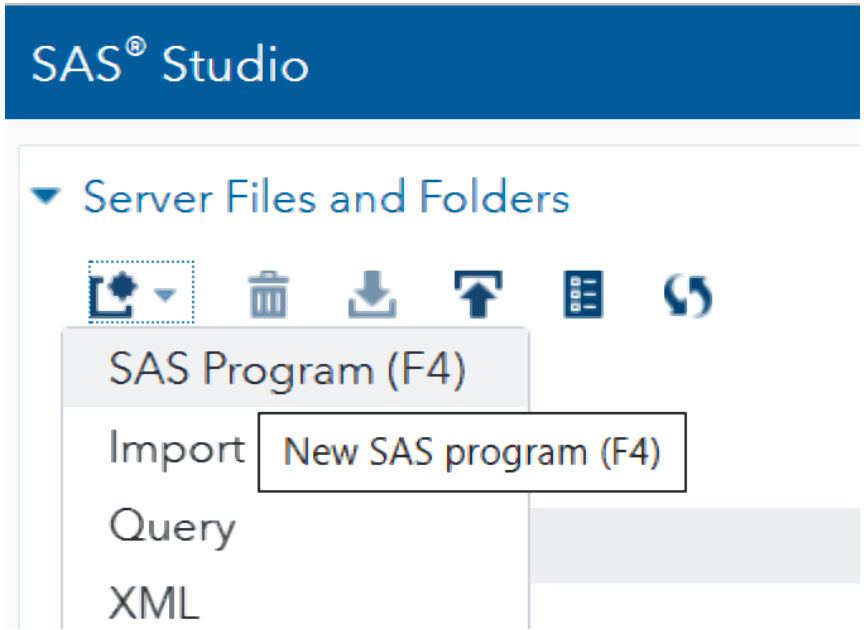
- **Create a new folder named tpsas** : right-Clic on **files** –>select –>**New** –> select **folder**



And **Upload** the Excel file CSSpending2005.xlsx by click on the Upload botton and brows your file as follows :



- **Create a new program file** : Clic on **Server Files and Folders**–>select –>**SAS Program F4**)



- **Create a library named tpsas to keep your files in** : type the following source code :

```
libname tpsas '/home/u59450248/tpsas/';
```

Type the following to import the excel file into sas studio and check for table contents:

```
1  libname tpsas '/home/u59450248/TPSAS2023/';
2
3  PROC IMPORT DATAFILE='/home/u59450248/CSSpending2005.xlsx'
4      DBMS=XLSX
5      OUT=TPSAS.CSS2005
6      replace;
7      GETNAMES=YES;
8      SHEET='DATA';
9  RUN;
10 * Check the contents of the data file;
11 PROC CONTENTS DATA=TPSAS.CSS2005;
12 RUN;
```

contents of the original table

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
2	V8	Num	8	BEST.	V8
3	V11	Num	8	BEST.	V11
4	V245	Num	8	BEST.	V245
5	V611	Num	8	BEST.	V611
6	V612	Num	8	BEST.	V612
7	V613	Num	8	BEST.	V613
8	V614	Num	8	BEST.	V614
9	V615	Num	8	BEST.	V615
10	V616	Num	8	BEST.	V616
11	V617	Num	8	BEST.	V617
12	V618	Num	8	BEST.	V618
14	V700	Num	8	BEST.	V700
13	V805	Num	8	BEST.	V805
1	V980	Num	8	BEST.	V980

- Create a temporary copy and assign labels of variables

```
data css2005;
    set tpsas.css2005;
    label V980='Household identifier'
        V8='Geographic zone'
        V11='Household size'
        V245='Socio-professional category of the main household maintainer'
        V611='Food expenditure'
        V612='Housing expenditure'
        V613='Clothing expenditure'
        V614='Hygiene and care expenditure'
        V615='Transportation expenditure'
        V616='Telecommunications expenditure'
        V617='Teaching expenditure'
        V618='Leisure expenditure'
        V805='Total expenditure'
        V700='household weight';
run;
```

contents of data table after assigning lables of variable

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
2	V8	Num	8	BEST.	Geographic zone
3	V11	Num	8	BEST.	Household size
4	V245	Num	8	BEST.	Socio-professional category of the main household maintainer
5	V611	Num	8	BEST.	Food expenditure
6	V612	Num	8	BEST.	Housing expenditure
7	V613	Num	8	BEST.	Clothing expenditure
8	V614	Num	8	BEST.	Hygiene and care expenditure
9	V615	Num	8	BEST.	Transportation expenditure
10	V616	Num	8	BEST.	Telecommunications expenditure
11	V617	Num	8	BEST.	Teaching expenditure
12	V618	Num	8	BEST.	Leisure expenditure
14	V700	Num	8	BEST.	household weight
13	V805	Num	8	BEST.	Total expenditure
1	V980	Num	8	BEST.	Household identifier

Prepare and assign format to categroical variable V8 and V245 (labels of levels)

```
proc format;
    value zone 1="Urban" 2="Rural";
    value csp
        1='Senior liberal profession'
        2='Medium liberal profession'
        3='Other employee'
        4='CEO, industry and commerce'
        5='Independent industry and trade'
        6='Non-agricultural workers'
        7='Farmer'
        8='Farm workers'
        9='Unemployed'
        10='Retreat'
        11='Other inactive'
        12='Non-household support';
run;
```

```
data css2005;
    set css2005;
    format V8 zone. V245 csp.;
run;
```

Use Proc Freq to obtain the frequency table of the Geographic Zone and the SPC variables and use V700 as weight

```
Proc freq data=css2005;
     table v8;
     weight V700;
run;
```

```
Proc freq data=css2005;
     table v245;
     weight V700;
run;
```

Geographic zone				
V8	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Urban	1517349	68.22	1517349	68.22
Rural	706800	31.78	2224149	100.00

Socio-professional category of the main household maintainer				
V245	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Senior liberal profession	134491	6.05	134491	6.05
Midum liberal profession	97224	4.37	231715	10.42
Other employee	191682	8.62	423397	19.04
CEO, industry and commerce	158875	7.14	582272	26.18
Independent industry and trade	76032	3.42	658304	29.60
Non-agricultural workers	581598	26.15	1239902	55.75
Farmer	206426	9.28	1446328	65.03
Farm workers	49922	2.24	1496250	67.27
Unemployed	42595	1.92	1538845	69.19
Retreat	308322	13.86	1847167	83.05
Other inactive	252983	11.37	2100150	94.42
Non-household support	123999	5.58	2224149	100.00

Create a Pivot-Table between SPC and Geographic Zone and test for the independence between the two variables

```
Proc freq data=css2005;
table v245*v8/nocol norow nopercent chisq;
weight v700;
run;
```

Table of V245 by V8			
	V8(Geographic zone)		
V245(Socio-professional category of the main household maintainer)	Urban	Rural	Total
Senior liberal profession	121740	12751	134491
Midum liberal profession	81082	16142	97224
Other employee	163164	28518	191682
CEO, industry and commerce	132914	25961	158875
Independent industry and trade	56383	19649	76032
Non-agricultural workers	378214	203384	581598
Farmer	41115	165311	206426
Farm workers	12975	36947	49922
Unemployed	22149	20446	42595
Retreat	267542	40780	308322
Other inactive	185186	67797	252983
Non-household support	54885	69114	123999
Total	1517349	706800	2224149

Statistics for Table of V245 by V8

Statistic	DF	Value	Prob
Chi-Square	11	440935	<.0001
Likelihood Ratio Chi-Square	11	439211	<.0001
Mantel-Haenszel Chi-Square	1	34035	<.0001
Phi Coefficient		0.44525	
Contingency Coefficient		0.40675	
Cramer's V		0.44525	

Use Proc Means to obtain descriptive statistics of the variables V611-V613 by Geographic Zone and use V700 as weight

```
proc means data=css2005  q1 mean median q3 std stderr sum maxdec=2;
class v8;
var v611-V613;
weight V700;
run;
```

Use Proc Means to obtain descriptive statistics of the variables V611-V613 by Geographic Zone and use V700 as weight

```
proc means data=css2005  q1 mean median q3 std stderr sum maxdec=2;
class v8;
var v611-V613;
weight V700;
run;
```

The MEANS Procedure

Geographic zone	N Obs	Variable	Label	Lower Quartile	Mean	Median	Upper Quartile	Std Dev	Std Error	Sum
Urban	7632	V611	Food expenditure	453.32	792.48	650.58	956.37	8688.42	7.05	1202461315.8
		V612	Housing expenditure	221.94	614.27	347.82	622.57	20364.04	16.53	932064745.91
		V613	Clothing expenditure	31.88	197.16	94.20	239.50	5140.68	4.17	299158185.38
Rural	4685	V611	Food expenditure	311.54	556.70	459.25	685.07	6534.00	7.77	393472172.58
		V612	Housing expenditure	89.66	241.29	141.05	248.28	4456.38	5.30	170543480.05
		V613	Clothing expenditure	16.00	100.28	48.50	119.00	2305.60	2.74	70878352.00

Use Proc Tabulate to obtain descriptive statistics of tatal expenditure by SPC and use V700 as weight

```
proc tabulate data=css2005;
class v245;
weight v700;
keylabel pctn='%' pctsum='% of sum';
var V805;
table V245 All='Total',(V805='')*(N pctn Mean median std sum pctsum)/box='';
run;
```

Use Proc Tabulate to obtain descriptive statistics of tatal expenditure by SPC and use V700 as weight

```
proc tabulate data=css2005;
class v245;
weight v700;
keylabel pctn='%' pctsum='% of sum';
var V805;
table V245 All='Total',(V805='')*(N pctn Mean median std sum pctsum)/box='';
run;
```

	Total expenditure						
	N	%	Mean	Median	Std	Sum	% of Total
Socio-professional category of the main household maintainer							
Senior liberal profession	630	5.11	4606.30	3566.00	67269.80	619505254.00	13.49
Medium liberal profession	525	4.26	3089.77	2385.00	41613.13	300399917.00	6.54
Other employee	1007	8.18	2084.80	1712.00	20043.23	399618916.00	8.70
CEO, industry and commerce	829	6.73	2287.01	1702.00	27069.80	363349082.00	7.91
Independent industry and trade	432	3.51	1614.33	1308.00	15536.49	122740700.00	2.67
Non-agricultural workers	3276	26.60	1381.51	1131.00	14261.88	803480651.00	17.50
Farmer	1288	10.46	1417.19	1106.00	16655.55	292545083.00	6.37
Farm workers	343	2.78	1086.07	856.00	9509.29	54218932.00	1.18
Unemployed	251	2.04	1076.06	794.00	12112.12	45834965.00	1.00
Retreat	1586	12.88	2698.46	1946.00	38917.15	831994187.00	18.12
Other inactive	1403	11.39	2072.76	1525.00	25282.35	524372652.00	11.42
Non-household support	747	6.06	1891.07	1297.00	25397.92	234491172.00	5.11
Total	12317	100.00	2064.86	1483.00	30299.31	4592551511.0	100.00

Use Proc Tabulate to obtain descriptive statistics of V611 and V612 by SPC and use V700 as weight

```
proc tabulate data=css2005;
class v245;
weight v700;
var V611-V612;
table V245 All='Total',(V611-V612)*(Q1 Mean median Q3)/box='';
run;
```

Use Proc Tabulate to obtain descriptive statistics of V611 and V612 by SPC and use V700 as weight

```
proc tabulate data=css2005;
class v245;
weight v700;
var V611-V612;
table V245 All='Total',(V611-V612)*(Q1 Mean median Q3)/box='';
run;
```

	Food expenditure				Housing expenditure			
	Q1	Mean	Median	Q3	Q1	Mean	Median	Q3
Socio-professional category of the main household maintainer								
Senior liberal profession	697.85	1088.66	950.11	1334.55	361.98	1371.75	707.06	1325.39
Medium liberal profession	562.53	869.92	759.84	996.25	272.18	835.14	513.83	872.04
Other employee	460.69	718.42	612.66	843.81	206.52	481.57	328.40	562.80
CEO, industry and commerce	455.56	757.28	647.53	897.63	205.04	491.79	306.11	545.68
Independent industry and trade	379.93	634.89	545.32	743.78	147.92	361.23	253.73	410.51
Non-agricultural workers	328.31	536.02	465.40	649.88	126.64	304.75	208.70	345.89
Farmer	337.99	627.86	498.03	767.45	92.10	261.22	152.80	269.10
Farm workers	283.37	481.68	407.70	597.63	81.65	211.53	138.44	241.70
Unemployed	230.27	434.09	347.45	511.92	92.08	236.39	169.70	249.03
Retreat	533.32	937.69	761.14	1106.88	225.22	625.25	363.21	656.07
Other inactive	428.49	792.04	629.89	1003.56	204.00	524.02	326.80	584.09
Non-household support	372.82	687.71	572.95	864.63	152.95	499.05	269.10	485.69
Total	399.26	717.55	589.09	875.15	158.94	495.74	275.00	512.12

Use Proc Tabulate to obtain the average total expenditure by Geographic Zone and SPC

```
title 'Descriptive Statistics of Total expenditure by Geographic Zone and SPC';
proc tabulate data=css2005;
    class v8 v245;
    weight v700;
    keylabel pctn='%';
    var V805;
    table (v245=''*V805='') V805='All', V8=''*(N pctn Mean median std) (N pctn
        Mean median std)/box='Socio-Professional Category';
run;
```

Use Proc Tabulate to obtain the average total expenditure by Geographic Zone and SPC

```
title 'Descriptive Statistics of Total expenditure by Geographic Zone and SPC';
proc tabulate data=css2005;
    class v8 v245;
    weight v700;
    keylabel pctn='%';
    var V805;
    table (v245=''*V805='') V805='All', V8=''*(N pctn Mean median std) (N pctn
        Mean median std)/box='Socio-Professional Category';
run;
```

Descriptive Statistics of Total expenditure by Geographic Zone and SPC

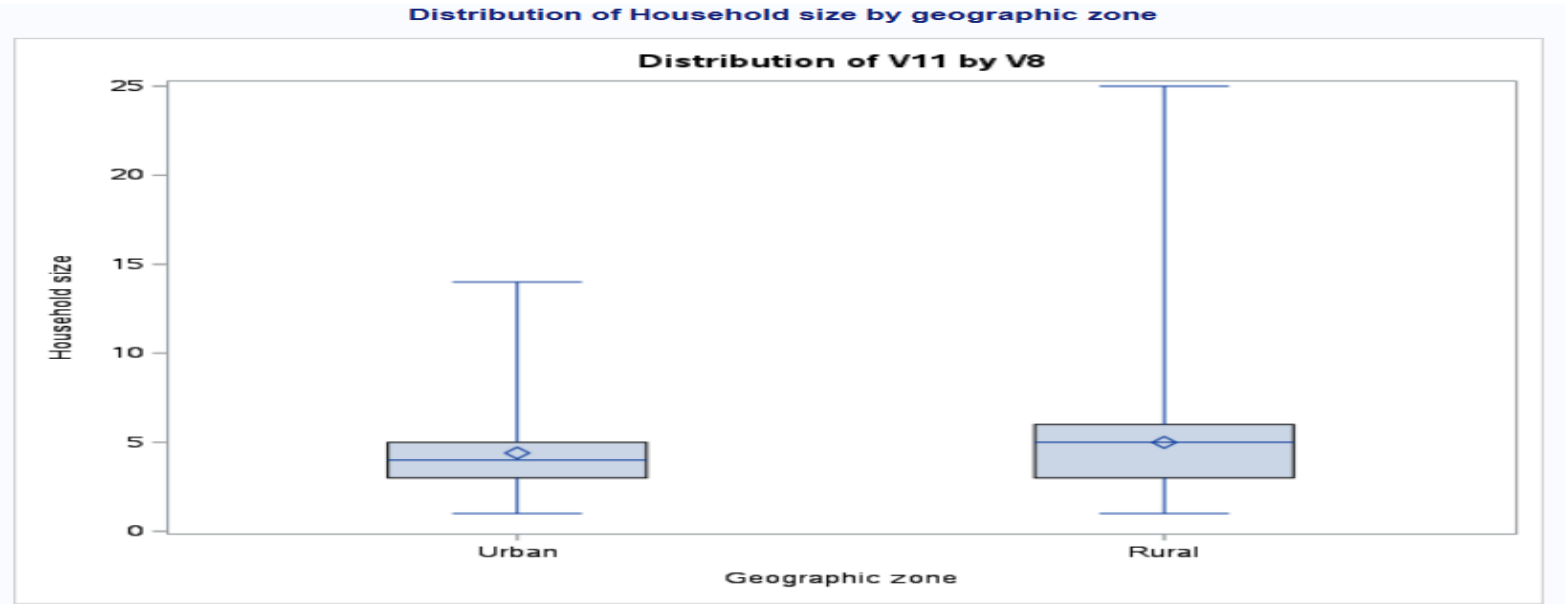
Socio-Professional Category	Urban					Rural					N	%	Mean	Median	Std
	N	%	Mean	Median	Std	N	%	Mean	Median	Std					
Senior liberal profession	546	4.43	4837.42	3712.00	70638.10	84	0.68	2399.61	1611.00	26517.50	630	5.11	4606.30	3566.00	67269.80
Medium liberal profession	425	3.45	3202.42	2442.00	44937.22	100	0.81	2523.94	2050.00	21315.99	525	4.26	3089.77	2385.00	41613.13
Other employee	818	6.64	2170.23	1777.00	21012.95	189	1.53	1596.03	1319.00	13721.50	1007	8.18	2084.80	1712.00	20043.23
CEO, industry and commerce	652	5.29	2442.64	1831.00	28664.05	177	1.44	1490.21	1249.00	17214.51	829	6.73	2287.01	1702.00	27069.80
Independent industry and trade	289	2.35	1823.94	1489.00	17367.08	143	1.16	1012.86	895.00	7308.59	432	3.51	1614.33	1308.00	15536.49
Non-agricultural workers	1957	15.89	1567.52	1283.00	16538.15	1319	10.71	1035.60	874.00	8433.02	3276	26.60	1381.51	1131.00	14261.88
Farmer	222	1.80	2068.55	1601.00	22680.85	1066	8.65	1255.19	1002.00	14423.31	1288	10.46	1417.19	1106.00	16655.55
Farm workers	87	0.71	1549.23	1304.00	11469.05	256	2.08	923.42	749.00	7884.65	343	2.78	1086.07	856.00	9509.29
Unemployed	111	0.90	1383.84	1040.00	15620.71	140	1.14	742.66	608.00	6269.91	251	2.04	1076.06	794.00	12112.12
Retreat	1283	10.42	2834.62	2031.00	42175.60	303	2.46	1805.14	1449.00	16543.29	1586	12.88	2698.46	1946.00	38917.15
Other inactive	934	7.58	2331.84	1731.00	28150.21	469	3.81	1365.09	1066.00	15361.13	1403	11.39	2072.76	1525.00	25282.35
Non-household support	308	2.50	2571.92	1763.00	33542.67	439	3.56	1350.40	1036.00	14346.94	747	6.06	1891.07	1297.00	25397.92
All	7632	61.96	2432.00	1755.00	35769.37	4685	38.04	1276.68	1000.00	13858.55	12317	100.00	2064.86	1483.00	30299.31

Use the Proc Boxplot to compare the distribution of the haousehold size between geographic zone

```
title "Distribution of Household size by geographic zone";
proc boxplot data=Css2005;
   plot v11*v8;
run;
```

Use the Proc Boxplot to compare the distribution of the haousehold size between geographic zone

```
title "Distribution of Household size by geographic zone";
proc boxplot data=Css2005;
    plot v11*v8;
run;
```

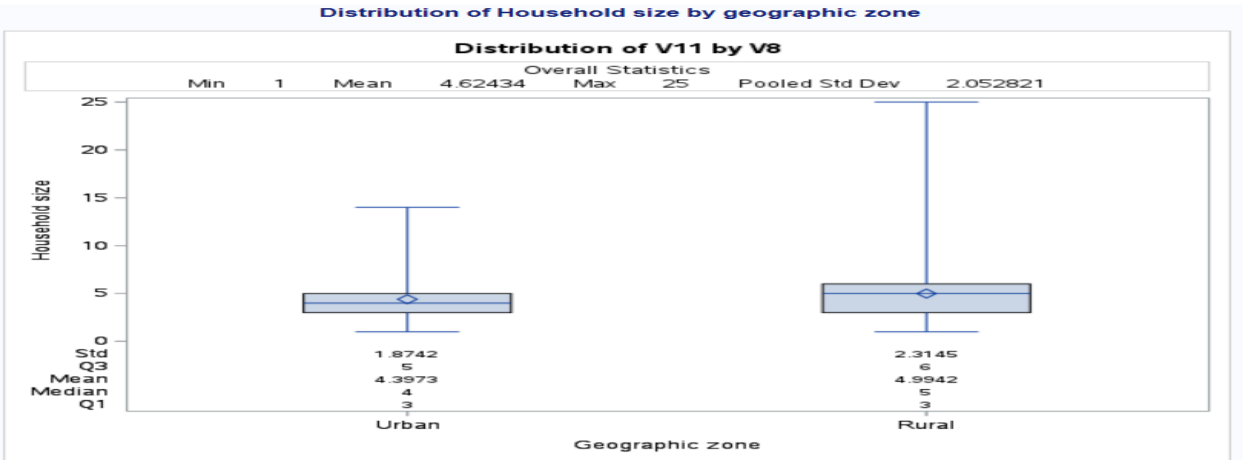


Use the Proc Boxplot to compare the distribution of the haousehold size between geographic zone and insert overall and individual statistics

```
proc boxplot data=Css2005;
    plot v11*v8;
    inset min mean max stddev /
        header = 'Overall Statistics'
        pos    = tm;
        insetgroup q1 q2 mean q3 stddev /
        header="Statistics By group statistics";
run;
```

Use the Proc Boxplot to compare the distribution of the haousehold size between geographic zone and insert overall and individual statistics

```
proc boxplot data=Css2005;
    plot v11*v8;
    inset min mean max stddev /
        header = 'Overall Statistics'
        pos    = tm;
        insetgroup q1 q2 mean q3 stddev /
        header="Statistics By group statistics";
run;
```



Use the Proc sgplot to plot a scatter plot of the leisure expenses and food expenses for the three first SPC and food expenses less 5000 and leisure expenses less than 2000 TND
Use the statement Where to filter data

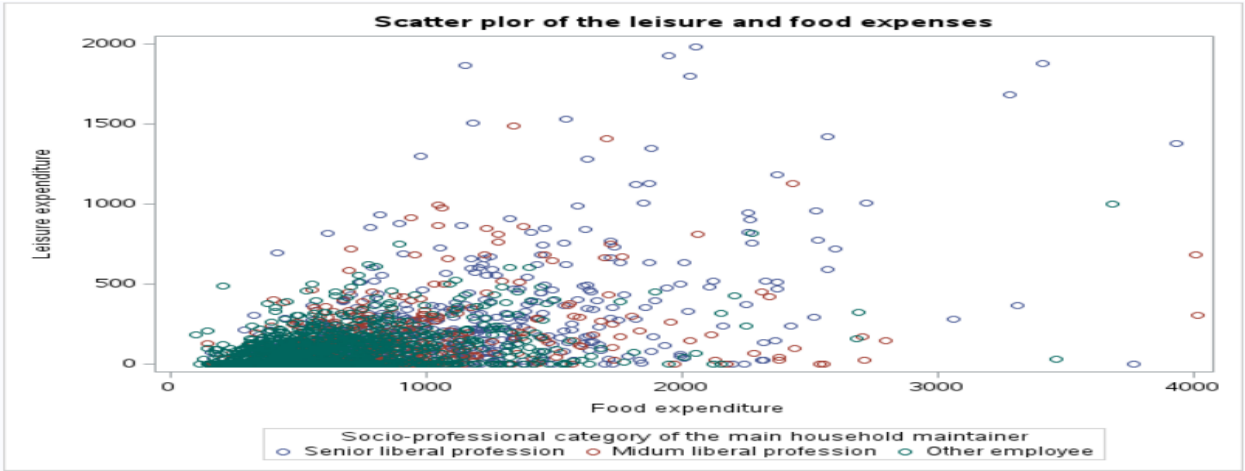
```
title "Scatter plor of the leisure and food expenses";

proc sgplot data=css2005;
    scatter x=V611 y=V618/group=V245;
    where v245 le 3 and v611 le 5000 and v618 le 2000;
run;
```

Use the Proc sgplot to plot a scatter plot of the leisure expenses and food expenses for the three first SPC and food expenses less 5000 and leisure expenses less than 2000 TND
Use the statement Where to filter data

```
title "Scatter plor of the leisure and food expenses";

proc sgplot data=css2005;
    scatter x=V611 y=V618/group=V245;
    where v245 le 3 and v611 le 5000 and v618 le 2000;
run;
```



Use **Proc CORR** to evaluate pairwaise correlation and scatter plot on SDGs database.

```
proc corr data=sdgs2021 plots=scatter;
    var sdg1_wpc sdg2_undernsh sdg2_stunting sdg2_wasting
        sdg2_obesity sdg3_incomeg sdg3_smoke sdg4_earlyedu
        sdg4_literacy sdg8_unemp;
run;
```

Use **Proc CORR** to evaluate pairwaise correlation and scatter plot on SDGs database.

```
proc corr data=sdgs2021 plots=scatter;
    var sdg1_wpc sdg2_undernsh sdg2_stunting sdg2_wasting
        sdg2_obesity sdg3_incomeg sdg3_smoke sdg4_earlyedu
        sdg4_literacy sdg8_unemp;
run;
```

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	sdg1_wpc	sdg2_undernsh	sdg2_stunting	sdg2_wasting	sdg2_obesity	sdg3_incomeg	sdg3_smoke	sdg4_earlyedu	sdg4_literacy	sdg8_unemp
sdg1_wpc Poverty headcount ratio at $1.90/day (%)	1.00000 161	0.78389 <.0001 147	0.68589 <.0001 158	0.36872 <.0001 158	-0.49833 <.0001 159	-0.24965 0.1420 36	-0.14981 0.3693 38	-0.52240 <.0001 147	-0.67239 <.0001 134	0.03031 0.7027 161
sdg2_undernsh Prevalence of undernourishment (%)	0.78389 <.0001 147	1.00000 170	0.66952 <.0001 165	0.40515 <.0001 165	-0.41487 <.0001 166	-0.20846 0.2224 36	-0.21715 0.1903 38	-0.56347 <.0001 153	-0.56620 <.0001 133	0.14974 0.0596 159
sdg2_stunting Prevalence of stunting in children under 5 years of age (%)	0.68589 <.0001 158	0.66952 <.0001 165	1.00000 186	0.63501 <.0001 186	-0.51591 <.0001 181	-0.31727 0.0594 36	-0.34190 0.0356 38	-0.57012 <.0001 167	-0.59508 <.0001 149	0.02168 0.7764 174
sdg2_wasting Prevalence of wasting in children under 5 years of age (%)	0.36872 <.0001 158	0.40515 <.0001 165	0.63501 <.0001 186	1.00000 188	-0.41440 <.0001 181	0.25709 0.1301 36	-0.00207 0.9902 38	-0.50495 <.0001 167	-0.47240 <.0001 149	0.13194 0.0827 174
sdg2_obesity Prevalence of obesity, BMI ≥ 30 (% of adult population)	-0.49833 <.0001 159	-0.41487 <.0001 166	-0.51591 <.0001 181	-0.41440 <.0001 181	1.00000 188	0.02974 0.8633 36	-0.07544 0.6526 38	0.30517 <.0001 170	0.45649 <.0001 151	0.17057 0.0236 176
sdg3_incomeg Gap in self-reported health status by income (percentage points)	-0.24965 0.1420 36	-0.20846 0.2224 36	-0.31727 0.0594 36	0.25709 0.1301 36	0.02974 0.8633 36	1.00000 36	0.10080 0.5586 36	0.10908 0.5328 35	0.32970 0.2497 14	-0.32059 0.0566 36
sdg3_smoke Daily smokers (% of population aged 15 and over)	-0.14981 0.3693 38	-0.21715 0.1903 38	-0.34190 0.0356 38	-0.00207 0.9902 38	-0.07544 0.6526 38	0.10080 0.5586 36	1.00000 38	-0.25400 0.1293 37	0.04650 0.8642 16	0.13876 0.4061 38
sdg4_earlyedu Participation rate in pre-primary organized learning (% of children aged 4 to 6)	-0.52240 <.0001 147	-0.56347 <.0001 153	-0.57012 <.0001 167	-0.50495 <.0001 167	0.30517 <.0001 170	0.10908 0.5328 35	-0.25400 0.1293 37	1.00000 174	0.52787 <.0001 138	-0.27248 0.0005 161
sdg4_literacy Literacy rate (% of population aged 15 to 24)	-0.67239 <.0001 134	-0.56620 <.0001 133	-0.59508 <.0001 149	-0.47240 <.0001 149	0.45649 <.0001 151	0.32970 0.2497 14	0.04650 0.8642 16	0.52787 <.0001 138	1.00000 153	0.09425 0.2529 149
sdg8_unemp Unemployment rate (% of total labor force, ages 15+)	0.03031 0.7027 161	0.14974 0.0596 159	0.02168 0.7764 174	0.13194 0.0827 174	0.17057 0.0236 176	-0.32059 0.0566 36	0.13876 0.4061 38	-0.27248 0.0005 161	0.09425 0.2529 149	1.00000 178

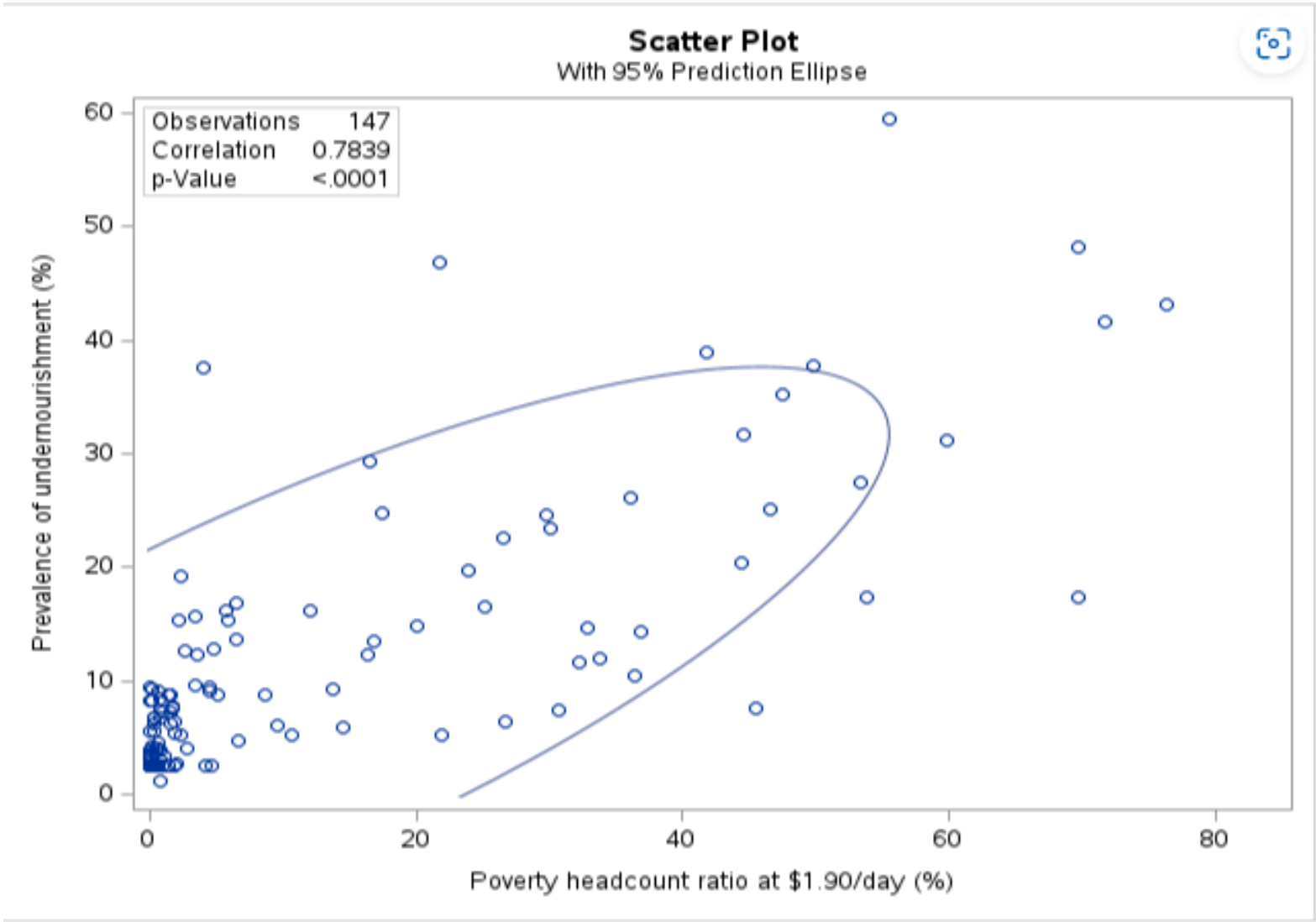


Figure: Scatter plot and correlation

Use **Proc CORR** to evaluate pairwaise correlation and scatter plot on SDGs database and histogram on diagonal.

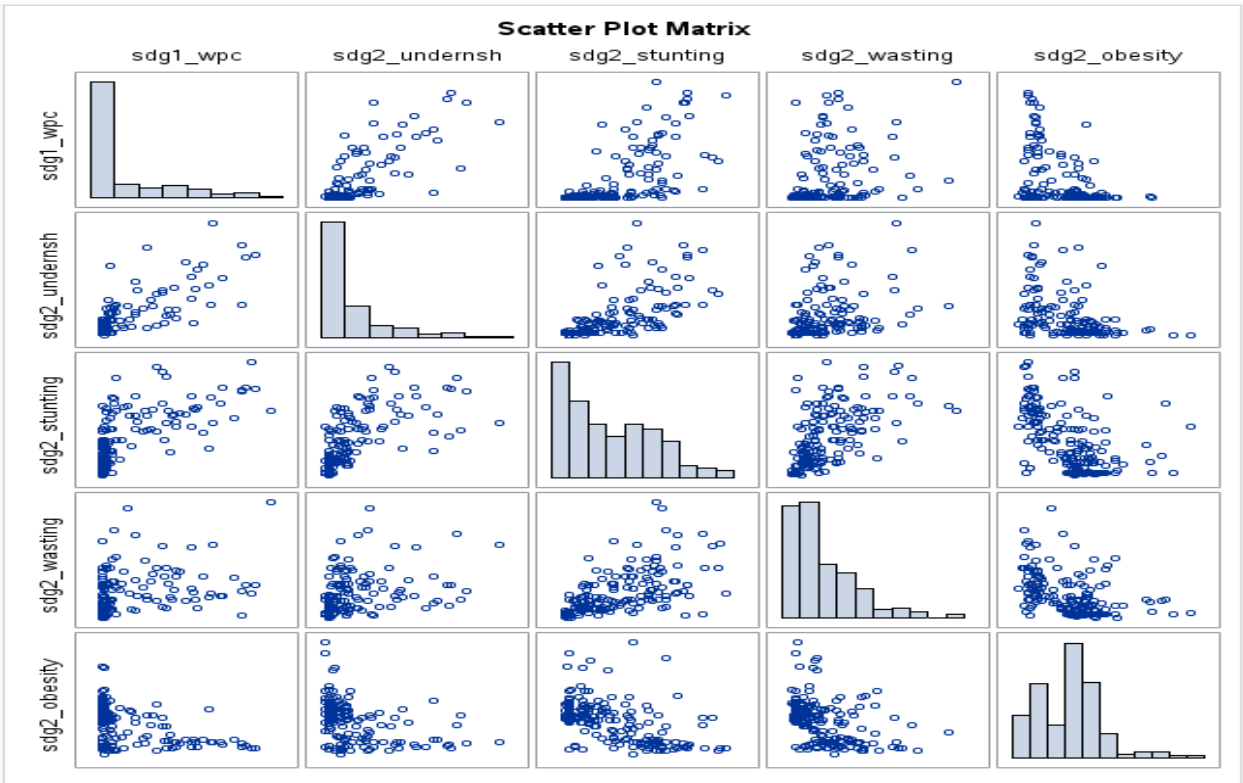
```
proc corr data=sdgs2021 plots=matrix(histogram) ;
    var sdg1_wpc sdg2_undernsh sdg2_stunting sdg2_wasting
        sdg2_obesity sdg3_incomeg sdg3_smoke sdg4_earlyedu
        sdg4_literacy sdg8_unemp;

run;
```

Use **Proc CORR** to evaluate pairwaise correlation and scatter plot on SDGs database and histogram on diagonal.

```
proc corr data=sdgs2021 plots=matrix(histogram) ;
    var sdg1_wpc sdg2_undernsh sdg2_stunting sdg2_wasting
        sdg2_obesity sdg3_incomeg sdg3_smoke sdg4_earlyedu
        sdg4_literacy sdg8_unemp;

run;
```



Principal Component Analysis : To be completed

Use **Proc Import** to import the two data files (Data1 and Data2 from the excel file tpsas_sql) and compare the data contents

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
2	age	Num	8	BEST.	age
1	id	Num	8	BEST.	id
4	men	Num	8	BEST.	men
3	seniority	Num	8	BEST.	seniority

Figure: Contents of Data1

Use **Proc Import** to import the two data files (Data1 and Data2 from the excel file tpsas_sql) and compare the data contents

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
2	age	Num	8	BEST.	age
1	id	Num	8	BEST.	id
4	men	Num	8	BEST.	men
3	seniority	Num	8	BEST.	seniority

Figure: Contents of Data1

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	ID	Num	8	BEST.		ID
4	marital	Char	1	$1.	$1.	marital
3	private	Num	8	BEST.		private
2	urban	Num	8	BEST.		urban
5	y	Num	8	BEST.		y

Figure: Contents of Data2

create a table from Data1 that contains id, age,seriorty for women

```
proc sql;
     create table women
     as select id, age, seniority
     from TP_SAS.data1
     where men=0;
run;
proc print dat=women(obs=5);
run;
```

create a table from Data1 that contains id, age,seriorty for women

```
proc sql;
     create table women
     as select id, age, seniority
     from TP_SAS.data1
     where men=0;
run;
proc print dat=women(obs=5);
run:
```

Obs	id	age	seniority
1	7	46	25
2	17	59	19
3	18	63	19
4	29	68	26
5	39	53	19

Figure: Five first observations of women data table

create a table from joining Data1 and Data2

```
proc sql;
    create table data
    as select * from TP_SAS.data1 as A, TP_SAS.data2 as B
    where A.id=B.id;
run;
proc print data =data (obs=5);
run;
```

create a table from joining Data1 and Data2

```
proc sql;
    create table data
    as select * from TP_SAS.data1 as A, TP_SAS.data2 as B
    where A.id=B.id;
run;
proc print data =data (obs=5);
run;
```

Obs	id	age	seniority	men	urban	private	marital	y
1	7	46	25	0	0	1	C	1
2	8	65	25	1	0	1	C	1
3	14	56	20	1	0	1	C	1
4	20	59	21	1	1	1	C	1
5	22	45	19	1	0	1	C	1

Figure: Five first observations of the data table

create a table from joining Data1 and Data2 (left joint)

```
proc sql;
    create table leftjoin
    as select * from TP_SAS.data1 as A
    left join TP_SAS.data2 as B
    on A.id=B.id;
run;
proc print data =leftjoin (obs=10);
run;
```

create a table from joining Data1 and Data2 (left joint)

```
proc sql;
    create table leftjoin
    as select * from TP_SAS.data1 as A
    left join TP_SAS.data2 as B
    on A.id=B.id;
run;
proc print data =leftjoin (obs=10);
run;
```

Obs	id	age	seniority	men	urban	private	marital	y
1	1	58	32	1	.	.		.
2	2	37	30	1	.	.		.
3	3	56	28	1	.	.		.
4	4	68	27	1	.	.		.
5	5	58	26	1	.	.		.
6	6	57	26	1	.	.		.
7	7	46	25	0	0	1	C	1
8	8	65	25	1	0	1	C	1
9	9	64	24	1	.	.		.
10	10	53	22	1	.	.		.

Figure: Five first observations of the data table

create a table from joining Data1 and Data2 (right joint)

```
proc sql;
    create table rightjoin
    as select * from TP_SAS.data1 as A
    right join TP_SAS.data2 as B
    on A.id=B.id;
run;
proc print data =rightjoin (obs=10);
run;
```

create a table from joining Data1 and Data2 (right joint)

```
proc sql;
    create table rightjoin
    as select * from TP_SAS.data1 as A
    right join TP_SAS.data2 as B
    on A.id=B.id;
run;
proc print data =rightjoin (obs=10);
run;
```

Obs	id	age	seniority	men	urban	private	marital	y
1	7	46	25	0	0	1	C	1
2	8	65	25	1	0	1	C	1
3	14	56	20	1	0	1	C	1
4	20	59	21	1	1	1	C	1
5	22	45	19	1	0	1	C	1
6	27	65	19	1	0	1	C	0
7	30	53	20	1	0	1	C	1
8	36	54	21	1	0	1	C	1
9	38	66	18	1	0	1	C	0
10	39	53	19	0	1	1	O	1

Figure: Five first observations of the data table

create a table descriptive statistics by region and sex for the variable "age"

```
title "Descripive statistics for Age by region and sex";
proc sql;
    select  men, urban, count(men) as N ,
    round(mean(age),0.01) as Mean,
    median(Age) as Median,
    round(stderr(age),0.0001) as SdtError
    from innerjoin
    group by urban,men ;
quit;
title "";
```

create a table descriptive statistics by region and sex for the variable "age"

```
title "Descripive statistics for Age by region and sex";
proc sql;
    select  men, urban, count(men) as N ,
    round(mean(age),0.01) as Mean,
    median(Age) as Median,
    round(stderr(age),0.0001) as SdtError
    from innerjoin
    group by urban,men ;
quit;
title "";
```

Descripive statistics for Age by region and sex

men	urban	N	Mean	Median	SdtError
Female	Rural	4	51.75	52.5	4.2303
Male	Rural	63	53.14	53	1.3781
Female	Urban	8	41.88	42.5	1.9127
Male	Urban	25	55.76	60	1.6576

MERCI POUR VOTRE ATTENTION

Pr. Mokhtar KOUKI
mokhtar.kouki@essai.ucar.tn