Running the First Program

Examining Frequency Distribution

(Data Management and Visualization Week 2 Assignment)

Expected Activities

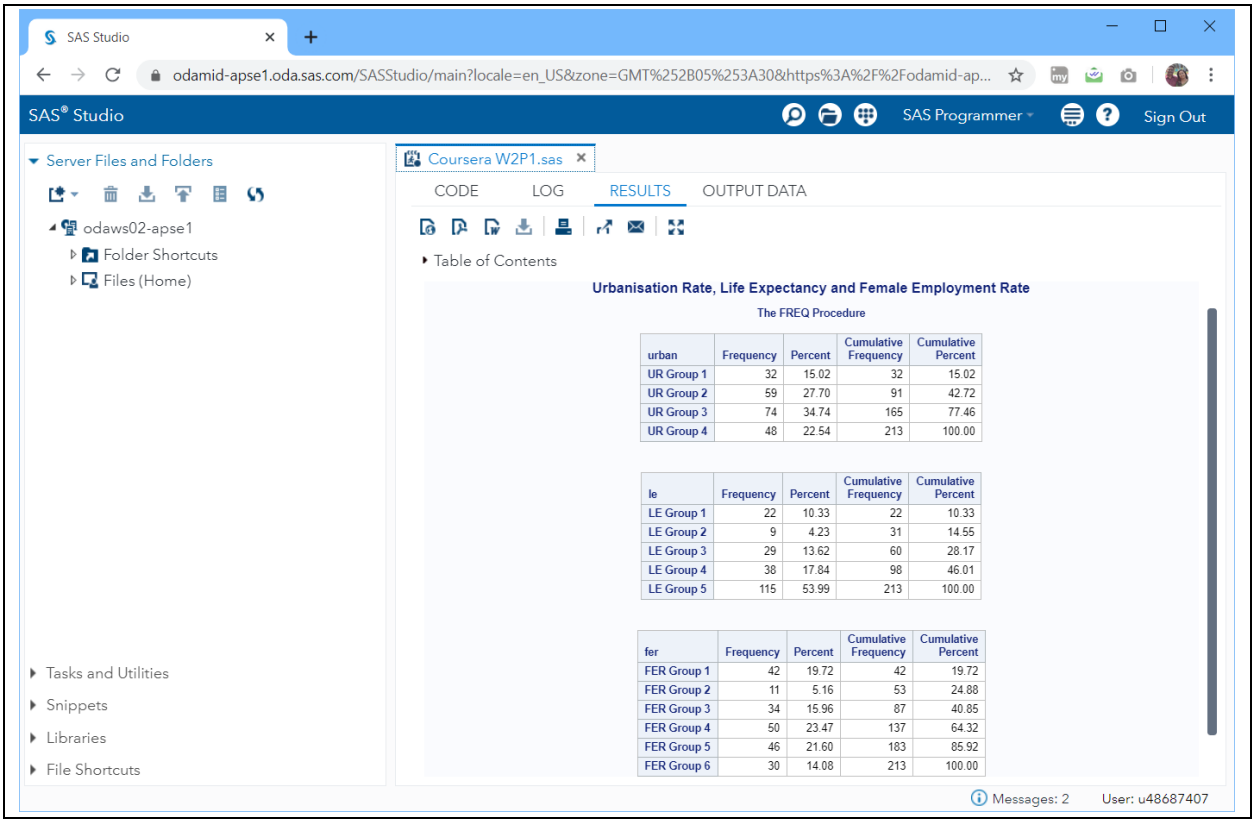
STEP 1: Run your first program

STEP 2: Run frequency distributions for your chosen variables and select columns, and possibly rows.

My First SAS Program

```
LIBNAME mydata "/courses/d1406ae5ba27fe300 " access=readonly;
DATA new; set mydata.gapminder;
KEEP country lifeexpectancy urbanrate femaleemployrate urban le fer;

/* Data Preparation or Management for variable urbanrate */

if urbanrate < 25 then urban = "UR Group 1";
if urbanrate >= 25 and urbanrate < 50 then urban = "UR Group 2";
if urbanrate >= 50 and urbanrate < 75 then urban = "UR Group 3";
if urbanrate >= 75 then urban = "UR Group 4";

/* Data Preparation or Management for variable lifeexpectancy */

if lifeexpectancy <  40 then le = "LE Group 1";
if lifeexpectancy >= 40 and lifeexpectancy < 50 then le = "LE Group 2";
if lifeexpectancy >= 50 and lifeexpectancy < 60 then le = "LE Group 3";
if lifeexpectancy >= 60 and lifeexpectancy < 70 then le = "LE Group 4";
if lifeexpectancy >= 70 then le = "LE Group 5";

/* Data Preparation or Management for variable femaleemployrate */

if femaleemployrate <  20 then fer = "FER Group 1";
if femaleemployrate >= 20 and femaleemployrate < 30 then fer = "FER Group 2";
if femaleemployrate >= 30 and femaleemployrate < 40 then fer = "FER Group 3";
if femaleemployrate >= 40 and femaleemployrate < 50 then fer = "FER Group 4";
if femaleemployrate >= 50 and femaleemployrate < 60 then fer = "FER Group 5";
if femaleemployrate >= 60 then fer = "FER Group 6";

PROC FREQ; TABLES urban le fer;
Title 'Frequency Tables';
Title2 'Urbanisation Rate, Life Expectancy and Female Employment Rate';
RUN;
```

Output Frequency Tables



urban	Frequency	Percent	Cumulative Frequency	Cumulative Percent
UR Group 1	32	15.02	32	15.02
UR Group 2	59	27.70	91	42.72
UR Group 3	74	34.74	165	77.46
UR Group 4	48	22.54	213	100.00

le	Frequency	Percent	Cumulative Frequency	Cumulative Percent
LE Group 1	22	10.33	22	10.33
LE Group 2	9	4.23	31	14.55
LE Group 3	29	13.62	60	28.17
LE Group 4	38	17.84	98	46.01
LE Group 5	115	53.99	213	100.00

fer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
FER Group 1	42	19.72	42	19.72
FER Group 2	11	5.16	53	24.88
FER Group 3	34	15.96	87	40.85
FER Group 4	50	23.47	137	64.32
FER Group 5	46	21.60	183	85.92
FER Group 6	30	14.08	213	100.00

Frequency distributions are created for the following three variables from the Gapminder dataset.

- urbanrate
- lifeexpectancy
- femaleemployrate

Urbanization Rate is classified into four different groups

- UR Group 1 (urbanrate < 25)
- UR Group 2 (25 <= urbanrate < 50)
- UR Group 3 (50 <= urbanrate < 75)
- UR Group 4 (urbanrate >= 75)

Majority of records are falling under UR Group 3 (~ 35%) followed by UR Group 2 (~28%). The lowest frequency is observed in UR Group 1(~15%).

Life Expectancy is classified into five different groups

- LE Group 1 (lifeexpectancy < 40)
- LE Group 2 (40 <= lifeexpectancy < 50)
- LE Group 3 (50 <= lifeexpectancy < 60)
- LE Group 4 (60 <= lifeexpectancy < 70)
- LE Group 5 (lifeexpectancy >= 70)

Majority of records are falling under LE Group 5 (~ 54%) and the lowest frequency is observed in LE Group 2 (~4%).

Female Employment Rate is classified into six different groups

- FER Group 1 (femaleemployrate < 20)
- FER Group 2 (20 <= femaleemployrate < 30)
- FER Group 3 (30 <= femaleemployrate < 40)
- FER Group 4 (40 <= femaleemployrate < 50)
- FER Group 5 (50 <= femaleemployrate < 60)
- FER Group 6 (femaleemployrate >= 60)

Majority of the data is falling under FER Group 4 (~ 23%) followed by FER Group 5 (~22%). The lowest frequency is observed in FER Group 2 (~5%).