Name: Zhijian Liu

Course: STAT-625

Homework#5

Exercise#1:

```
1. Program
    ∃Data Voter;
      Input Age Party : $1. (Quest1-Quest4)($1. +1);
      label Quest1='The president is doing a good job'
             Quest2='Congress is doing a good job'
             Quest3='Taxes are too high'
             Quest4='Government should cut spending';
      Datalines;
      23 D 1 1 2 2
      45 R 5 5 4 1
      67 D 2 4 3 3
      39 R 4 4 4 4
      19 D 2 1 2 1
      75 D 3 3 2 3
      57 R 4 3 4 4
      run;
   Log
   1
          Data Voter;
          Input Age Party : $1. (Quest1-Quest4)($1. +1);
label Quest1='The president is doing a good job'
Quest2='Congress is doing a good job'
Quest3='Taxes are too high'
    2
    3
    4
5
    6
                  Quest4='Government should cut spending';
          Datalines;
   NOTE: The data set WORK.VOTER has 7 observations and 6 variables.
NOTE: DATA statement used (Total process time):
           real time
                                     0.00 seconds
           cpu time
                                     0.01 seconds
    15
    16
          run;
2. Program
   proc print data = Voter;
     run;
   ∃proc freq data = Voter;
          table (Quest1-Quest4);
     run;
```

The SAS System

Obs	Age	Party	Quest1	Quest2	Quest3	Quest4
1	23	D	1	1	2	2
2	45	R	5	5	4	1
3	67	D	2	4	3	3
4	39	R	4	4	4	4
5	19	D	2	1	2	1
6	75	D	3	3	2	3
7	57	R	4	3	4	4

The FREQ Procedure

	The president is doing a good job					
Quest1	Frequency	Percent	Cumulative Frequency	Cumulative Percent		
1	1	14.29	1	14.29		
2	2	28.57	3	42.86		
3	1	14.29	4	57.14		
4	2	28.57	6	85.71		
5	1	14.29	7	100.00		

	Congress is doing a good job					
Quest2	Frequency	Percent	Cumulative Frequency	Cumulative Percent		
1	2	28.57	2	28.57		
3	2	28.57	4	57.14		
4	2	28.57	6	85.71		
5	1	14.29	7	100.00		

	Taxes are too high						
Quest3	Quest3 Frequency Percent Cumulative Frequency Percent						
2	3	42.86	3	42.86			
3	1	14.29	4	57.14			
4	3	42.86	7	100.00			

	Government should cut spending					
Quest4	Frequency	Percent	Cumulative Frequency	Cumulative Percent		
1	2	28.57	2	28.57		
2	1	14.29	3	42.86		
3	2	28.57	5	71.43		
4	2	28.57	7	100.00		

3. Program

```
□proc format;
     value Age 0-25 = '0-25'
               26-50 = '26-50'
               51-75 = '51-75'
               76 - high = '>75';
     value $Quest '1' = 'Strongly Disagree'
                   '2' = 'Disagree'
                   '3' = 'No opinion'
                   '4' = 'Agree'
                   '5' = 'Strongly Agree';
 run;
∃Data Voter;
 Input Age Party : $1. (Quest1-Quest4)($1. +1);
 label Quest1='The president is doing a good job'
       Quest2='Congress is doing a good job'
       Quest3='Taxes are too high'
       Quest4='Government should cut spending';
 Format Age Age.
        Quest1-Quest4 $Quest.;
 Datalines;
 23 D 1 1 2 2
 45 R 5 5 4 1
 67 D 2 4 3 3
 39 R 4 4 4 4
 19 D 2 1 2 1
 75 D 3 3 2 3
 57 R 4 3 4 4
 run;
```

```
1
         proc format;
              value Age 0-25 = '0-25'
26-50 = '26-50'
51-75 = '51-75'
   2
3
4
   5
                           76 - high = '>75';
   NOTE: Format AGE has been output.
6 value $Quest '1' = 'Strongly Disagree'
                               '2' = 'Disagree'
   7
                              '3' = 'No opinion'
'4' = 'Agree'
'5' = 'Strongly Agree';
   8
   9
   10
   NOTE: Format $QUEST has been output.
   11
         run;
   NOTE: PROCEDURE FORMAT used (Total process time):
                                   0.05 seconds
           real time
           cpu time
                                   0.01 seconds
   12
         Data Voter;
          Input Age Party : $1. (Quest1-Quest4)($1. +1);
   13
   14
          label Quest1='The president is doing a good job'
                 Quest2='Congress is doing a good job'
   15
                 Quest3='Taxes are too high'
Quest4='Government should cut spending';
   16
   17
   18
         Format Age Age.
                  Quest1-Quest4 $Quest.;
   19
   20
         Datalines;
   NOTE: The data set WORK.VOTER has 7 observations and 6 variables.
NOTE: DATA statement used (Total process time):
           real time
                                   0.01 seconds
           cpu time
                                   0.01 seconds
   28
   29
         run;
4. Program
   ∃proc format;
          value $Questb '1', '2' = 'Generally Disagree'
                           '3' = 'No Option'
                           '4','5' = 'Generally Agree';
     title 'Frequencies for Questions 1 to 4';
   ∃proc freq data=Voter;
          table Quest1-Quest4;
          format Quest1-Quest4 $Questb.;
     run;
   Log
```

```
50
      proc format;
           value $Questb '1','2' = 'Generally Disagree'
'3' = 'No Option'
'4','5' = 'Generally Agree';
51
52
NOTE: Format $QUESTB is already on the library WORK.FORMATS.
NOTE: Format $QUESTB has been output.
      title 'Frequencies for Questions 1 to 4';
NOTE: PROCEDURE FORMAT used (Total process time):
                                0.00 seconds
0.00 seconds
       real time
       cpu time
      proc freq data=Voter;
56
57
           table Quest1-Quest4;
           format Quest1-Quest4 $Questb.;
58
59
NOTE: Writing HTML Body file: sashtml.htm
NOTE: There were 7 observations read from the data set WORK.VOTER.
NOTE: PROCEDURE FREQ used (Total process time):
       real time
                                0.31 seconds
0.15 seconds
       cpu time
```

Frequencies for Questions 1 to 4

The FREQ Procedure

The president is doing a good job						
Quest1 Frequency Percent Cumulative Frequency Percent						
Generally Disagree	3	42.86	3	42.86		
No Option	1	14.29	4	57.14		
Generally Agree	3	42.86	7	100.00		

Congress is doing a good job						
Quest2 Frequency Percent Cumulative Frequency Percent						
Generally Disagree	2	28.57	2	28.57		
No Option	2	28.57	4	57.14		
Generally Agree	3	42.86	7	100.00		

Taxes are too high						
Quest3 Frequency Percent Cumulative Frequency Percent						
Generally Disagree	3	42.86	3	42.86		
No Option	1	14.29	4	57.14		
Generally Agree	3	42.86	7	100.00		

Government should cut spending						
Quest4 Frequency Percent Cumulative Frequency Percent						
Generally Disagree	3	42.86	3	42.86		
No Option	2	28.57	5	71.43		
Generally Agree	2	28.57	7	100.00		

5. Program

```
libname folder 'C:\Users\zl1409a\Desktop';
∃data folder.Voter;
      Input Age Party : $1. (Quest1-Quest4)($1. +1);
  Datalines;
  23 D 1 1 2 2
  45 R 5 5 4 1
  67 D 2 4 3 3
  39 R 4 4 4 4
  19 D 2 1 2 1
  75 D 3 3 2 3
  57 R 4 3 4 4
 run;
Log
     libname folder 'C:\Users\zl1409a\Desktop';
NOTE: Libref FOLDER was successfully assigned as follows:
Engine: V9
      Physical Name: C:\Users\z11409a\Desktop
     data folder.Voter;
61
          Input Age Party : $1. (Quest1-Quest4)($1. +1);
62
63
     Datalines;
NOTE: The data set FOLDER.VOTER has 7 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time
                             0.00 seconds
      cpu time
                             0.00 seconds
71
72
     run;
```

6. Program

```
libname folder 'C:\Users\z11409a\Desktop';
 option fmtsearch = (folder);
Dproc format library = folder;
     value Age 0-25 = '0-25'
               26-50 = '26-50'
               51-75 = '51-75'
               76 - high = '>75';
     value $Quest '1' = 'Strongly Disagree'
                  '2' = 'Disagree'
                   '3' = 'No opinion'
                  '4' = 'Agree'
                  '5' = 'Strongly Agree';
 run;
∃data folder.Voter;
     set folder. Voter;
     label Quest1 = 'The president is doing a good job'
           Quest2 = 'Congress is doing a good job'
           Quest3= 'Taxes are too high'
           Quest4= 'Government should cut spending';
     format Age Age.
            Quest1-Quest4 $Quest.;
 run;
Dproc format library=folder fmtlib;
```

```
libname folder 'C:\Users\z11409a\Desktop';
NOTE: Libref FOLDER was successfully assigned as follows:
        Engine:
                            V9
        Physical Name: C:\Users\z11409a\Desktop
      option fmtsearch =(folder);
3
4
5
6
      proc format library = folder;
            value Age 0-25 = '0-25'
                         26-50 = '26-50'
51-75 = '51-75'
76 - high = '>75';
7
NOTE: Format AGE is already on the library FOLDER.FORMATS.

NOTE: Format AGE has been written to FOLDER.FORMATS.

8 value $Quest '1' = 'Strongly Disagree'

9 '2' = 'Disagree'
                              '3' = 'No opinion'
10
                              '4' = 'Agree'
11
12 '5' = 'Strongly Agree';
NOTE: Format $QUEST is already on the library FOLDER.FORMATS.
NOTE: Format $QUEST has been written to FOLDER.FORMATS.
13
     run;
NOTE: PROCEDURE FORMAT used (Total process time):
                                   0.01 seconds
        real time
        cpu time
                                   0.01 seconds
14
      data folder. Voter;
           set folder.Voter;
label Quest1 = 'The president is doing a good job'
Quest2 = 'Congress is doing a good job'
15
16
17
                    Quest3= 'Taxes are too high'
Quest4= 'Government should cut spending';
18
19
20
            format Age Age.
                     Quest1-Quest4 $Quest.;
21
22
      run:
NOTE: There were 7 observations read from the data set FOLDER.VOTER.
NOTE: The data set FOLDER.VOTER has 7 observations and 6 variables.
NOTE: DATA statement used (Total process time):
        real time
                                   0.01 seconds
        cpu time
                                   0.01 seconds
23
      proc format library=folder fmtlib;
24
      run;
```

		TH: 5 NUMBER OF VALUES: 4 40 DEFAULT LENGTH: 5 FUZZ: STD
START	END	LABEL (VER. V7 V8 29SEP2018:16:48:11)
0	25	0-25
26	50	26-50
51	75	51 - 75
76	HIGH	>75

	•	TH: 17 NUMBER OF VALUES: 5 40 DEFAULT LENGTH: 17 FUZZ: 0
START	END	LABEL (VER. V7 V8 29SEP2018:16:48:11)
1 2 3 4 5	1 2 3 4 5	Strongly Disagree Disagree No opinion Agree Strongly Agree

Exercise#2

1. Program

```
data RESULTS;
    Input E1 E2 E3 CS $7.+1 ID $1.;
    Datalines;
    0.844 0.76 0.913 Stat101 1
    0.889 0.73 0.93 Stat101 2
    0.90 0.77 0.84 Stat101 3
    0.74 0.69 0.61 Stat101 4
    0.94 0.76 0.31 Stat101 5
    0.87 0.73 0.85 Stat101 6
    0.26 0.74 0.90 Stat101 7
;
run;
```

```
1
        data RESULTS;
   2
             Input E1 E2 E3 CS $7.+1 ID $1.;
   3
             Datalines;
   NOTE: The data set WORK.RESULTS has 7 observations and 5 variables. NOTE: DATA statement used (Total process time):
         real time 0.01 seconds
         cpu time
                               0.01 seconds
   11
        run;
   12
2. Program
    title 'The Data Porportion of RESULTS';
   ∃proc print data = RESULTS;
    title 'The Descriptor Porportion of RESULTS';
   ∃proc contents data = RESULTS varnum;
    run;
   Log
        title 'The Data Porportion of RESULTS';
   13
        proc print data = RESULTS;
   NOTE: Writing HTML Body file: sashtml.htm
   15
       run;
   NOTE: There were 7 observations read from the data set WORK.RESULTS.
   NOTE: PROCEDURE PRINT used (Total process time):
         real time
cpu time
                               0.23 seconds
                               0.12 seconds
        title 'The Descriptor Porportion of RESULTS';
        proc contents data = RESULTS varnum;
   17
        run;
   NOTE: PROCEDURE CONTENTS used (Total process time):
                               0.09 seconds
0.01 seconds
         real time
         cpu time
```

The Data Porportion of RESULTS

Obs	E1	E2	E3	CS	ID
1	0.844	0.76	0.913	Stat101	1
2	0.889	0.73	0.930	Stat101	2
3	0.900	0.77	0.840	Stat101	3
4	0.740	0.69	0.610	Stat101	4
5	0.940	0.76	0.310	Stat101	5
6	0.870	0.73	0.850	Stat101	6
7	0.260	0.74	0.900	Stat101	7

The Descriptor Porportion of RESULTS

The CONTENTS Procedure

Data Set Name	WORK.RESULTS	Observations	7
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	09/29/2018 17:02:48	Observation Length	32
Last Modified	09/29/2018 17:02:48	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

This table describe the information of the data sets. The name of the data set is "WORK.RESULTS". The number of observations is 7, and there are 5 variables. The length of observation is 32.

Engine/Host Dependent Information			
Data Set Page Size	65536		
Number of Data Set Pages	1		
First Data Page	1		
Max Obs per Page	2039		
Obs in First Data Page	7		
Number of Data Set Repairs	0		
ExtendObsCounter	YES		
Filename	C:\Users\zl1409a\AppData\Local\Temp\SAS Temporary Files_TD4624_B13-15_\results.sas7bdat		
Release Created	9.0401M0		
Host Created	X64_7PRO		

This table provides more detail of the data. It path is

"C:\Users\zl1409a\AppData\Local\Temp\SAS Temporary Files_TD4624_B13-

15_\results.sas7bdat", and data set page size is 65536.

Va	Variables in Creation Order				
#	Variable	Туре	Len		
1	E1	Num	8		
2	E2	Num	8		
3	E3	Num	8		
4	CS	Char	7		
5	ID	Char	1		

This table describe the variables by showing variable names, their types and their length.

3. Program

```
∃data RESULTS;
     Input E1 E2 E3 CS $7.+1 ID $1.;
     label E1 = 'Exam 1 grade'
           E2 = 'Exam 2grade'
           E3 = 'Exam 3grade'
           CS ='Lass Section'
           ID ='Student ID';
     Datalines;
     0.844 0.76 0.913 Stat101 1
     0.889 0.73 0.93 Stat101 2
     0.90 0.77 0.84 Stat101 3
     0.74 0.69 0.61 Stat101 4
     0.94 0.76 0.31 Stat101 5
     0.87 0.73 0.85 Stat101 6
     0.26 0.74 0.90 Stat101 7
 run;
```

```
Log
   19
         data RESULTS;
   20
              Input E1 E2 E3 CS $7.+1 ID $1.;
              label E1 = 'Exam 1 grade'
E2 = 'Exam 2grade'
   21
   22
                    E3 = 'Exam 3grade'
   23
                     CS = 'Lass Section'
   24
                     ID ='Student ID';
   25
   26
              Datalines;
   NOTE: The data set WORK.RESULTS has 7 observations and 5 variables. NOTE: DATA statement used (Total process time):
          real time
                                  0.01 seconds
          cpu time
                                  0.01 seconds
4. Program
   ∃data RESULTS;
         Input E1 E2 E3 CS $7.+1 ID $1.;
         label E1 = 'Exam 1 grade'
                E2 = 'Exam 2grade'
                E3 = 'Exam 3grade'
                CS ='Lass Section'
                ID ='Student ID';
         Format E1-E3 percent5.;
         Datalines;
         0.844 0.76 0.913 Stat101 1
         0.889 0.73 0.93 Stat101 2
         0.90 0.77 0.84 Stat101 3
         0.74 0.69 0.61 Stat101 4
         0.94 0.76 0.31 Stat101 5
         0.87 0.73 0.85 Stat101 6
         0.26 0.74 0.90 Stat101 7
     run;
   Log
   1
         data RESULTS;
              Input E1 E2 E3 CS $7.+1 ID $1.;
label E1 ='Exam 1 grade'
E2 ='Exam 2grade'
   2
3
4
5
6
7
                     E3 ='Exam 3grade'
                     CS = 'Lass Section'
                     ID ='Student ID';
   8
              Format E1-E3 percent5.;
              Datalines;
   NOTE: The data set WORK.RESULTS has 7 observations and 5 variables.
   NOTE: DATA statement used (Total process time):
          real time
                                  0.00 seconds
                                  0.00 seconds
          cpu time
    17
   18
         run;
```

```
∃proc format;
     value scale low-<0.5 = 'F'
                  0.5 - < 0.6 = 'E'
                  0.6 - < 0.7 = 'D'
                  0.7 - < 0.8 = 'C'
                  0.8 - < 0.9 = 'B'
                  0.9-1 = 'A';
∃data RESULTS;
     Input E1 E2 E3 CS $7.+1 ID $1.;
     label E1 = 'Exam 1 grade'
           E2 = 'Exam 2grade'
           E3 = 'Exam 3grade'
           CS ='Lass Section'
            ID ='Student ID';
     Format E1-E3 scale.;
     Datalines;
     0.844 0.76 0.913 Stat101 1
     0.889 0.73 0.93 Stat101 2
     0.90 0.77 0.84 Stat101 3
     0.74 0.69 0.61 Stat101 4
     0.94 0.76 0.31 Stat101 5
     0.87 0.73 0.85 Stat101 6
   0.26 0.74 0.90 Stat101 7
 run;
□proc freq data=RESULTS;
     table E1-E3;
 run;
Log
```

```
proc format;
52
          value scale low-<0.5 = 'F'
                        0.5-<0.6 = 'E'
0.6-<0.7 = 'D'
53
54
55 0.7-(0.8 = 'C'

56 0.8-(0.9 = 'B'

57 0.9-1 = 'A';

NOTE: Format SCALE is already on the library WORK.FORMATS.
NOTE: Format SCALE has been output.
NOTE: PROCEDURE FORMAT used (Total process time):
       real time
                              0.00 seconds
       cpu time
                              0.00 seconds
     data RESULTS;
Input E1 E2 E3 CS $7.+1 ID $1.;
label E1 = 'Exam 1 grade'
58
59
60
                 E2 = 'Exam 2grade'
61
                 E3 = 'Exam 3grade'
62
                 CS = 'Lass Section'
63
                  ID ='Student ID';
64
65
          Format E1-E3 scale.;
66
          Datalines;
NOTE: The data set WORK.RESULTS has 7 observations and 5 variables.
NOTE: DATA statement used (Total process time):
       real time
                              0.01 seconds
                              0.01 seconds
       cpu time
74
      run;
75
76
      proc freq data=RESULTS;
77
         table E1-E3;
NOTE: There were 7 observations read from the data set WORK.RESULTS.
NOTE: PROCEDURE FREQ used (Total process time):
       real time
                              0.11 seconds
                              0.01 seconds
       cpu time
```

The FREQ Procedure

	Exam 1 grade					
E1	Frequency	Percent	Cumulative Frequency	Cumulative Percent		
F	1	14.29	1	14.29		
С	1	14.29	2	28.57		
В	3	42.86	5	71.43		
Α	2	28.57	7	100.00		

	Exam 2grade					
E2	Frequency	Cumulative Frequency				
D	1	14.29	1	14.29		
С	6	85.71	7	100.00		

	Exam 3grade					
E3	Frequency	Percent	Cumulative Frequency	Cumulative Percent		
F	1	14.29	1	14.29		
D	1	14.29	2	28.57		
В	2	28.57	4	57.14		
Α	3	42.86	7	100.00		

6. Program

```
□proc format;
     value ranges low-<0.6 = 'Weak'
                  0.6-0.8 = 'Medium'
                 0.8<-high = 'Good';
∃data RESULTS;
     Input E1 E2 E3 CS $7.+1 ID $1.;
     label E1 = 'Exam 1 grade'
           E2 = 'Exam 2grade'
           E3 ='Exam 3grade'
           CS ='Lass Section'
           ID ='Student ID';
     Format E1-E3 ranges.;
     Datalines;
     0.844 0.76 0.913 Stat101 1
     0.889 0.73 0.93 Stat101 2
     0.90 0.77 0.84 Stat101 3
     0.74 0.69 0.61 Stat101 4
     0.94 0.76 0.31 Stat101 5
     0.87 0.73 0.85 Stat101 6
  0.26 0.74 0.90 Stat101 7
 run;
□proc freq data=RESULTS;
    table E1-E3;
 run;
Log
```

```
26
      proc format;
           value ranges low-<0.6 = 'Weak'
0.6-0.8 = 'Medium'
27
28
29
                           0.8<-high = 'Good';
NOTE: Format RANGES has been output.
NOTE: PROCEDURE FORMAT used (Total process time):
                                0.01 seconds
       real time
       cpu time
                                0.01 seconds
30
      data RESULTS;
31
           Input E1 E2 E3 CS $7.+1 ID $1.;
32
           label E1 = 'Exam 1 grade'
                  E2 = 'Exam 2grade'
E3 = 'Exam 3grade'
CS = 'Lass Section'
33
34
35
36
                  ID ='Student ID';
           Format E1-E3 ranges.;
37
38
           Datalines;
NOTE: The data set WORK.RESULTS has 7 observations and 5 variables.
NOTE: DATA statement used (Total process time): real time 0.01 seconds
                                0.01 seconds
0.01 seconds
       cpu time
46
47
      run;
48
      proc freq data=RESULTS;
49
       table E1-E3;
50
      run;
NOTE: Writing HTML Body file: sashtml.html
NOTE: There were 7 observations read from the data set WORK.RESULTS.
NOTE: PROCEDURE FREQ used (Total process time):
       real time
                                0.28 seconds
0.09 seconds
       cpu time
```

The FREQ Procedure

Exam 1 grade					
E1	Frequency	Percent	Cumulative Frequency	Cumulative Percent	
Weak	1	14.29	1	14.29	
Medium	1	14.29	2	28.57	
Good	5	71.43	7	100.00	

Exam 2grade				
E2 Frequency Percent Cumulative Cumulative Percent				
Medium	7	100.00	7	100.00

	Exam 3grade					
E3 Frequency Percent Cumulative Frequency Percent						
Weak	1	14.29	1	14.29		
Medium	1	14.29	2	28.57		
Good	5	71.43	7	100.00		

7. Program

```
libname folder 'C:\Users\z11409a\Desktop';

proc format library = folder;

value ranges low-<0.6 = 'Weak'

0.6-0.8 = 'Medium'

0.8<-high = 'Good';

value scale low-<0.5 = 'F'

0.5-<0.6 = 'E'

0.6-<0.7 = 'D'

0.7-<0.8 = 'C'

0.8-<0.9 = 'B'

run;
```

```
libname folder 'C:\Users\z11409a\Desktop';
NOTE: Libref FOLDER was successfully assigned as follows:
        Engine:
                             V9
        Physical Name: C:\Users\z11409a\Desktop
      proc format library = folder;
value ranges low-(0.6 = 'Weak'
0.6-0.8 = 'Medium'
0.8<-high = 'Good';
80
81
82
83
NOTE: Format RANGES has been written to FOLDER.FORMATS.

84 value scale low-<0.5 = 'F'
85 0.5-<0.6 = 'E'
                             0.6-<0.7 = 'D'
86
                             0.7-<0.8 = 'C'
87
                             0.8-<0.9 = 'B'
0.9-1 = 'A';
88
NOTE: Format SCALE has been written to FOLDER.FORMATS.
90
       run;
NOTE: PROCEDURE FORMAT used (Total process time): real time 0.09 \ {\rm seconds}
                                    0.01 seconds
        cpu time
```



