Introduction to Statistics and Data Science using *eStat*

Chapter 4 Data Summary Using Tables and Measures

4.2 Contingency Table for Two Variables

Jung Jin Lee
Professor of Soongsil University, Korea
Visiting Professor of ADA University, Azerbaijan

4.2 Contingency Table for Two Variables

- **Cross table or contingency table** is to summarize two categorical variables
  => table with rows and columns using possible values of two categorical variables,
  => examine data of two variables to count frequency of corresponding cells.
  => percentage of each cell for the sum of rows, for the sum of columns, for the total number of data
  => studying their associated characteristics

- Statistical analysis for cross table
  => independence test in Chapter 11.

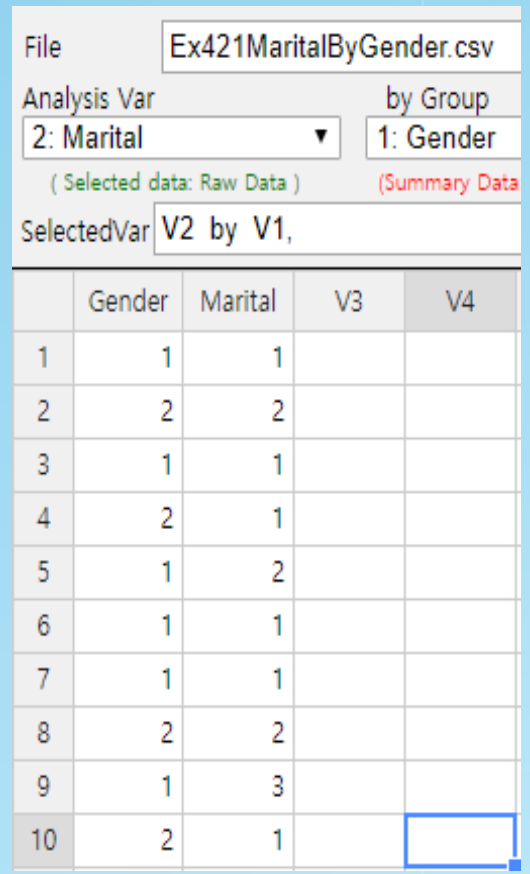
4.2 Contingency Table for Two Variables

[EX 4.2.1] The following table shows the survey data on gender (1: Male, 2: Female) and marital status (1: Single, 2: Married, 3: Other) which are used in [Example 2.2.3]. Find a cross table on marital status by gender.

Gender	Marital
1	1
2	2
1	1
2	1
1	2
1	1
1	1
2	2
1	3
2	1

<Answer>

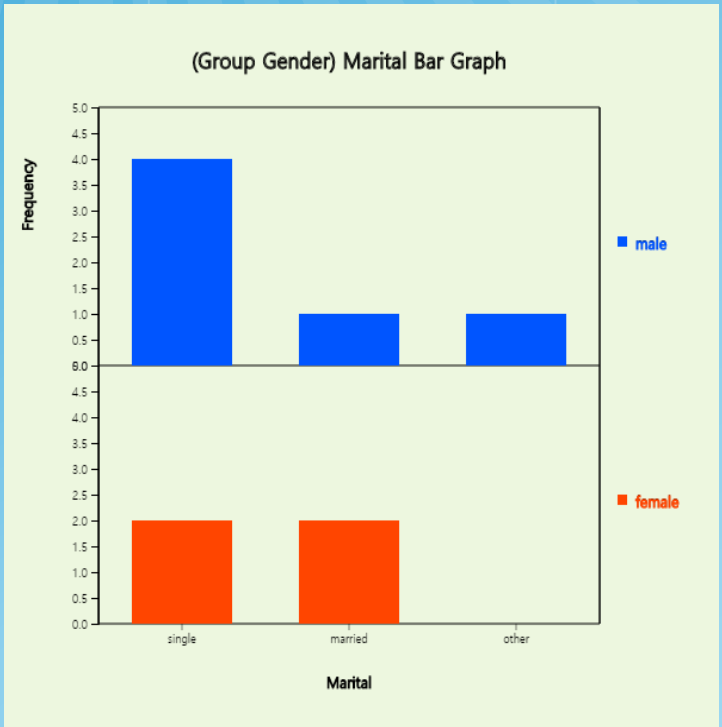
- Enter data in 『eStat』.
- Use [Edit Var] button to enter the variable name 'Gender' and the value labels 'Male' for 1 and 'Female' for 2.
- In the same way, enter the variable name 'Marital' and the value labels 'Single' for 1, 'Married' for 2 and 'Other' for 3.
- Data should be saved in JSON format by clicking on the icon .



	Gender	Marital	V3	V4
1	1	1		
2	2	2		
3	1	1		
4	2	1		
5	1	2		
6	1	1		
7	1	1		
8	2	2		
9	1	3		
10	2	1		

4.2 Contingency Table for Two Variables

- **Click on ‘Marital’ ('Analysis Var'), then ‘Gender’ ('by group'). Then you will see a bar chart of marital status by gender.**
- **Click ‘Frequency Table’ icon to display the cross table of marital status by gender in the log window. In the cross table, ‘by group’ variable becomes the row variable and ‘Analysis Var’ becomes the column variable.**



Cross Table	Col Variable	(Marital)		
Row Variable (Gender)	single	married	other	Total
male	4 66.7%	1 16.7%	1 16.7%	6 100%
female	2 50.0%	2 50.0%	0 0.0%	4 100%
Total	6 60.0%	3 30.0%	1 10.0%	10 100%
	Missing Observations	0		
Independence Test				
Sum of χ^2 value	1.667	deg of freedom	2	p-value 0.4346

4.2 Contingency Table for Two Variables

❖ Cross-table for two quantitative variables,

=> divide intervals in each variable

=> advisable to use statistical package such as R, SPSS, SAS

❖ One variable is categorical and the other is a continuous

=> Use ‘Histogram’ [Frequency Table]' of the 『eStat』

4.2 Contingency Table for Two Variables

[Ex 4.2.2] The data on the gender and age of a middle school teacher is at Ex ⇨ 02English ⇨ 032Continous_TeacherAgeByGender.csv. Use histogram module of 『eStat』 to create a cross table of age by gender.

<Ans>

- **Load the data and enter the value labels of 'Gender' as 'Male' for 1 and 'Female' for 2.**
- **After clicking the histogram icon with the mouse, select the 'Age' variable ('Analysis Var') and then the 'Gender' variable ('by group'). Histogram will appear.**

File: 032Continous_TeacherAgeByGen

Analysis Var: --- by Group: ---

(Selected data: Raw Data)

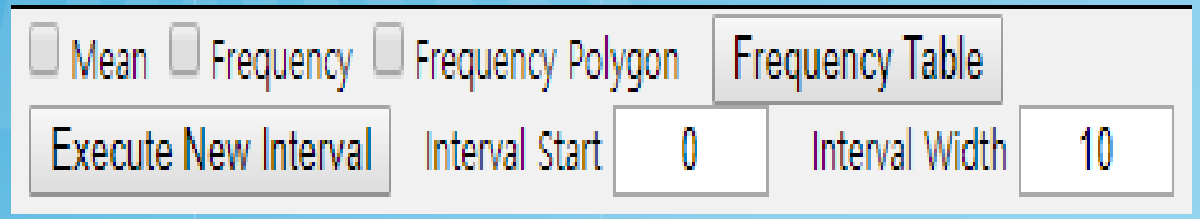
SelectedVar: V2 by V1,

	Gender	Age	V3	V4	V5
1	1	26			
2	1	34			
3	2	28			
4	2	39			
5	1	32			
6	1	36			
7	2	41			
8	2	42			
9	1	26			
10	1	25			
11	2	33			
12	2	43			
13	1	54			
14	1	49			
15	2	56			
16	2	31			
17	2	27			
18	1	42			
19	2	32			

4.2 Contingency Table for Two Variables

<Ans>

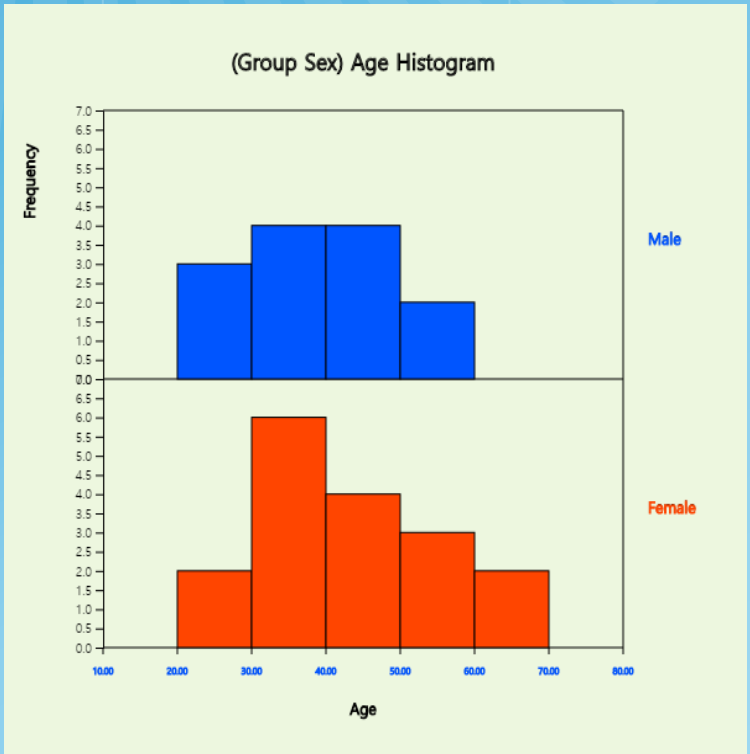
- **Click on the 'Frequency Table' below the graph, the cross table will appear in the log window.**



Histogram Frequency Table	Group Name	(Gender)	
Interval (Age)	Group 1 (Male)	Group 2 (Female)	Total
1 [25.00, 30.43)	3 (23.1%)	2 (11.8%)	5 (16.7%)
2 [30.43, 35.86)	3 (23.1%)	4 (23.5%)	7 (23.3%)
3 [35.86, 41.29)	1 (7.7%)	3 (17.6%)	4 (13.3%)
4 [41.29, 46.71)	3 (23.1%)	3 (17.6%)	6 (20.0%)
5 [46.71, 52.14)	1 (7.7%)	1 (5.9%)	2 (6.7%)
6 [52.14, 57.57)	1 (7.7%)	2 (11.8%)	3 (10.0%)
7 [57.57, 63.00)	1 (7.7%)	2 (11.8%)	3 (10.0%)
Total	13 (100%)	17 (100%)	30 (100%)

4.2 Contingency Table for Two Variables

- If the histogram interval is to be readjusted from 20 to 10 years apart, the histogram will be shown when you set the 'Start Interval' to 20 in the graph options and press the 'Run Interval' button.
- Click on the optional 'Frequency distribution table' to display the segmented frequency distribution table.



Histogram Frequency Table	Group Name	(Sex)	
Interval (Age)	Group 1 (Male)	Group 2 (Female)	Total
1 [20.00, 30.00)	3 (23.1%)	2 (11.8%)	5 (16.7%)
2 [30.00, 40.00)	4 (30.8%)	6 (35.3%)	10 (33.3%)
3 [40.00, 50.00)	4 (30.8%)	4 (23.5%)	8 (26.7%)
4 [50.00, 60.00)	2 (15.4%)	3 (17.6%)	5 (16.7%)
5 [60.00, 70.00)	0 (0.0%)	2 (11.8%)	2 (6.7%)
Total	13 (100%)	17 (100%)	30 (100%)

Thank you