Chapter 4

Data Summary Using Tables and Measures

Professor Jung Jin Lee

Table of Contents

Chapter 4 Data Summary with Tables and Measure

4.1 Frequency Table for Single Variable

4.2 Contingency Table for Two Variables

4.3 Summary Measures for Quantitative Variable

- 4.3.1 Measure of Central Tendency
- 4.3.2 Measure of Dispersion

4.1 Frequency Table for Single Variable

- Frequency table is simply a summary of the frequency at which the measurements appear.
  - to summarize categorical data
  - shows relative frequencies (i.e., percent), and the cumulative relative frequencies
- Based on these frequency tables, bar chart, pie chart and band graph are drawn.
- The frequency table can be also used for continuous data by dividing data into the intervals. After examine the frequency of each interval and draw up a frequency table.
- By using the frequency table, we can test a goodness of fit of data which is described at Chapter 11.

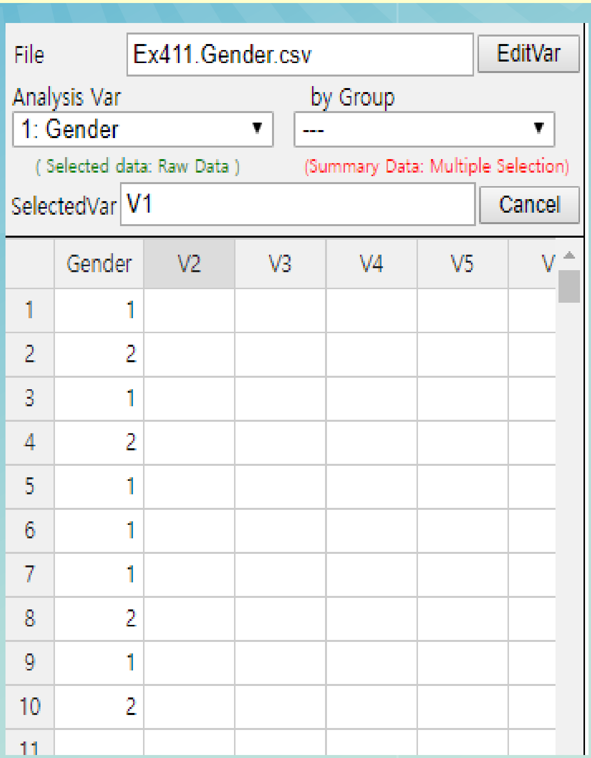
4.1 Frequency Table for Single Variable

[EX 4.1.1] Using the following gender data (1:Male, 2:Female), create the frequency table using 『eStat』.

Gender
1
2
1
2
1
1
1
2
1
2

〈Answer〉

- Enter data in『eStat』. Edit variable name, variable value using ‘EditVar’.



File: Ex411.Gender.csv

Analysis Var: 1: Gender

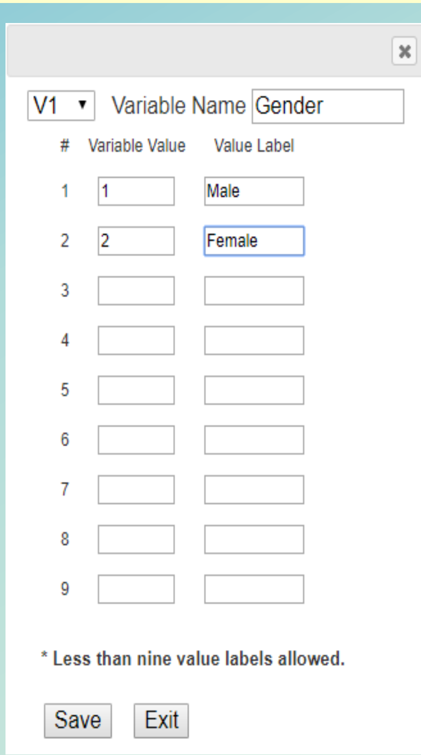
by Group: ---

(Selected data: Raw Data)

(Summary Data: Multiple Selection)

SelectedVar: V1

	Gender	V2	V3	V4	V5	V
1	1					
2	2					
3	1					
4	2					
5	1					
6	1					
7	1					
8	2					
9	1					
10	2					
11						



V1 Variable Name: Gender

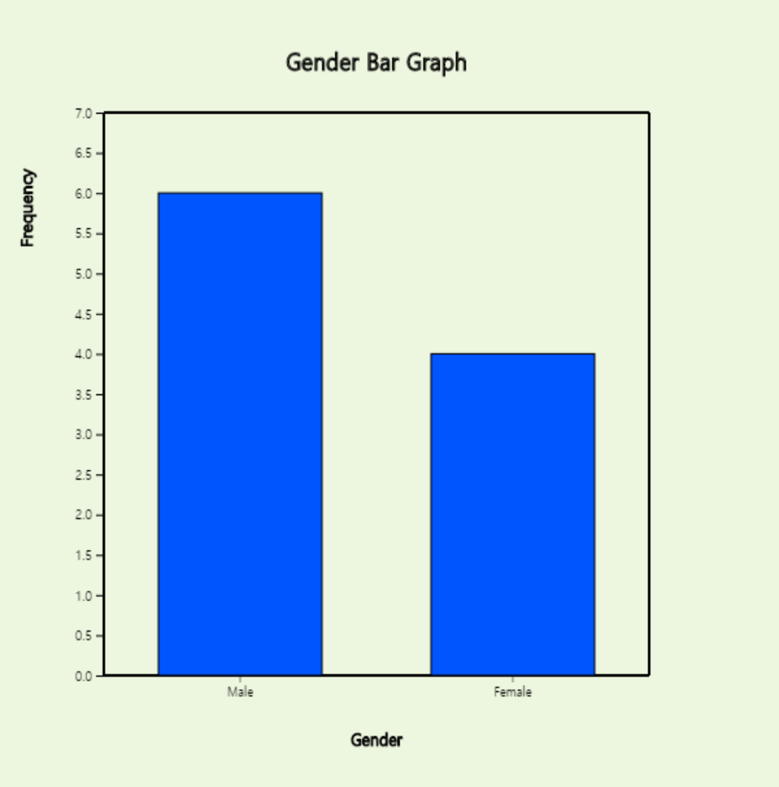
#	Variable Value	Value Label
1	1	Male
2	2	Female
3		
4		
5		
6		
7		
8		
9		

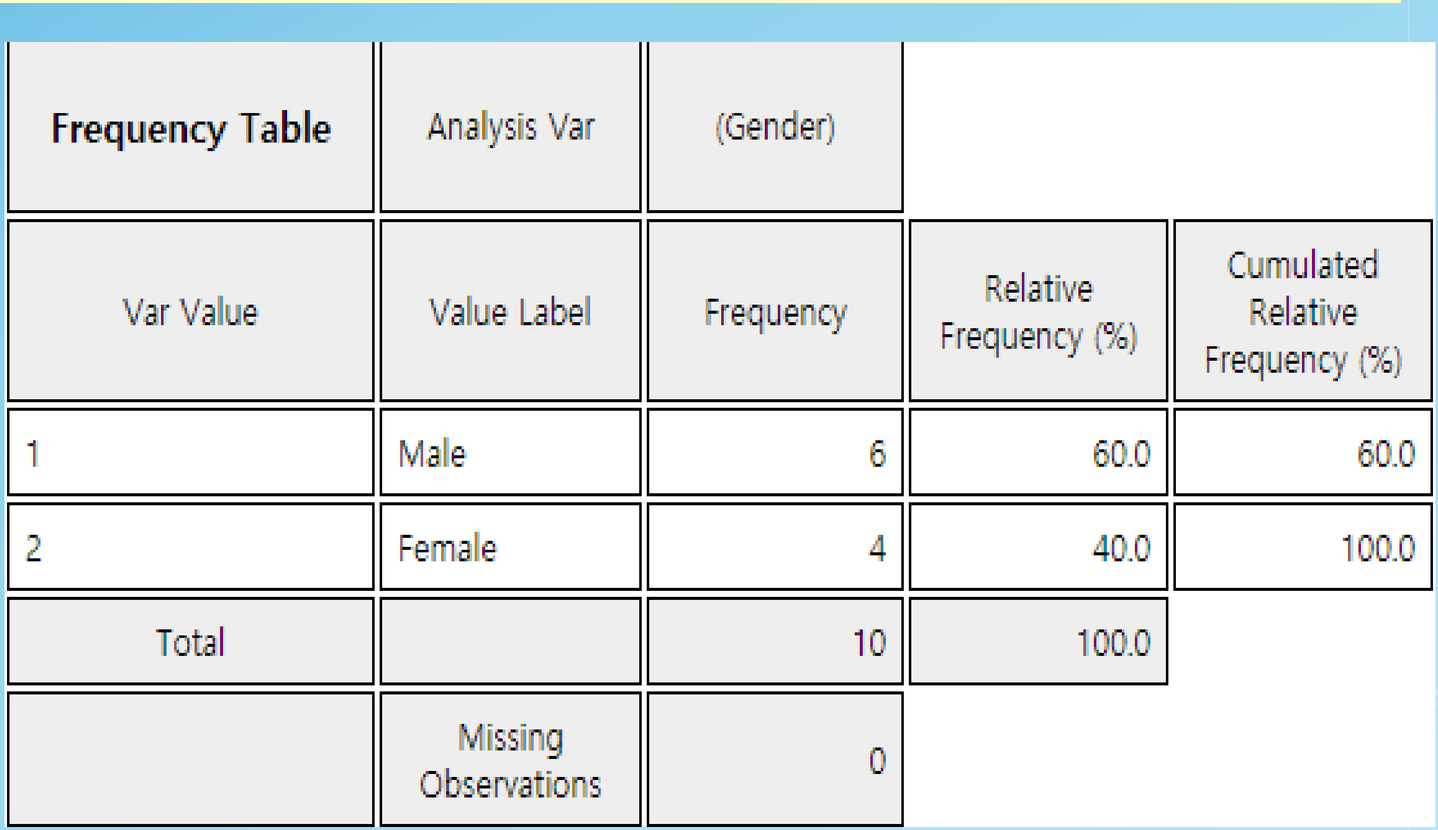
* Less than nine value labels allowed.

- (Note) After editing variable value, data should be saved as a JSON format to reload it again.

4.1 Frequency Table for Single Variable

- If you select the gender as the 'Analysis Var' in the variable selection box , a bar chart of the gender is drawn
- Then if you click the frequency table icon , the frequency table of the gender will appear in the log window





Frequency Table	Analysis Var	(Gender)		
Var Value	Value Label	Frequency	Relative Frequency (%)	Cumulated Relative Frequency (%)
1	Male	6	60.0	60.0
2	Female	4	40.0	100.0
Total		10	100.0	
	Missing Observations	0		

4.1 Frequency Table for Single Variable

4.1.2 Frequency Table of Continuous data

- In order to find a frequency table for continuous data, the data is divided into intervals, and the frequency of data belonging to each interval is investigated.
- Generally, we set up the intervals that do not overlap with each other and prepare the frequency table which shows the number of data in each interval.
- For this purpose, the maximum and maximum values are first obtained to determine the range of the data and then determine the number of intervals.

4.1 Frequency Table for Single Variable

4.1.2 Frequency Table of Continuous data

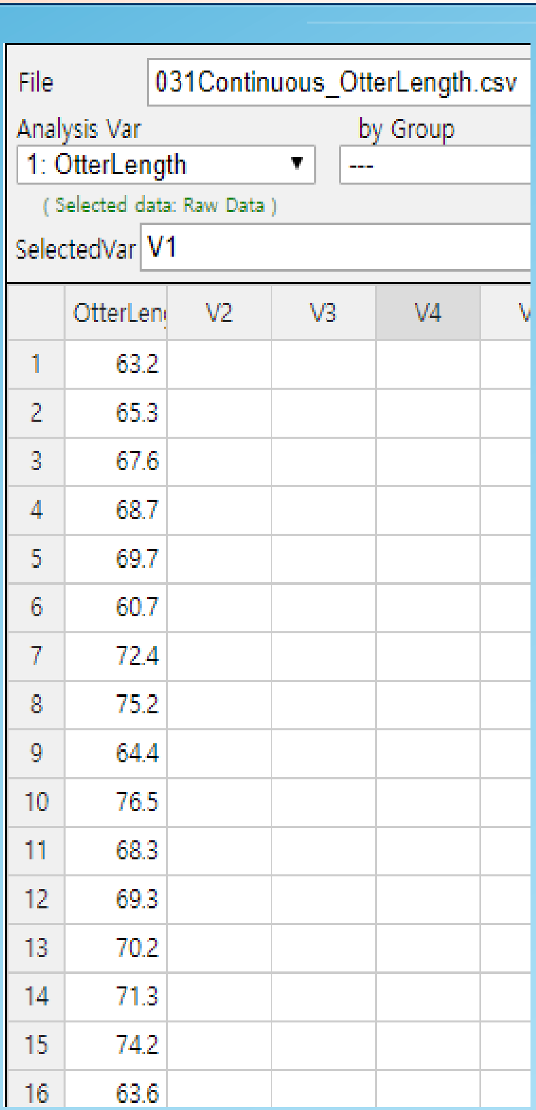
- 'How many intervals are you going to do? ' is an analyst's choice. Typically, the number of intervals is between 5 and 10 depending on the number of data.
- When the number of intervals is determined, the range of data (=maximum−maximum) is divided by the number of intervals to calculate the width of the interval.
- The start and end points of each interval are usually determined from '~ greater than equal ≤’ to ‘~ less than (〈)’.

4.1 Frequency Table for Single Variable

[Ex 4.1.2] The data of the otter length can be found at 『eStat』 Ex ⇨ 02English ⇨ 031Continuous_OtterLength.csv. Draw a histogram and its frequency table of the otter length by using 『eStat』

〈Ans〉

- Load the data by clicking Ex ⇨ 02English ⇨ 031Continuous_OtterLength.csv.
- Click the histogram icon and then select the variable name 'OtterLength' to draw a histogram.



File: 031Continuous_OtterLength.csv

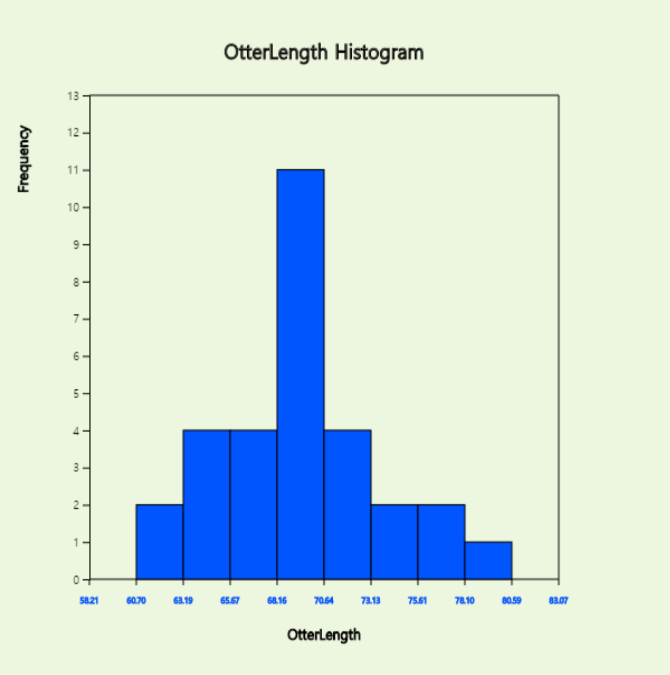
Analysis Var: 1: OtterLength

by Group: ---

(Selected data: Raw Data)

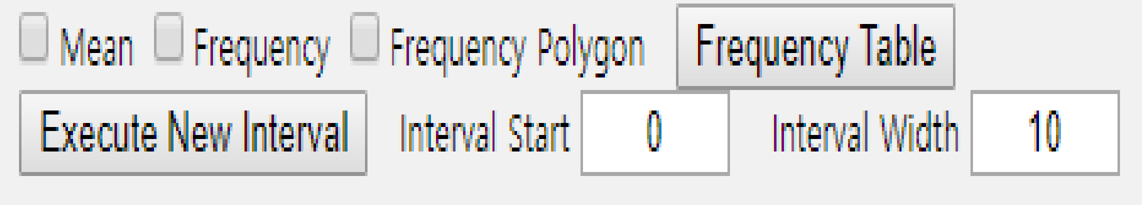
SelectedVar: V1

	OtterLen	V2	V3	V4	V
1	63.2				
2	65.3				
3	67.6				
4	68.7				
5	69.7				
6	60.7				
7	72.4				
8	75.2				
9	64.4				
10	76.5				
11	68.3				
12	69.3				
13	70.2				
14	71.3				
15	74.2				
16	63.6				



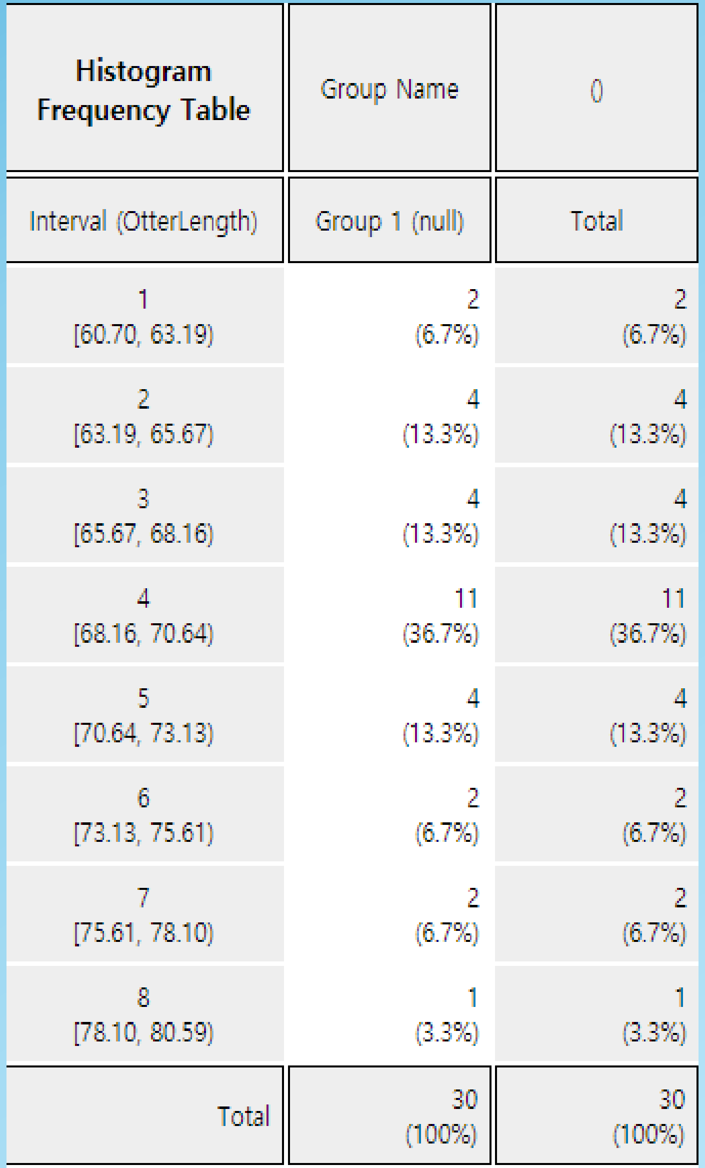
4.1 Frequency Table for Single Variable

- Click on the [Frequency Table] button in the options below the histogram.
- Then the frequency table of the histogram intervals is shown in the log window.



Interval Start 0

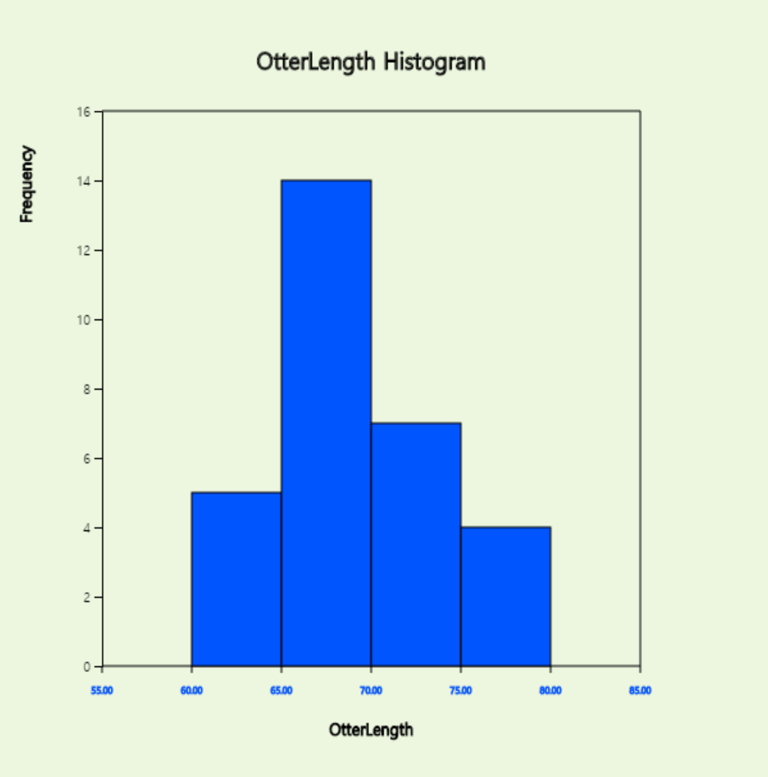
Interval Width 10

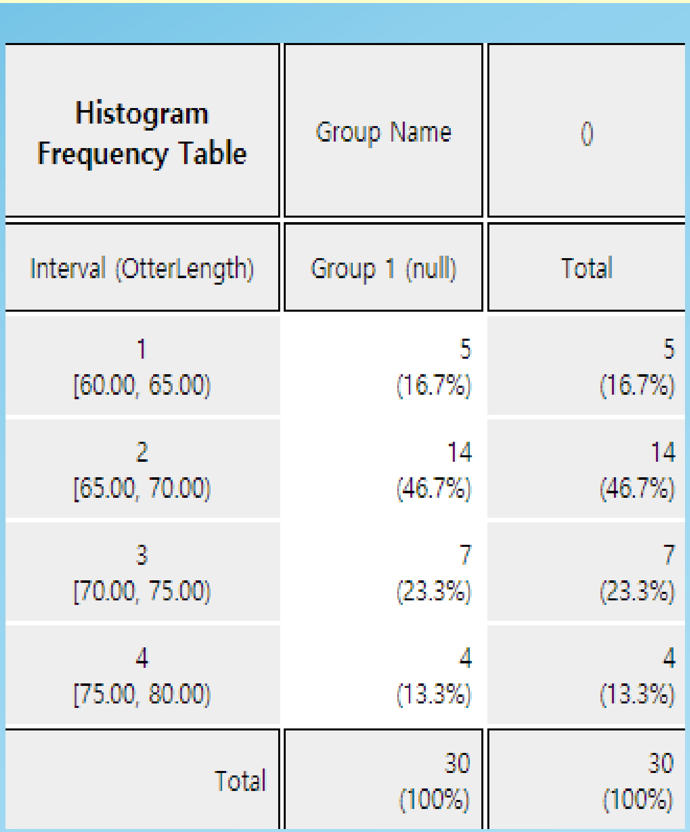


Histogram Frequency Table	Group Name	0
Interval (OtterLength)	Group 1 (null)	Total
1 [60.70, 63.19)	2 (6.7%)	2 (6.7%)
2 [63.19, 65.67)	4 (13.3%)	4 (13.3%)
3 [65.67, 68.16)	4 (13.3%)	4 (13.3%)
4 [68.16, 70.64)	11 (36.7%)	11 (36.7%)
5 [70.64, 73.13)	4 (13.3%)	4 (13.3%)
6 [73.13, 75.61)	2 (6.7%)	2 (6.7%)
7 [75.61, 78.10)	2 (6.7%)	2 (6.7%)
8 [78.10, 80.59)	1 (3.3%)	1 (3.3%)
Total	30 (100%)	30 (100%)

4.1 Frequency Table for Single Variable

- In order to adjust the histogram interval from 60kg with interval length of 5kg, set the ‘Interval Start’ to 60 and ‘Interval Width’ to 5 in the graph options.
- Press [Execute] New Interval] button to display the adjusted histogram.
- Click on [Frequency Table] button to reveal the new frequency table.





Histogram Frequency Table	Group Name	()
Interval (OtterLength)	Group 1 (null)	Total
1 [60.00, 65.00)	5 (16.7%)	5 (16.7%)
2 [65.00, 70.00)	14 (46.7%)	14 (46.7%)
3 [70.00, 75.00)	7 (23.3%)	7 (23.3%)
4 [75.00, 80.00)	4 (13.3%)	4 (13.3%)
Total	30 (100%)	30 (100%)

4.2 Contingency Table for Two Variables

- Cross table or contingency table is a very effective to summarize two categorical variables and studying their associated characteristics, similar to a frequency table of single variable.
- Cross table divides a table into rows and columns to create cells by using the possible variable values of the two variables, and then examine the data values of row and column variable for each data to examine the frequency of data belonging to the corresponding cells.
- For analysis, the percentage of each cell for the sum of rows, the percentage of each cell for the sum of columns, and the percentage of each cell for the total number of data are also shown below the frequency of each cell.

4.2 Contingency Table for Two Variables

- Cross tables are made for categorical data, but if continuous data is transformed into categorical by using intervals, you can create cross tables.
- If we examine the frequency distribution of a cross-table, we can see the association between the two variables.
- We can do statistical analyses such as the independence test of row and column variable, or homogeneity test, which are described in Chapter 11.

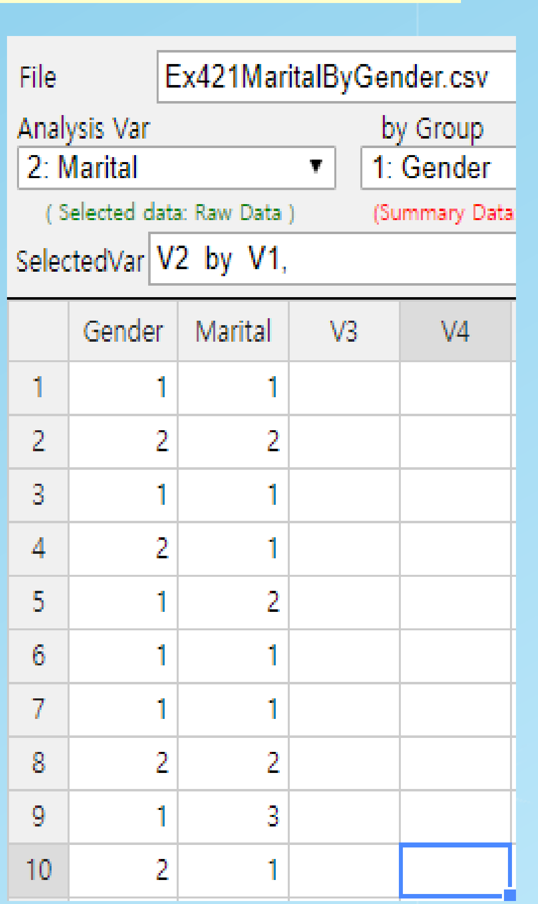
4.2 Contingency Table for Two Variables

[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 gender and marital status 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.
- The data should be saved in JSON format by clicking on the icon .



File: Ex421MaritalByGender.csv

Analysis Var: 2: Marital (Selected data: Raw Data)

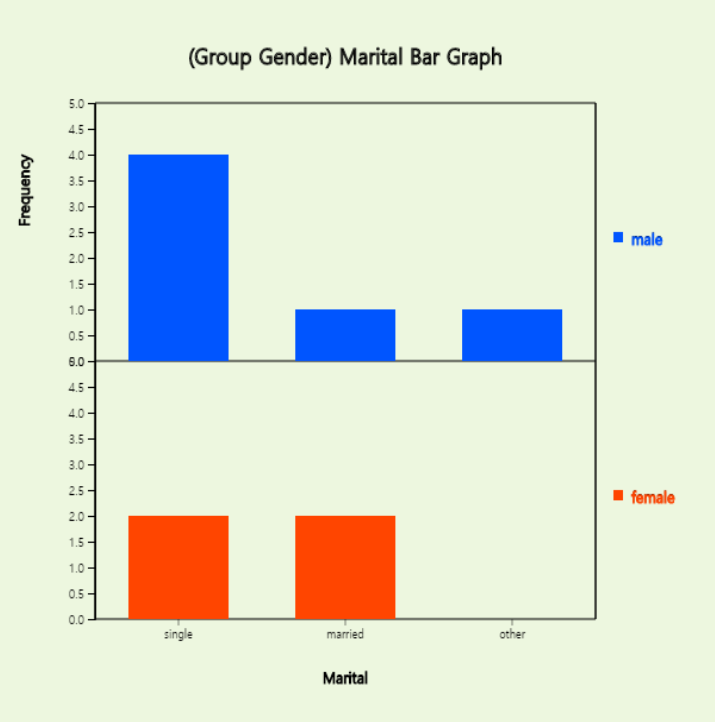
by Group: 1: Gender (Summary Data

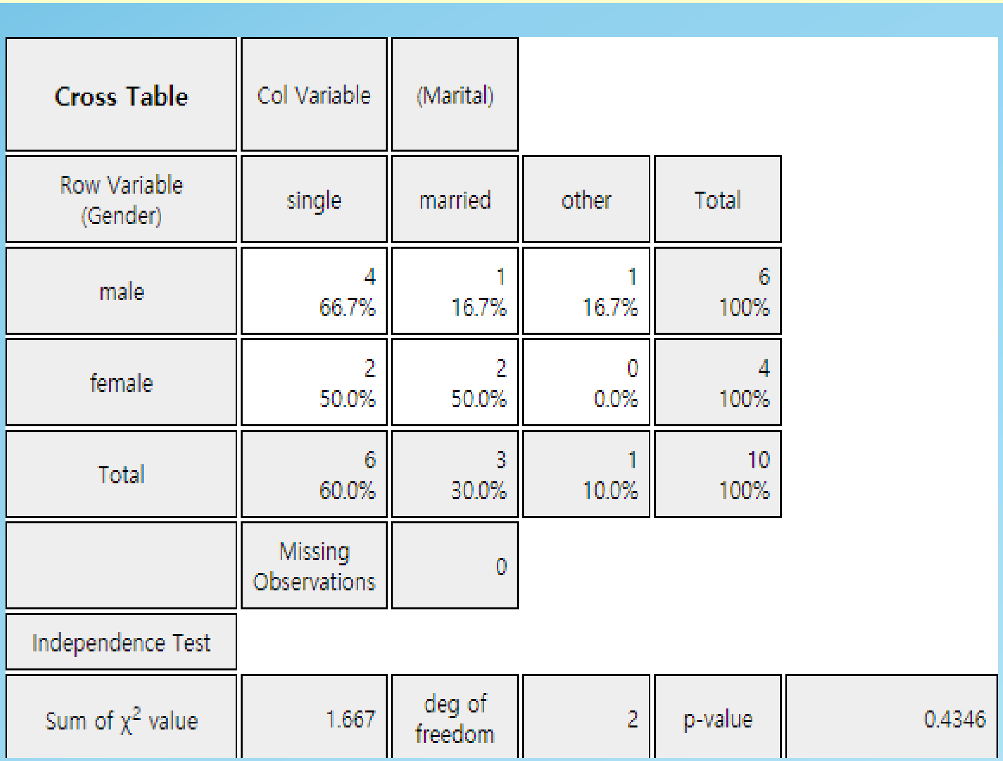
SelectedVar: V2 by V1,

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

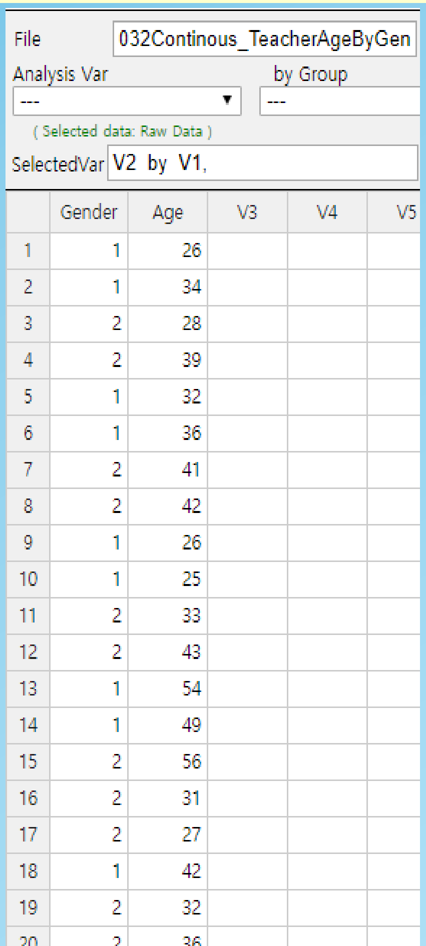
- To create a cross-table for two continuous variables, divide the intervals in each variable such as when creating a frequency table of single continuous variable.
- If both variables are continuous variables, it is advisable to use software such as Excel or R, SPSS, etc.
- If one variable is a categorical group variable and the other is a continuous variable, cross tables can be created using the [Frequency Table]' of the 『eStat』 histogram module.

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.



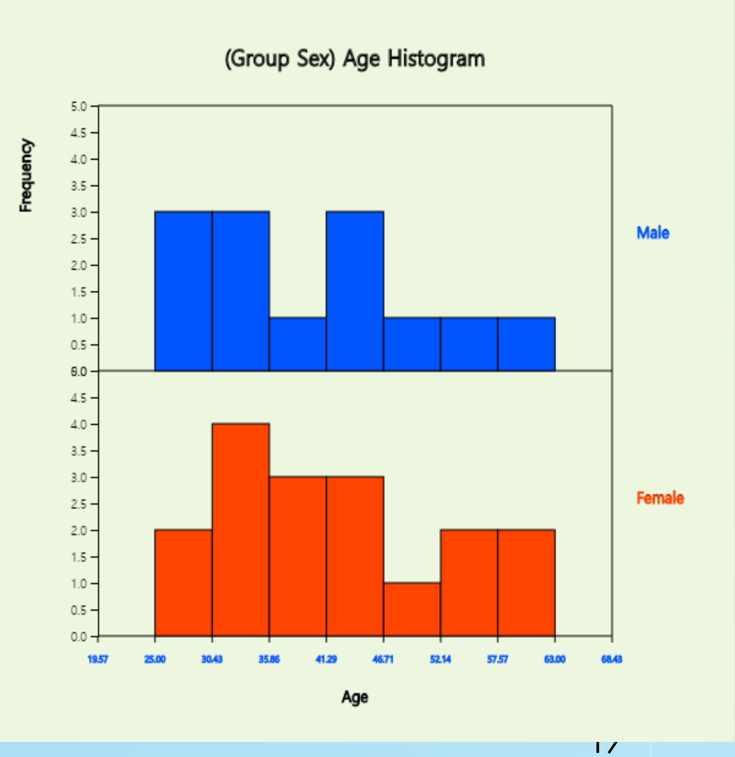
File 032Continous_TeacherAgeByGen

Analysis Var --- by Group ---

(Selected data: Raw Data)

SelectedVar V2 by V1,

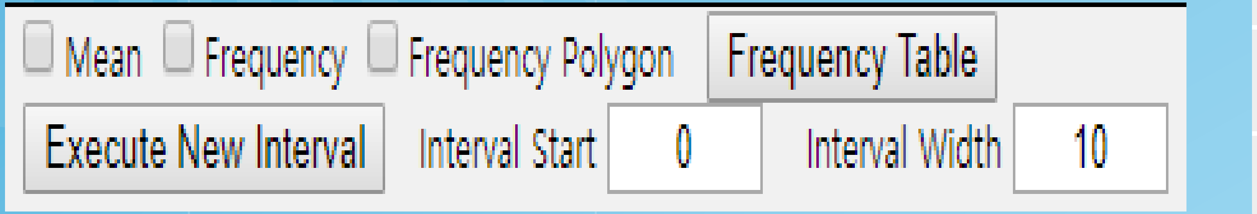
	Gender	Age	V3	V4	V5
1	1	26			
2	1	34			
3	2	28			
4	2	39			
5	1	32			
6	1	36			
7	2	41			
8	2	42			
9	1	26			
10	1	25			
11	2	33			
12	2	43			
13	1	54			
14	1	49			
15	2	56			
16	2	31			
17	2	27			
18	1	42			
19	2	32			
20	2	36			



4.2 Contingency Table for Two Variables

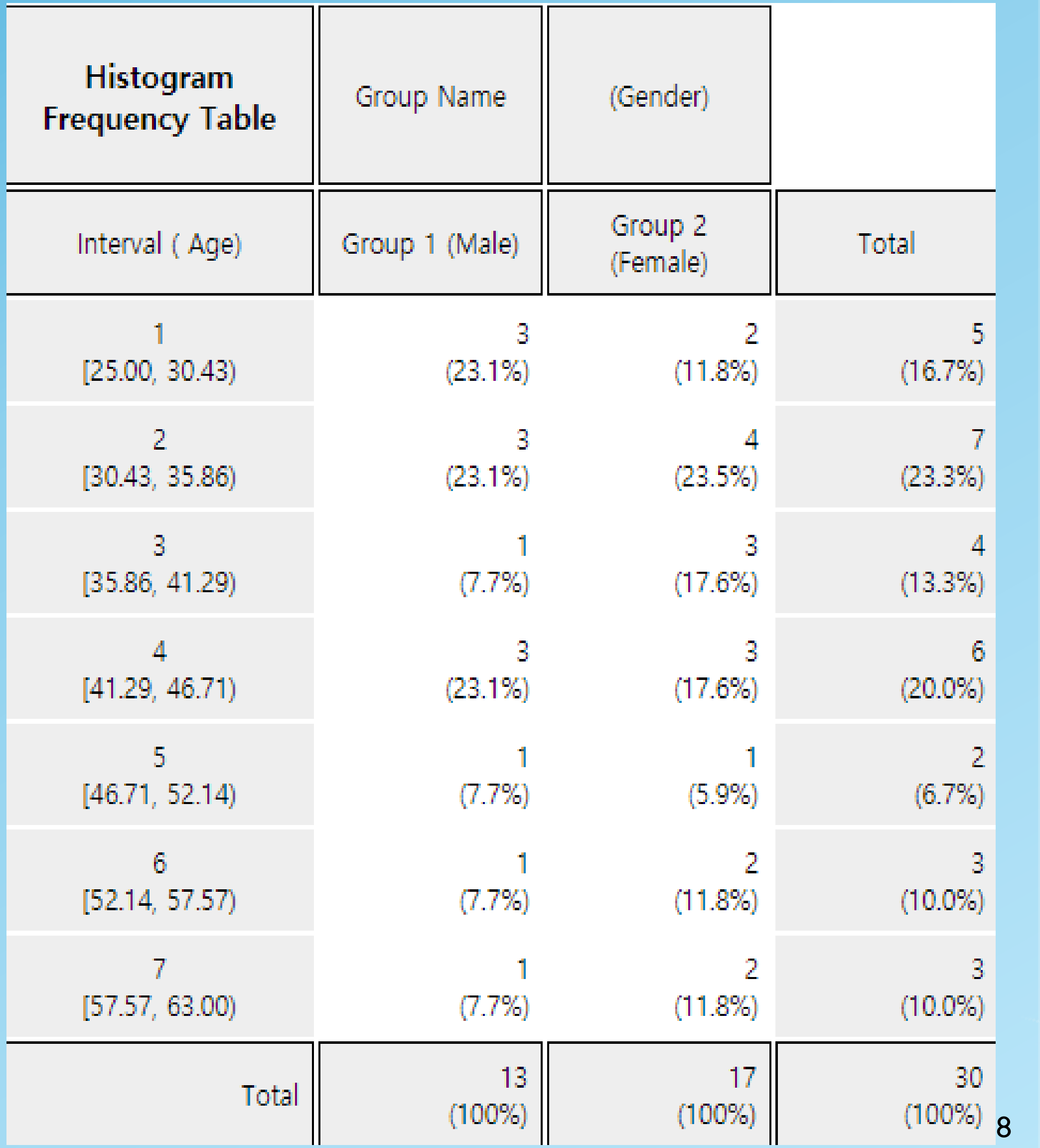
〈Ans〉

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



Interval Start 0

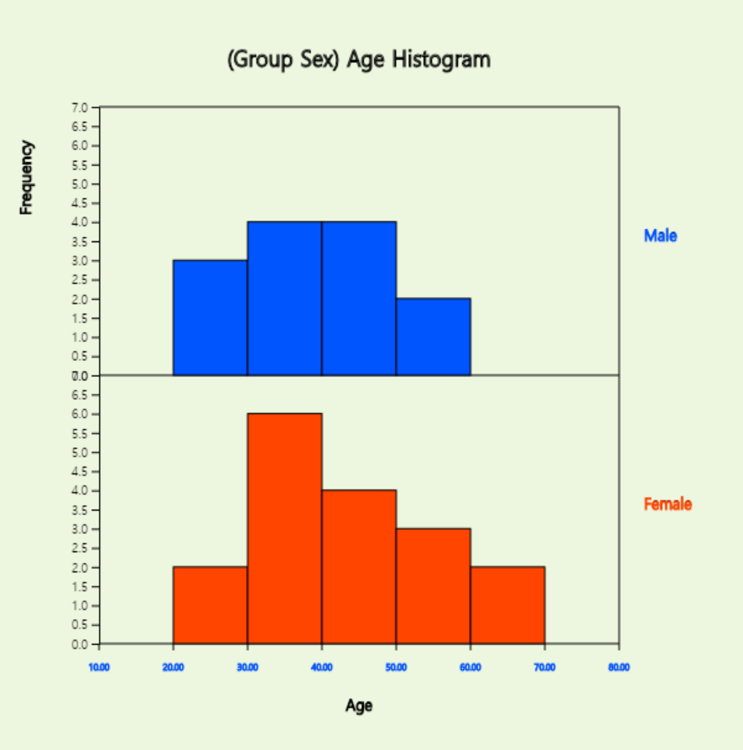
Interval Width 10

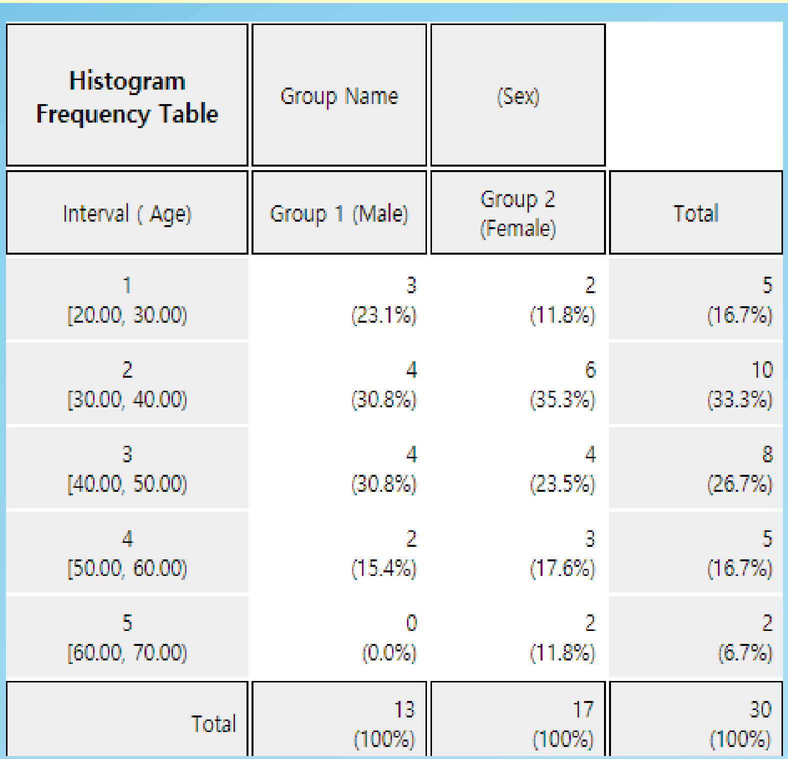


Histogram Frequency Table	Group Name	(Gender)	
Interval (Age)	Group 1 (Male)	Group 2 (Female)	Total
1 [25.00, 30.43)	3 (23.1%)	2 (11.8%)	5 (16.7%)
2 [30.43, 35.86)	3 (23.1%)	4 (23.5%)	7 (23.3%)
3 [35.86, 41.29)	1 (7.7%)	3 (17.6%)	4 (13.3%)
4 [41.29, 46.71)	3 (23.1%)	3 (17.6%)	6 (20.0%)
5 [46.71, 52.14)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

4.3 Summary Measures for Quantitative Variable

4.3.1 Measure of the central location

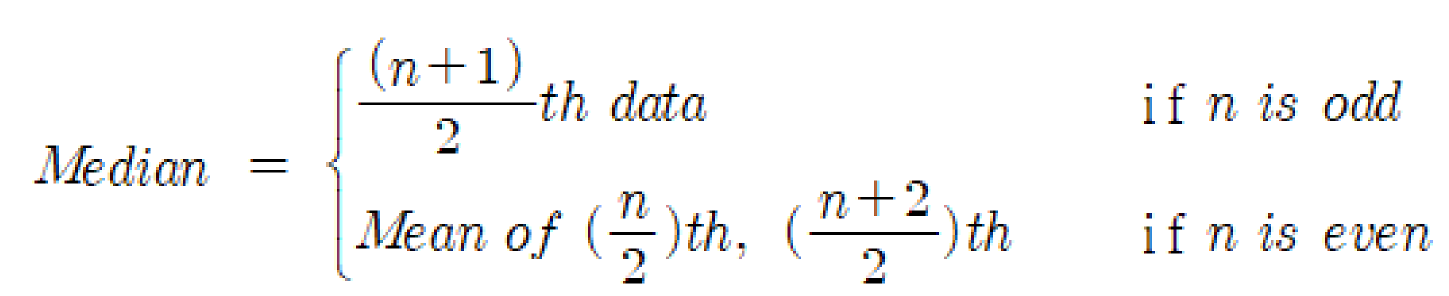
- Measure of the central location of data includes average, median, and mode
- The most commonly used is the mean (also called average).

$$Mean = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

– Mean of a population data is referred to as a population mean, μ
– Mean of a sample datat is referred as a sample mean $\bar{x}$
– Mean is influenced by extreme points whose value is very large or small.
– However, sample mean has a good characteristic to estimate population mean.

4.3 Summary Measures for Quantitative Variable

- Median is the value placed centrally when data is listed in order of size
  - Sample median m, population median M
  - If the number of data n id odd, median is the value of (n+1)/2 th
  - if the number of data n is even, median is the mean of n/2 th and (n/2+1) th



$$Median = \begin{cases} \frac{(n+1)}{2}th\ data & if\ n\ is\ odd \\ Mean\ of\ (\frac{n}{2})th,\ (\frac{n+2}{2})th & if\ n\ is\ even \end{cases}$$

- The median value is not sensitive even if there is an extreme point.
- It is used more often as a measure of the central location than the average if there is an extreme point.

4.3 Summary Measures for Quantitative Variable

- Mode is the most frequently occurred value among data.
- If data is continuous, it is unreasonable to simply set a mode value as the most frequently occurred value because most of the continuous data occurred only once or twice.
- In continuous data cases, we divide data into several intervals and find frequencies for each interval, and then the middle value of the interval which has the highest frequency is set to the mode.

4.3 Summary Measures for Quantitative Variable

[Ex 4.3.1] (Mean and Median)

Quiz scores of seven students in the data science class are as follows;

5, 6, 3, 7, 9, 4, 8

Find the mean and median of this sample by using 『eStat』 and compare them.

〈Answer〉

- The sample mean is as follows:

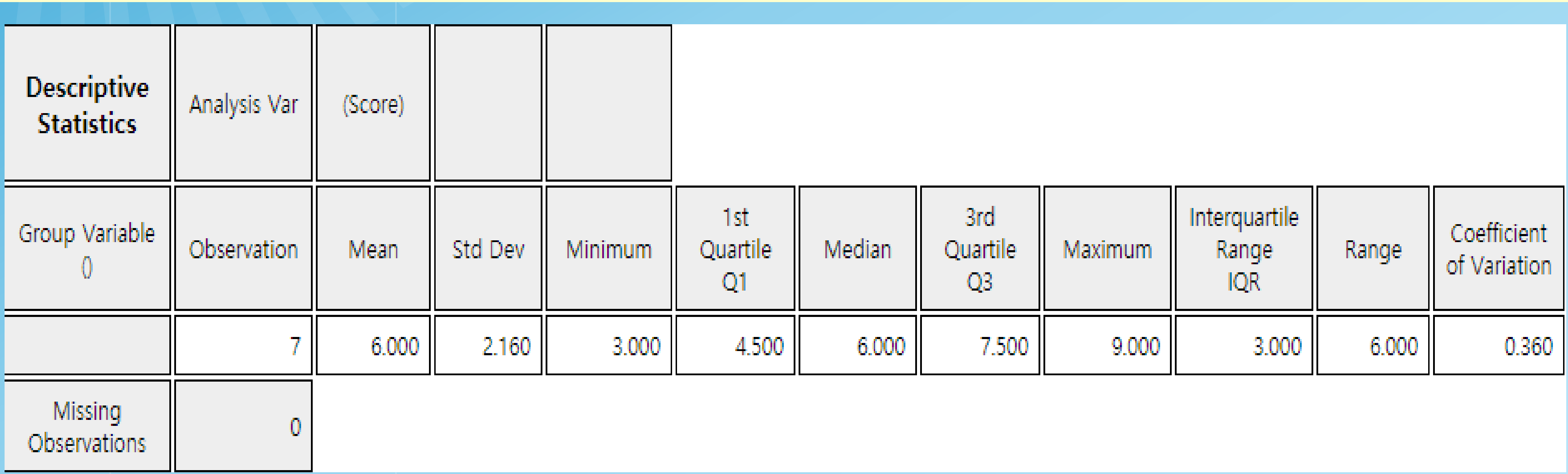
$$\bar{x} = (5 + 6 + 3 + 7 + 9 + 4 + 8) \,/\, 7 = 6$$

- In order to find the sample median, first arrange data in ascending order as
  3, 4, 5, 6, 7, 8, 9
- Since the sample size is an odd number, median is (n+1)/2th data which is (7+1)/2th that is m = 6,

4.3 Summary Measures for Quantitative Variable

〈Answer of Ex 4.3.1〉

- To obtain the mean and median values using『eStat』, enter the data in column V1 of the sheet and click the Descriptive Statistics icon .
- This will result in the log window as follows. It shows not only mean and median, but also other statistics such as the standard deviation, minimum, and maximum etc.



Descriptive Statistics	Analysis Var	(Score)									
Group Variable ()	Observation	Mean	Std Dev	Minimum	1st Quartile Q1	Median	3rd Quartile Q3	Maximum	Interquartile Range IQR	Range	Coefficient of Variation
	7	6.000	2.160	3.000	4.500	6.000	7.500	9.000	3.000	6.000	0.360
Missing Observations	0										

4.3 Summary of Data by Using Measures

[Ex 4.3.2] (Mode)
If the frequency table of a library visitor's age is as shown in Table 4.3.1.
Find the mode of the age by using the table.

Age Interval	Frequency (%)
[20.00, 30.00)	2 (6.7%)
[30.00, 40.00)	7 (23.3%)
[40.00, 50.00)	7 (23.3%)
[50.00, 60.00)	9 (30.0%)
[60.00, 70.00)	3 (10.0%)
[70.00, 80.00)	2 (6.7%)
Total	30 (100%)

〈Answer〉

- The interval [50.00, 60.00) has the highest frequency which is 9 and median is the mid value of the interval [50.00, 60.00) is 55.

4.3 Summary Measures for Quantitative Variable

- There are several variants to compensate the disadvantage of simple mean, one of which is the trimmed mean.
- This is to list the data in order and then average the data except for a constant number of large and small values respectively in order to eliminate the extremes.
- The trimmed mean is often used to prevent biased judging by referees in sports such as gymnastics and figure skating at the Olympics.
- In extreme cases, you may remove the top few percent data instead of the maximum and the bottom few percent data instead of the minimum.

4.3 Summary Measures for Quantitative Variable

- Another variant is a weighted mean in which each measurement is multiplied by a constant weight to obtain the mean.

$$Weighted\ \ Mean = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^{n} w_ix_i}{\sum_{i=1}^{n} w_i}$$

- The grade point average for college students which uses the weight of credit hours is an example of the weighted mean.
- The price index which uses the weights of the total amount of sales of the goods is another example of the weighted mean.

4.3 Summary Measures for Quantitative Variable

[Ex 4.3.3] An Olympic Gymnastics Game was judged by eight referees and their results are as follows. 9.0 9.5 9.3 7.2 10.0 9.1 9.4 9.0

Obtain the mean and median of this data. Also, find the trimmed mean which exclude the maximum and maximum and compare them.

〈Answer〉

- This data is not a sample but a population of eight. The mean is as follows.

  $\mu = (9.0 + 9.5 + 9.3 + 7.2 + 10.0 + 9.1 + 9.4 + 9.0) / 8 = 72.5/ 8 = 9.063$
- To find the median, arrange the data in ascending order.

  7.2 9.0 9.0 9.1 9.3 9.4 9.5 10.0
- Since n=8 is an even number, median is the average of $(n/2) = 8/2 =$ 4th data (=9.1) and $(n/2 + 1) = (8/2 + 1) =$ 5th data (=9.3). $M = (9.1 + 9.3)/2 = 9.2$.
- The trimmed mean is the average of the remaining numbers except the maximum of 7.2 and the maximum value of 10.0.

  Trimmed mean $= (9.0 + 9.0 + 9.1 + 9.3 + 9.4 + 9.5) / 6 = = 55.3/6 = 9.217$
- median or trimmed mean is more representative of the data than the mean.

4.3 Summary Measures for Quantitative Variable

[Ex 4.3.4]
A student took three courses in Korean (two credits), Math (four credits), and English (3 credits) this semester, and got A in Korean, B in math and C in English. Obtain the mean, and weighted mean if A is rated 4 points, B is 3 points, and C is 2 points 2.

〈Answer〉

- Mean = (4 + 3 + 2) / 3 = 3

- $$\text{Weighted Mean} = \frac{2\times 4+4\times 3+3\times 2}{2+4+3}=\frac{8+12+6}{9}=2.89$$

- Weighted mean is less than arithmetic mean because, although the Korean language (two credits) score A was good, it was relatively poor grade B in English (three credits).

4.3 Summary Measures for Quantitative Variable

4.3.2 Measure of Dispersion

- Measuring the degree of data dispersion in numerical values is called a measure of dispersion. The measure of dispersion commonly used is the variance or standard deviation, and other measures include mean absolute deviation, range, and inter-quartile range.
- Variance is the sum of the squared distances from data to the mean, and then divided by the number of data.

  − If the data are spread widely around the mean, the variance will increase

  − If the data is concentrated around the mean, the variance will be small

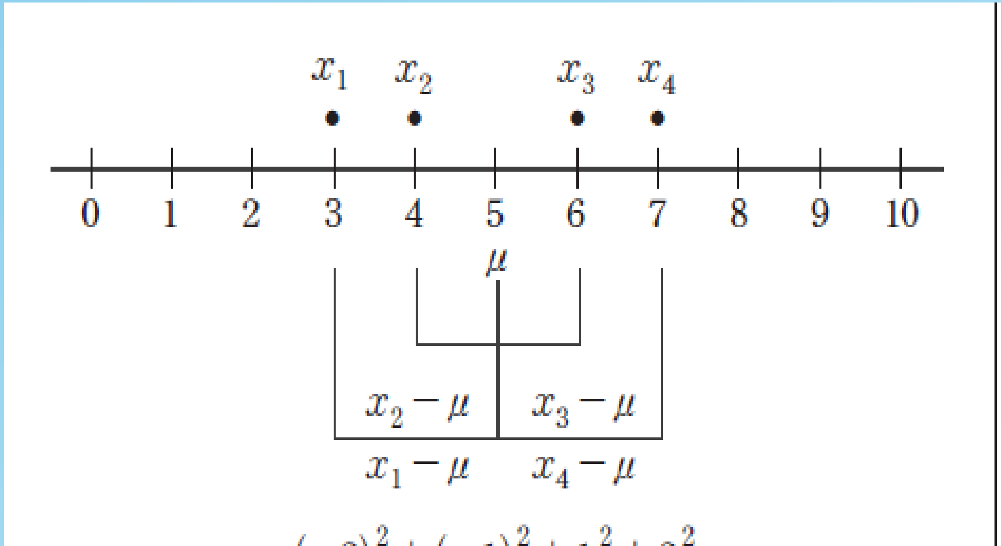
4.3 Summary Measures for Quantitative Variable

Population variance $\sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$ (N: number of population data)

Sample variance $S^2 = \dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$ (n : number of sample data)

✓ There are important reasons for using n-1 instead n when calculating the sample variance (Refer Chapter 6)

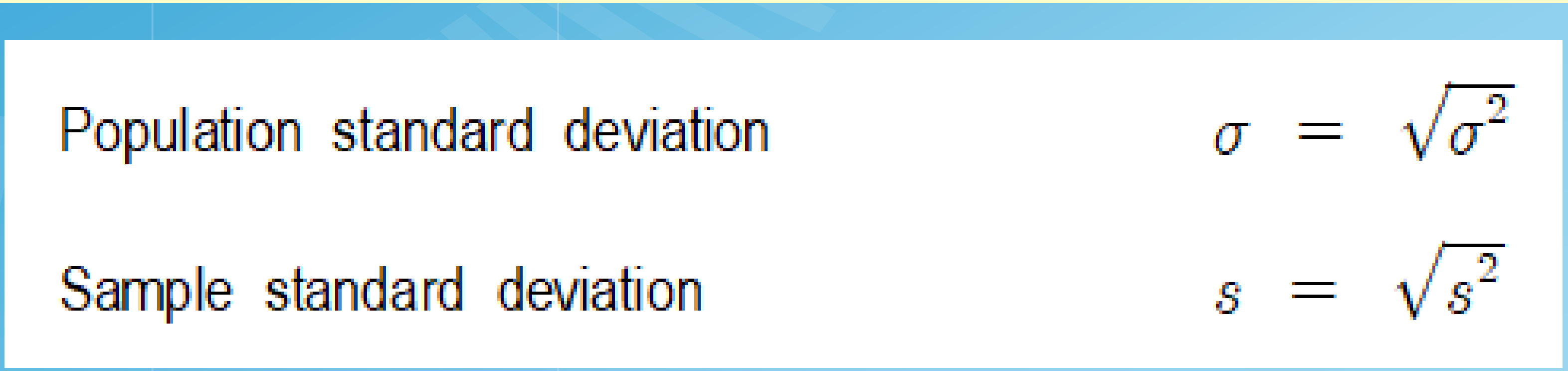
- The variance is the mean of the sum of squared distances from the mean to each measured value.



$$\sigma^2 = \frac{(-2)^2+(-1)^2+1^2+2^2}{4} = 2{,}5$$

4.3 Summary Measures for Quantitative Variable

- Standard deviation is defined as the square root of the variance.
  - The standard deviation of the population is denoted as σ.
  - The standard deviation of the sample is denoted as s.



Population standard deviation $\sigma = \sqrt{\sigma^2}$

Sample standard deviation $s = \sqrt{s^2}$

(

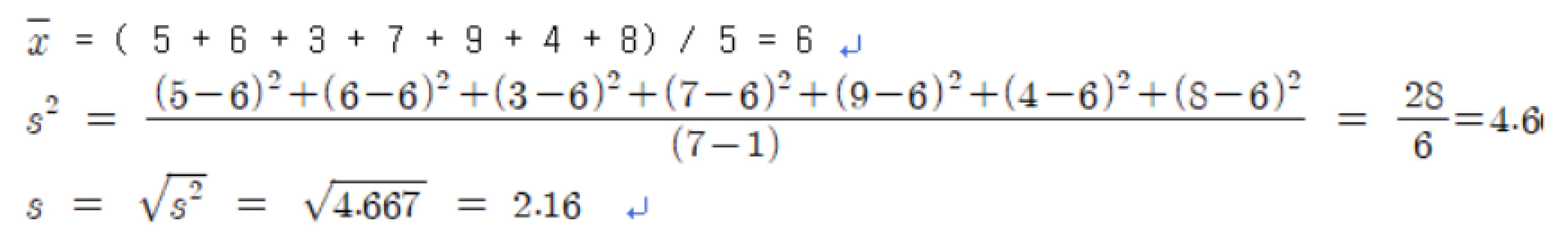
- Variance is not easy to interpret because it is the mean of the squared distance.
- Standard deviation is the square root of the variance, which allows it to be interpreted as a measure of the mean distance from each value to the mean.

4.3 Summary Measures for Quantitative Variable

[Example 4.3.5] Calculate mean and standard deviation from sample data
5, 6, 3, 7, 9, 4, 8.

〈Answer〉

- Note that this data is sample.



$$\bar{x} = (5 + 6 + 3 + 7 + 9 + 4 + 8) / 5 = 6$$

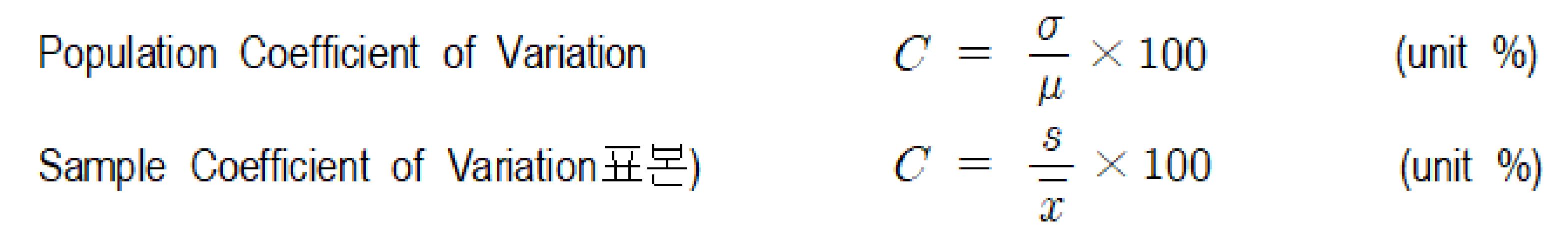
$$s^2 = \frac{(5-6)^2+(6-6)^2+(3-6)^2+(7-6)^2+(9-6)^2+(4-6)^2+(8-6)^2}{(7-1)} = \frac{28}{6} = 4.6$$

$$s = \sqrt{s^2} = \sqrt{4.667} = 2.16$$

- 『eStat』 calculates the sample standard deviation.

4.3 Summary Measures for Quantitative Variable

- If the units of measurements for more than one data are different from each other, comparing standard deviations is meaningless.
- Coefficient of variation which is the division of the standard deviation by its mean is used to compare several sets of data.



Population Coefficient of Variation $C = \frac{\sigma}{\mu} \times 100$ (unit %)

Sample Coefficient of Variation표본) $C = \frac{s}{\overline{x}} \times 100$ (unit %)

[Ex 4.3.6] The average weekly sales of a company was 1.36 billion dollar and the standard deviation was 0.28 billion dollar. When the same data was made in monthly sales, the average was 5.4.4 billion dollar and the standard deviation was 0.5 billion dollar. Calculate and compare the coefficient of variation .

〈Answer〉

- The coefficient of variation in weekly sales is (0.28 / 1.36) × 100 = 20.6%.
- The coefficient of variable in monthly sales is (0.50 / 5.44) × 100 = 9.2%.
- The change in monthly sales is smaller than the change in weekly sales.

4.3 Summary Measures for Quantitative Variable

- Range indicates the difference from the maximum value of the data minus the maximum value. The range is easy to calculate, but is not a good measure of the dispersion if there are extreme points.

  Range = Maximum − Maximum

- Inter-quartile range is a measure to complement the disadvantage of the range. The p percentile means roughly the pth percent data when data is arranged in order from small to large.

  p percentile = there are p% of observations less than(≤) this value

  ((100−p)% of observations above(≥) this value .

- The 25 percentile of the data is called the 1st quartile (Q1), the 50 percentile is called the 2nd quartile(Q2) or the median, and the 75 percentile is called the 3rd quartile (Q3). The inter-quartile range (IQR) is the 3rd quartile minus the 1st quartile.

  inter-quartile range (IQR) = Q3 − Q1

4.3 Summary Measures for Quantitative Variable

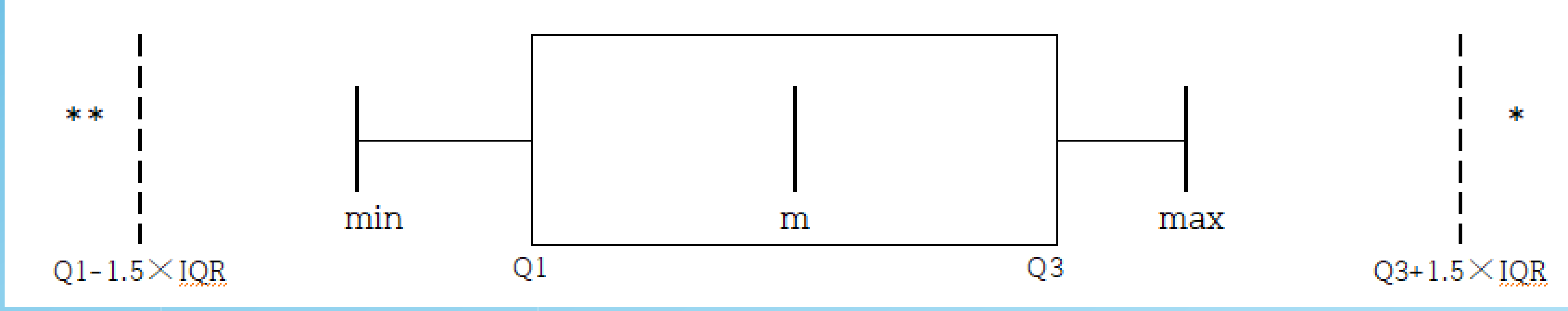
[Ex 4.3.7] For data 5, 6, 3, 7, 9, find the range and inter−quartile range.

⟨Answer⟩

- Range = maxi(9) - min(3) = 6.
- Arrange data in ascending order.
  (3, 5, 6, 7, 9)
- Median is (5+1)/2 th data which is 6.
- Divide sorted data into two parts
  (3,5,6) (6,7,9) ⟨= note that median 6 included in both parts
- Median of the (3,5,6) which is 5 is the 25 percentile (Q1).
- Median of the (6,7,9) which is 7 is the 75 percentile (Q3).
- IQR is Q3 − Q1 = 7 − 5 = 2.

4.3 Summary Measures for Quantitative Variable

- Box−whiskers plot is a method to show the quartiles of data.
- The box−whiskers plot mark Q1 and Q3 at a horizontal line and connects with a square box. Then displays the median (Q2) at the location proportional to Q1 and Q3 in the box. Connect a line with the minimum value which is greater than (minimum − 1.5*IQR) and the box. Similarly, connect a line with the maximum value which is less than (maximum + 1.5*IQR) and the box.
- Using the box graph, you can see the symmetry of the data distribution, the central location of the data, and the degree of dispersion.
- Data that crosses the line between (minimum − 1.5*IQR) and (maximum + 1.5*IQR)are sometimes considered extreme.

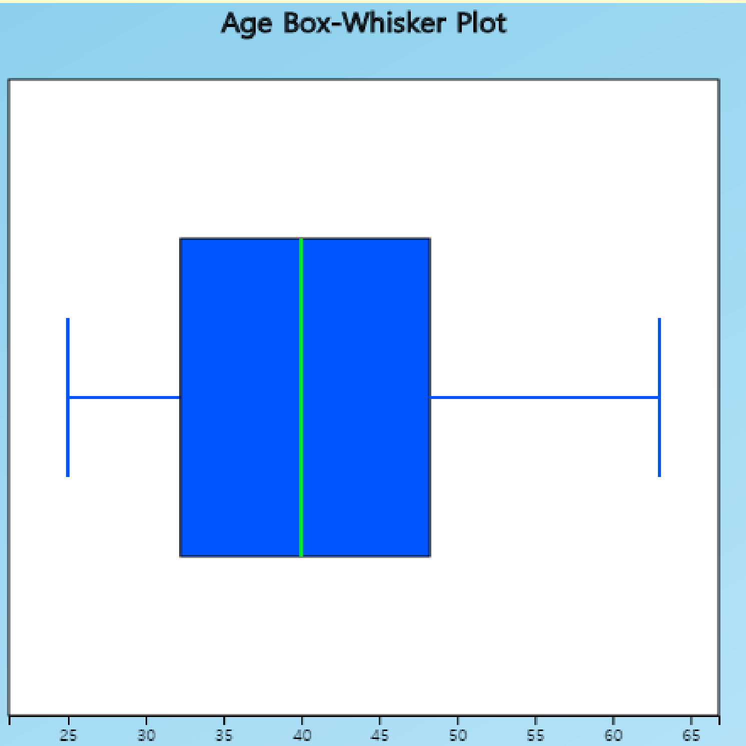


4.3 Summary Measures for Quantitative Variable

[Ex 4.3.8]] The data 032Continous_TeacherAgeByGender.csv in 『eStat』 is a teacher's age in a middle school.
1) Draw a box plot of age and examine median, range, quartiles and IQR.
2) Draw a box plot of age by gender and compare median, range, quartiles and IQR.

〈Answer〉

- After loading the data in 『eStat』, enter the value label of 'Gender' as 'Male' for 1 and 'Female' for 2 at [EditVar] button.
- Clicking on the box graph icon and then the 'age' variable
- Based on the median, we can see that the upper value is more scattered.

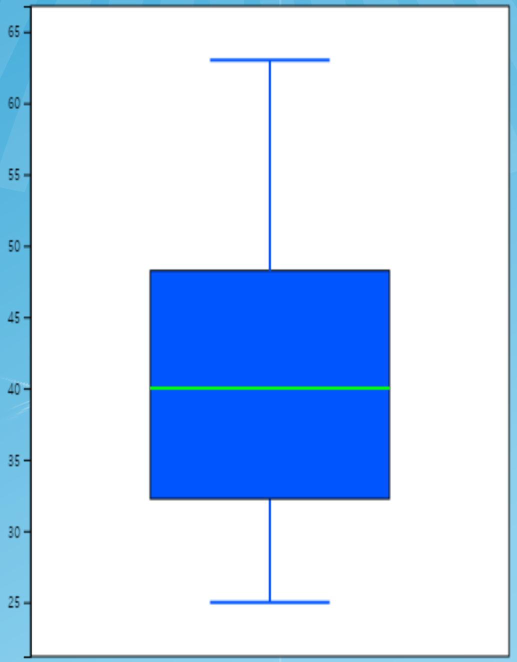


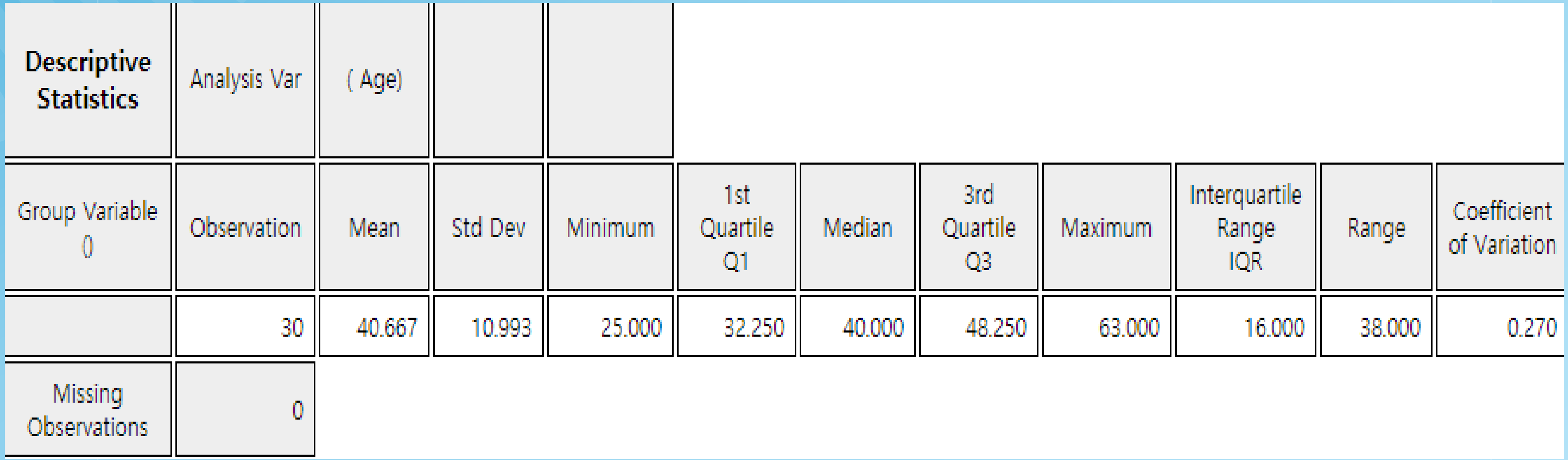
4.3 Summary Measures for Quantitative Variable

〈Answer of Ex 4.3.8〉

- Click the [Descriptive Statistics] button in the graph options to display the basic statistics of the ages
- Select 'Vertical' from the options below the graph for a vertical box graph

Age Box-Whisker Plot





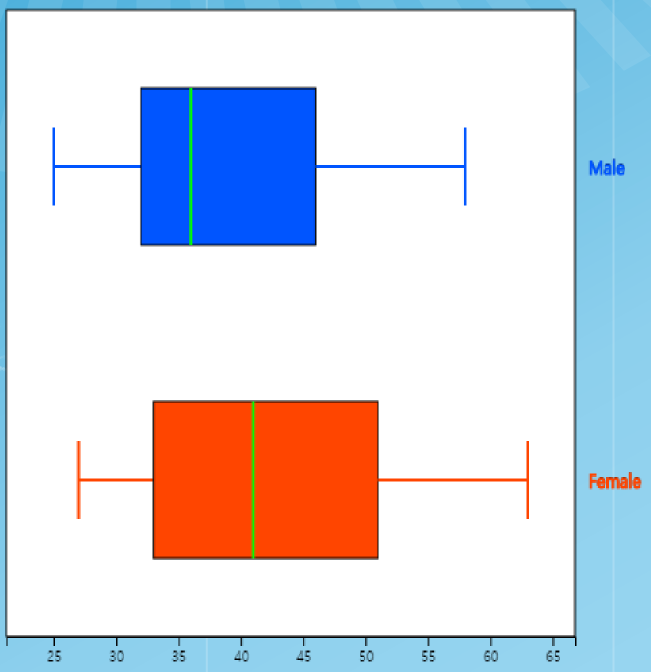
Descriptive Statistics	Analysis Var	(Age)									
Group Variable ()	Observation	Mean	Std Dev	Minimum	1st Quartile Q1	Median	3rd Quartile Q3	Maximum	Interquartile Range IQR	Range	Coefficient of Variation
	30	40.667	10.993	25.000	32.250	40.000	48.250	63.000	16.000	38.000	0.270
Missing Observations	0										

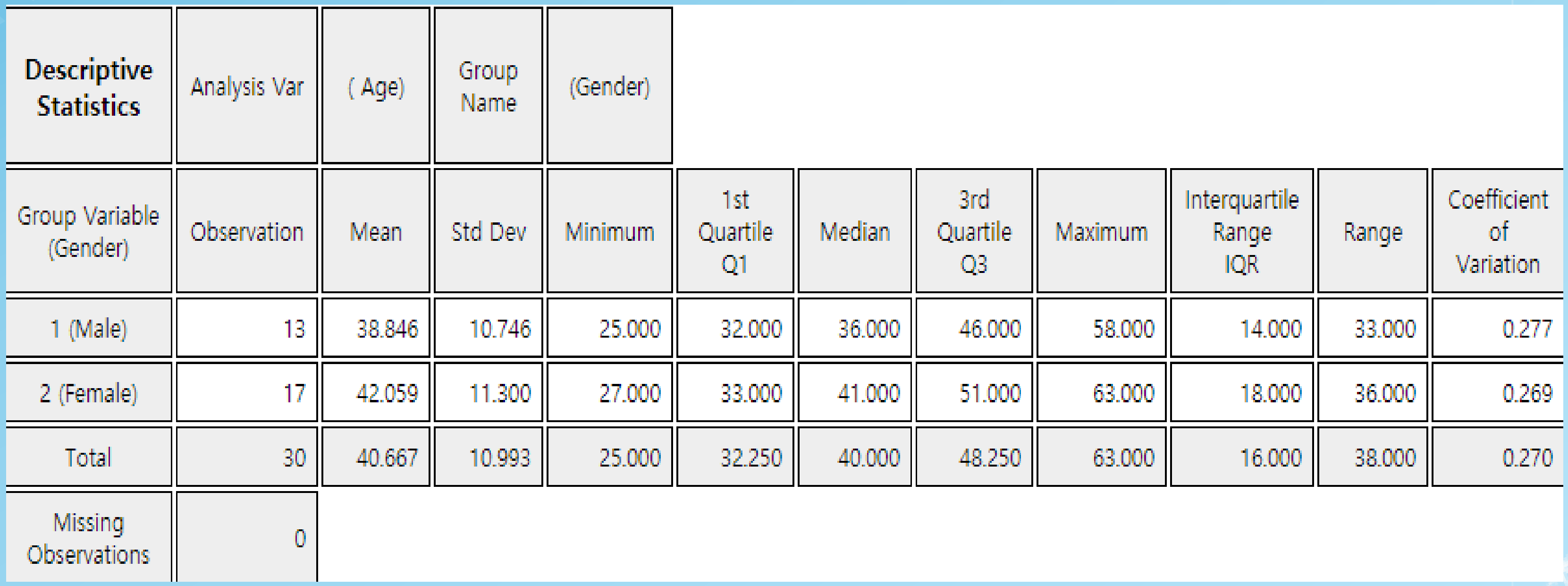
4.3 Summary Measures for Quantitative Variable

⟨Answer of Ex 4.3.8⟩

- Click on a 'gender' variable with the 'age' variable selected, a horizontal box plot by gender appears.
- The dispersion of female teachers' ages is greater than that of male teachers.

(Group Gender) Age Box-Whisker Plot





Descriptive Statistics	Analysis Var	(Age)	Group Name	(Gender)							
Group Variable (Gender)	Observation	Mean	Std Dev	Minimum	1st Quartile Q1	Median	3rd Quartile Q3	Maximum	Interquartile Range IQR	Range	Coefficient of Variation
1 (Male)	13	38.846	10.746	25.000	32.000	36.000	46.000	58.000	14.000	33.000	0.277
2 (Female)	17	42.059	11.300	27.000	33.000	41.000	51.000	63.000	18.000	36.000	0.269
Total	30	40.667	10.993	25.000	32.250	40.000	48.250	63.000	16.000	38.000	0.270
Missing Observations	0										

4.4 Summary

- Frequency table for categorical and continuous data
- Contingency table for categorical and continuous data
- Summary measure of continuous variable
  - Measure of central tendency : Mean, Median, Mode
  - Measure of dispersion: variance, standard deviation range, inter-quartile range, coefficient variation
  - box plot

Thank you