Introduction to Statistics and Data Science using *eStat*

Chapter 4 Data Summary Using Tables and Measures

4.3 Summary Measures for Quantitative Variable - Measure of Central Tendency -

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

4.3 Summary Measures for Quantitative Variable

4.3.1 Measure of the central location

- Measure of the central location of data : mean, median, and mode

- The most commonly used is the mean (also called average).

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

  - Mean of a population data is referred to as a population mean, μ
  - Mean of a sample data is referred as a sample mean, $\bar{x}$

- Mean is influenced by extreme points whose value is very large or small.
- 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 $(\frac{n+1}{2})^{th}$ data
  - if the number of data n is even, median is the mean of $(\frac{n}{2})^{th}$ and $(\frac{n+2}{2})^{th}$ data

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

- The median value is not sensitive even if there is an extreme point.
  => good measure of the central location 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 find a mode value.
  => divide data into several intervals and find frequencies for each interval,
  => middle value of the interval which has the highest frequency is 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:**

$$\bar{x} = \frac{5 + 6 + 3 + 7 + 9 + 4 + 8}{7} = 6$$

- **In order to find the sample median, first arrange data in ascending order**

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

- **Since the sample size is an odd number, median is $(\frac{n+1}{2})^{th}$ data which is $(\frac{7+1}{2})^{th}$ 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

- **Trimmed mean** Is to compensate the disadvantage of simple mean.
  => list data in order
  => remove certain portion of large and small values to eliminate extremes.
  => take an average of the remaining data
- It is often used to prevent biased judging by referees in sports such as gymnastics and figure skating at the Olympics.

4.3 Summary Measures for Quantitative Variable

- **Weighted 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 scores are as follows.

9.0 9.5 9.3 7.2 10.0 9.1 9.4 9.0

Find mean, median, trimmed mean which exclude the maximum and minimum.

<Answer>

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

  μ = (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 $(\frac{n}{2})^{th}$ = $(\frac{8}{2})^{th}$ = 4th data (=9.1) and $(\frac{n+2}{2})^{th}$ = $(\frac{8+2}{2})^{th}$= 5th data (=9.3). M = (9.1 + 9.3)/2 = 9.2.

- **Trimmed mean is the average of the remaining numbers except the minimum 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, grades he received as follows.

Korean A
Math B
English C

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
- $Weighted\ Mean = \frac{2\times4+4\times3+3\times2}{2+4+3} = \frac{8+12+6}{9} = 2.89$
- 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).

Thank you