Muhammad Farras Arira (23623006)
MS 5100: Research Methodology
November 1, 2023

Homework

The source code used in this homework can be found here :
https://github.com/farrasarira/Homework-Statistics-for-Research

1. Description of body height data of all participants in the Research Methodology course

   The description of body height data, statistically, is summarized in the table 1 below. In order to know the central tendency of the data, we use mean and median. From the table we can see that the mean and median the value is not much different. Besides that, we can know that the shortest person in the class has 148 cm of body height and the tallest perosn in the class has 186 cm of body height. The standard deviation tells us that dispersion of the data around the mean, in this case the value is 7.22 cm. And lastly from the negative value of skewness, we can know that the body height of the student is leaning toward tall body height.

	Value [cm]
mean	169.01
median	170
minimum value	148
maximum value	186
standard deviation	7.22
skewness	-0.497

Table 1: Summary of body height data of all participants in the Research Methodology course

   Description of typing speed data of all participants in the Research Methodology course

   In the dataset, the typing speed of the student is quantified by words-per-minute (WPM). The summary of the typing speed of all student in the class is shown by the table 2 below. The mean and median are used to study the central tendency of the data. We can see that the mean and the median value is quite different. This different between mean and median, ussually occurs due to outlier data. From the summary, we can also know that the slowest typing speed is 25 WPM and the fastest typing speed is 273 WPM. The standard deviation of this data is 28.49. If we compare to the body height's standard deviation, the typing speed data will have spreader data. Lastly, from the skewness value, we can learn that the typing speed data is leaning towards low WPM. This statements will be proved by the visualization of the data distribution below.

	Value [WPM]
mean	51.78
median	45
minimum value	25
maximum value	273
standard deviation	28.49
skewness	5.10

Table 2: Summary of type speed data of all participants in the Research Methodology course

2. Visualization of the distribution of body height data

   The distribution of body heihgt data is visualized by histogram as depicted on figure 1. From this histogram, we can see that the body height data is tend towards the higher value. This proves our skewness calculation above, where if the data is tending toward higher data the sknewness should be negative value.

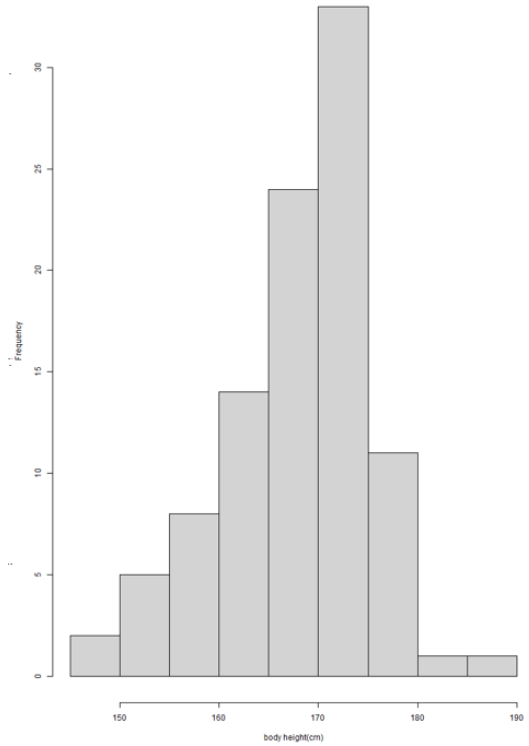

Figure 1: Histogram of the body height data

Visualization of the distribution of typing speed data

The distribution of the typing speed data is visualized by the histogram in figure 2. There are several things that we can know from the distribution. First, the histogram shows that the data has tendency towards lower value, which means the skewness of the data should be positive value. It proves our calculation on the previous part. The second thing that we can know from the histogram is there is an outlier in the data. The outlier is located in the highest value of the data.

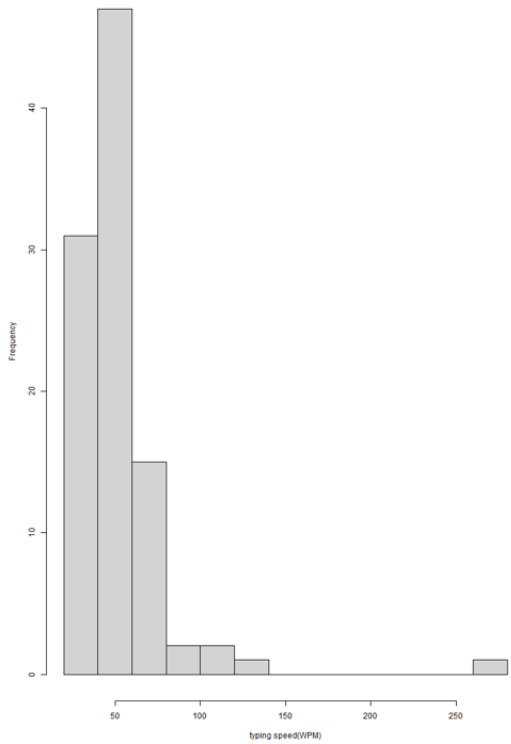

Figure 2: Histogram of the typing speed data

3. Quartiles and Quantiles of Data

The quartile and 0.1, 0.2, 0.8, 0.9 quantiles of the body height data is shown by the table 3. From the quartile data, we can calculate interquartile quartile range (IQR) by using $IQR = Q_3 - Q_1$, which is 8 cm. This IQR value can be used to describe the variability of the data.

	Value [cm]
Q_1	165
Q_2	170
Q_3	173
0.1-quantile	160
0.2-quantile	165
0.8-quantile	175
0.9-quantile	178

Table 3: Quartiles and quantiles of body height data

The quartile and 0.1, 0.2, 0.8, 0.9 quantiles of the typing speed data is shown by the table 4. The IQR of the typing speed data is 18 WPM, which is low value. However if we see the 0.9-quantile value, which is 72.4, we can know that the variablity of the data is quite high. This phenomena is due to the outlier data as already explained in the previous section.

	Value [WPM]
Q_1	38.5
Q_2	45.0
Q_3	56.5
0.1-quantile	32.0
0.2-quantile	36.0
0.8-quantile	61.0
0.9-quantile	72.4

Table 4: Quartiles and quantiles of typing speed data

4. Comparing the distribution of the body height data with the average body height of the Indonesian Population

The comparison of the distribution of the body height data with average body height of Indonesian is done by using the following hypotheses.

$$\mathcal{H}_0 : \mu = \mu_0$$
$$\mathcal{H}_1 : \mu \neq \mu_0$$

where μ is average of the student's body height and μ_0 is the average of the Indonesian population body height. Based on the data obtained from the reference, the average body height of Indonesian population is 158 cm. In other word, in this hypothesis test, we want to whather there is a significant different between average body height of the student with the Indonesian population. In order to test the hypothesis, we utilize the t-test. Based on the calculation using t-test, the p-value is lower that 2.2×10^{-16}. This p-value is lower than our significance level of 0.05. Hence, we reject the $\mathcal{H}_0$. This means that there is significant different between average of body height of the student with Indonesian population.

Comparing the distribution of the body height data with the average body height of the Global Population

The comparison of the distribution of the body height data with average body height of global population is done by using the following hypotheses.

$$\mathcal{H}_0 : \mu = \mu_0$$
$$\mathcal{H}_1 : \mu \neq \mu_0$$

where μ is average of the student's body height and μ_0 is the average of the global population body height. Based on the data obtained from the reference, the average body height of global population is 168.5 cm. In order to test the hypothesis, we utilize the t-test. Based on the calculation using t-test, the p-value is 0.4836. This p-value is higher than our significance level of 0.05. Hence, we fail to reject the $\mathcal{H}_0$ hypotheses. This means that there is no significant different between average of body height of the student with global population.

5. Comparing the distribution of the student's typing speed data with the mean typing speed of all people who took the test

   The comparison of the distribution of the student's typing speed data with the mean typing speed of all people who took the test is done by using the following hypotheses.

$$\mathcal{H}_0 : \mu = \mu_0$$
$$\mathcal{H}_1 : \mu \neq \mu_0$$

   where μ is average of the student's typing speed and μ_0 is the the mean typing speed of all people who took the test. Based on the data obtained from the typing test website, the mean of typing speed is 40.426 WPM. In order to test the hypothesis, we utilize the t-test. Based on the calculation using t-test, the p-value is 0.0001398. This p-value is lower than our significance level of 0.05. Hence, we reject the $\mathcal{H}_0$ hypotheses. This means that there is significant different between average of typing speed of the student with the mean typing speed of all people who took the test.

6. Correlation between typing speed and body height

   In order to know the correlation between the typing speed and body height, we utilize Pearson correlation, which measure linear relationship between the two datasets. Based on the Pearson correlation formula, for this case, we obtain the Pearson correlation coefficient 0.052. The Person correlation coefficient can tell us the "level" of relationship between two datasets. The Person correlation has value between -1 and 1, where -1 tells there is an exact negative linear relationship and 1 tells there is an eact positive linear relationship. If the Pearson correlation falls near 0, it means there is no correlation exist. In our case, since the Person correlation coefficient is 0.052, we can conclude that there no correlation between typing speed and body height.

7. Comparing the typing speed of regular student and the non-regular student

   The comparison of the typing speed of regular student and the non-regular student is done using the followin hypotheses.

$$\mathcal{H}_0 : \mu_1 = \mu_2$$
$$\mathcal{H}_1 : \mu_1 \neq \mu_2$$

   where μ_1 is average of the regular student's typing speed and μ_2 is the average of the non-regular student's typing speed. In order to test the hypothesis, we utilize the t-test. Based on the calculation using t-test, the p-value is 0.9222. This p-value is higher than our significance level of 0.05. Hence, we fail reject the $\mathcal{H}_0$ hypotheses. This means that there is no significant different between average of typing speed of the regular student with the non-regular student.