1. This question is solved in Excel. The details are given below.

   a. First of all, the number of intervals for WP is determined by using the below formulas:

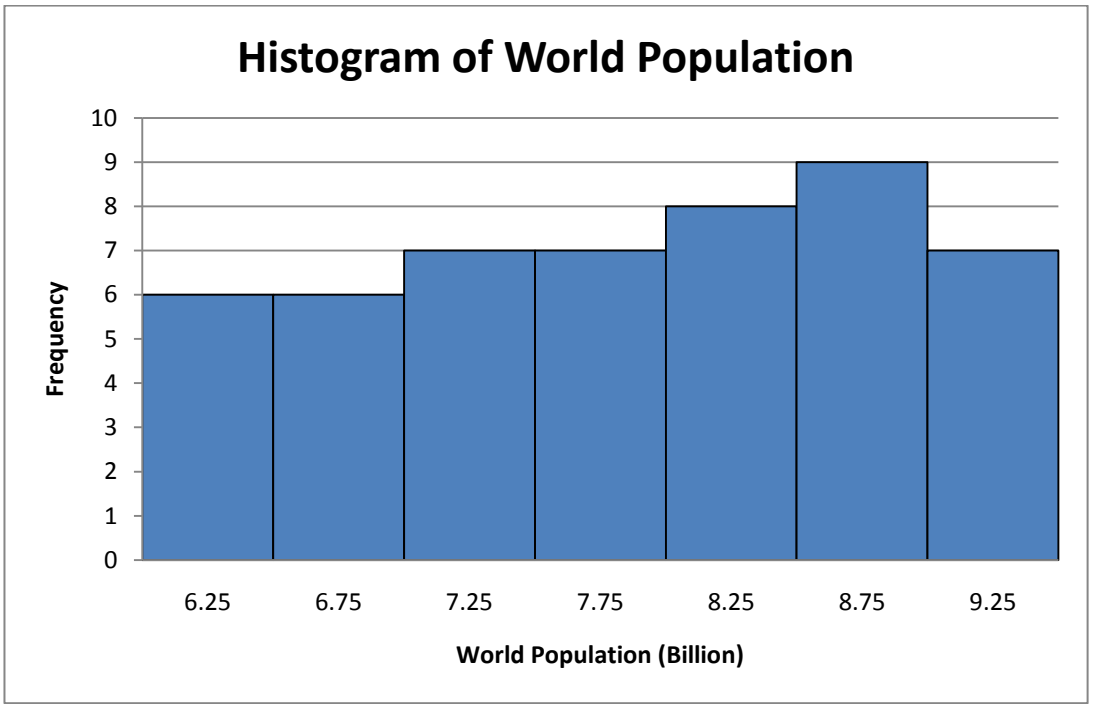
$$n_c = n^{1/2} \Rightarrow 7.071$$
$$n_c = 1 + 3.3log_{10}(n) \Rightarrow 6.607$$
$$n_c = \frac{r * n^{1/2}}{2 * iqr} \Rightarrow 3.521$$

where n=50, r=9.325-6.091=3.234 and iqr=8.707-7.015=1.692. (Note that the non – grouped ranges and the interquartile ranges are used for these formulas because the categorization is not done at this step.)

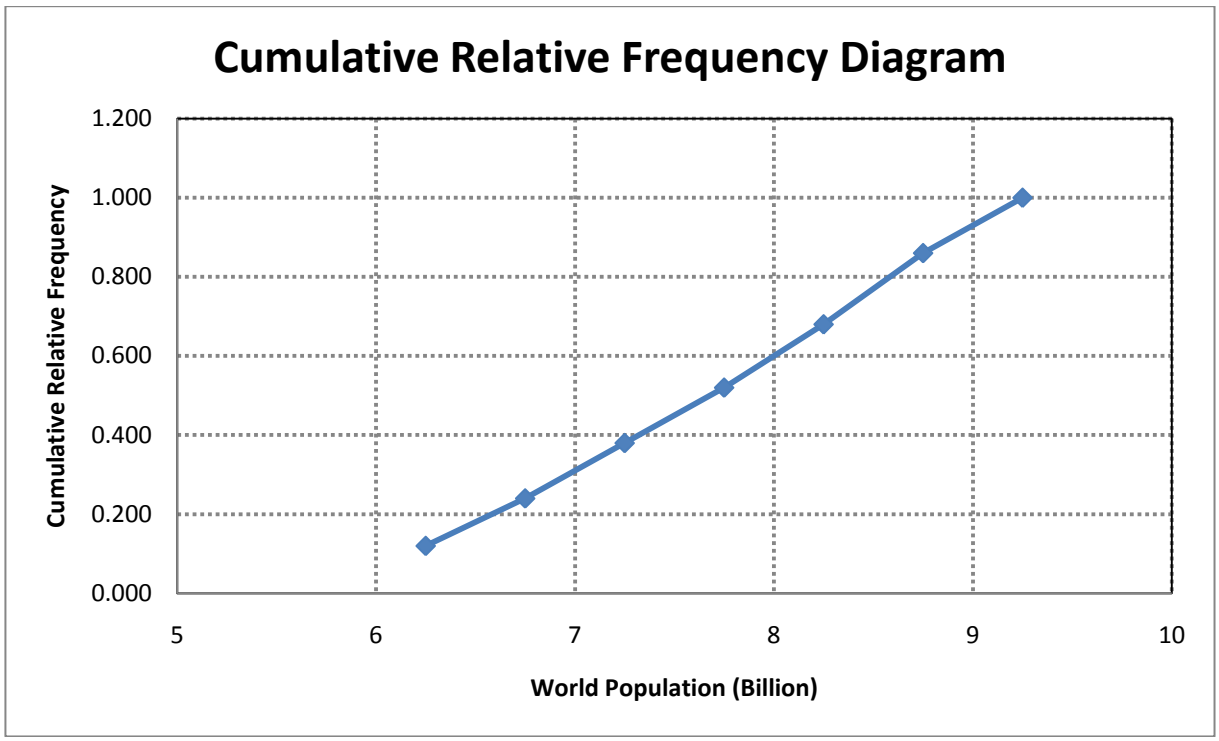
Therefore, it is a good idea to use seven intervals starting from 6 and having a size of 0.5. The other steps are done in Excel medium and given below table.

| Class Interval | Class Mark | Frequency | Total Frequency | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|---|---|---|
| 6-6.5 | 6.250 | 6 | 6 | 0.120 | 0.120 |
| 6.5-7 | 6.750 | 6 | 12 | 0.120 | 0.240 |
| 7-7.5 | 7.250 | 7 | 19 | 0.140 | 0.380 |
| 7.5-8 | 7.750 | 7 | 26 | 0.140 | 0.520 |
| 8-8.5 | 8.250 | 8 | 34 | 0.160 | 0.680 |
| 8.5-9 | 8.750 | 9 | 43 | 0.180 | 0.860 |
| 9-9.5 | 9.250 | 7 | 50 | 0.140 | 1.000 |

Note that in the table above, the intervals are taken as [a,b) in calculating the frequencies. After that, the histogram for the world population is drawn.




The cumulative frequency diagram for the world population is also drawn in Excel and shown below.




The number of intervals for WPC is determined by using the below formulas:

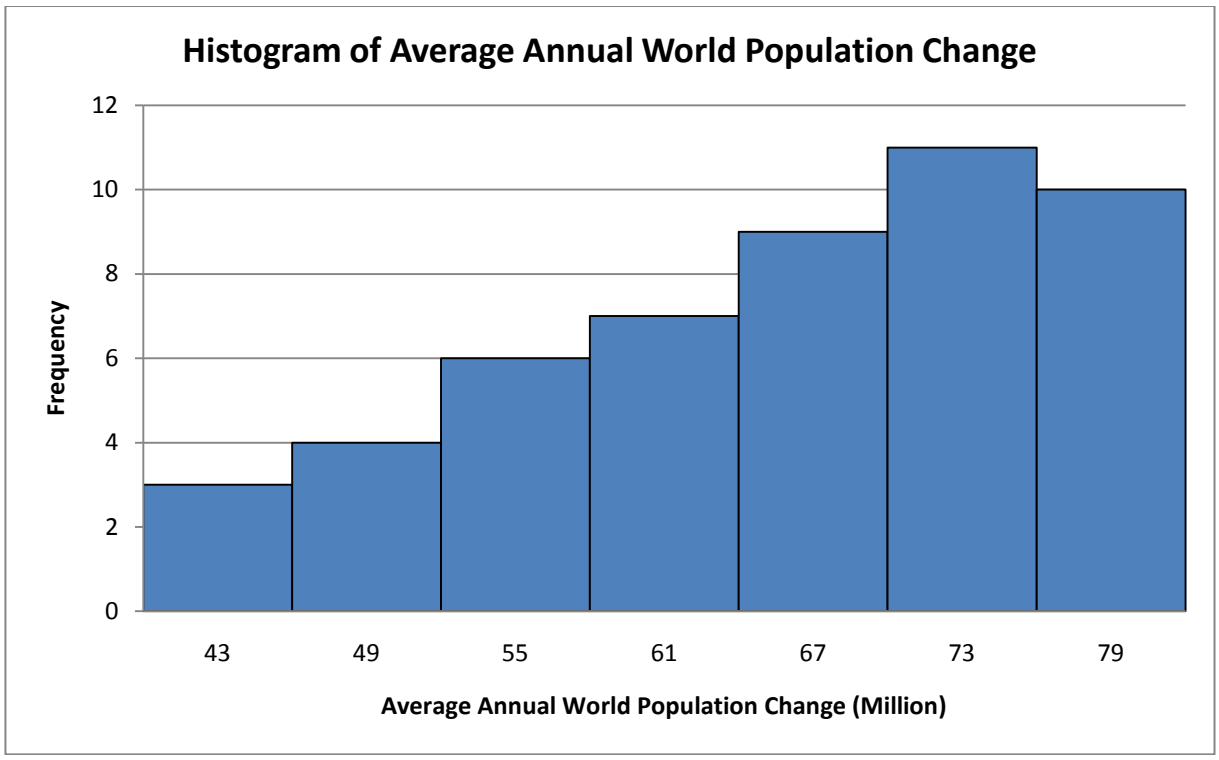
$$n_c = n^{1/2} \Rightarrow 7.071$$
$$n_c = 1 + 3.3log_{10}(n) \Rightarrow 6.607$$
$$n_c = \frac{r * n^{1/2}}{2 * iqr} \Rightarrow 3.733$$

where n=50, r=78.770-43.218=35.552 and iqr=75.461-57.921=17.54.

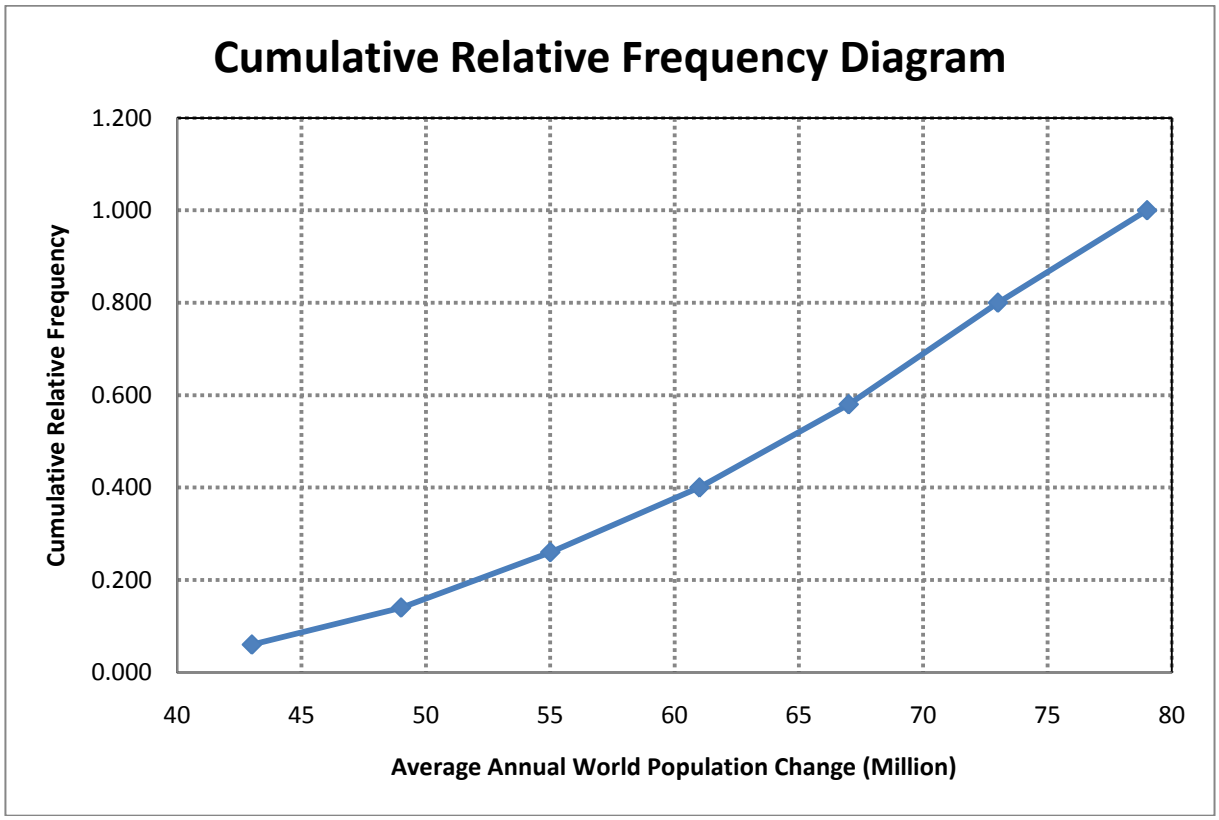
Therefore, it is a good idea to use seven intervals starting from 40 and having a size of 6. The other steps are done in Excel medium and given below table.

| Class Interval | Class Mark | Frequency | Total Frequency | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|---|---|---|
| 40-46 | 43.000 | 3 | 3 | 0.060 | 0.060 |
| 46-52 | 49.000 | 4 | 7 | 0.080 | 0.140 |
| 52-58 | 55.000 | 6 | 13 | 0.120 | 0.260 |
| 58-64 | 61.000 | 7 | 20 | 0.140 | 0.400 |
| 64-70 | 67.000 | 9 | 29 | 0.180 | 0.580 |
| 70-76 | 73.000 | 11 | 40 | 0.220 | 0.800 |
| 76-82 | 79.000 | 10 | 50 | 0.200 | 1.000 |

The histogram for the average annual world population change is shown below.




The cumulative frequency diagram for the average annual world population change is drawn in Excel and shown below.



b. The mean, median, mode, standard deviation and coefficient of variation for the world population (WP) are calculated in Excel and summarized below.

| Class Interval | Class Mark | Frequency | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 6-6.5 | 6.250 | 6 | 37.500 | 234.375 |
| 6.5-7 | 6.750 | 6 | 40.500 | 273.375 |
| 7-7.5 | 7.250 | 7 | 50.750 | 367.938 |
| 7.5-8 | 7.750 | 7 | 54.250 | 420.438 |
| 8-8.5 | 8.250 | 8 | 66.000 | 544.500 |
| 8.5-9 | 8.750 | 9 | 78.750 | 689.063 |
| 9-9.5 | 9.250 | 7 | 64.750 | 598.938 |
| | | | 392.500 | 3128.625 |

| | |
|---|---|
| ***Mean*** | 7.850 |
| ***Median*** | 7.500 |
| ***Mode*** | 8.750 |
| ***Standard Deviation*** | 0.985 |
| ***Coefficient of Variation*** | 0.125 |

The mean, median, mode, standard deviation and coefficient of variation for the average annual world population change (WPC) are also calculated in Excel and summarized below.

| Class Interval | Class Mark | Frequency | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|---|
| 40-46 | 43 | 3 | 129 | 5547 |
| 46-52 | 49 | 4 | 196 | 9604 |
| 52-58 | 55 | 6 | 330 | 18150 |
| 58-64 | 61 | 7 | 427 | 26047 |
| 64-70 | 67 | 9 | 603 | 40401 |
| 70-76 | 73 | 11 | 803 | 58619 |
| 76-82 | 79 | 10 | 790 | 62410 |
| | | | 3278 | 220778 |

| | |
|---|---|
| ***Mean*** | 65.560 |
| ***Median*** | 67.000 |
| ***Mode*** | 73.000 |
| ***Standard Deviation*** | 10.947 |
| ***Coefficient of Variation*** | 0.167 |

c. The histogram for both the world population and the average annual world population indicate that the data shows a little dispersion. This means that data are gathered around the mean. Therefore, the coefficients of variation are expected to be a small value, which is the case.

d. There are no outliers in the world population data set as the lower quartile value is 7.25 and the upper quartile is 8.75.

$$Q_1=7.25 \text{ and } Q_3=8.75 \Rightarrow iqr=Q_3-Q_1=1.5$$

The smallest data = 6.091 > $Q_1$-1.5iqr = 7.25-1.5*1.5 = 5
Check

The largest data = 9.325 < $Q_3$+1.5iqr = 8.75+1.5*1.5 = 11
Check

There are no outliers in the average annual world population change data set as the lower quartile value is 55 and the upper quartile is 73.

$$Q_1=55 \text{ and } Q_3=73 \Rightarrow iqr=Q_3-Q_1=18$$

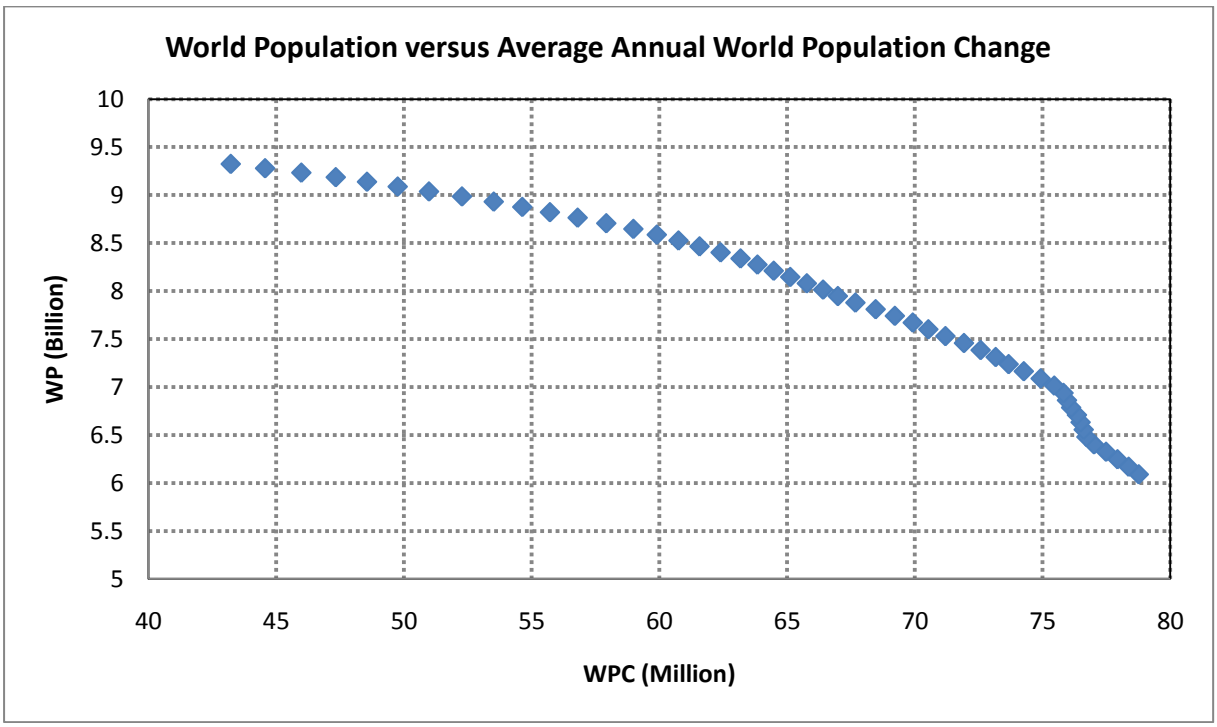
The smallest data = 43.218 > $Q_1$-1.5iqr =55-1.5*18 = 28
Check

The largest data = 78.770 < $Q_3$+1.5iqr =73+1.5*18 = 100
Check

a. The correlation coefficient is calculated by using the built-in functions of Excel. The scatter diagram for the world population and the average annual world population change is drawn in Excel and shown below.

| ***Correlation Coefficient ($r_{xy}$)*** | -0.961 |
|---|---|

The absolute value of the correlation coefficient is close to one, so there is a strong linear correlation between the world population data set and the average annual world population change data set. This inference can easily be proved by drawing the scatter diagram of these data sets.

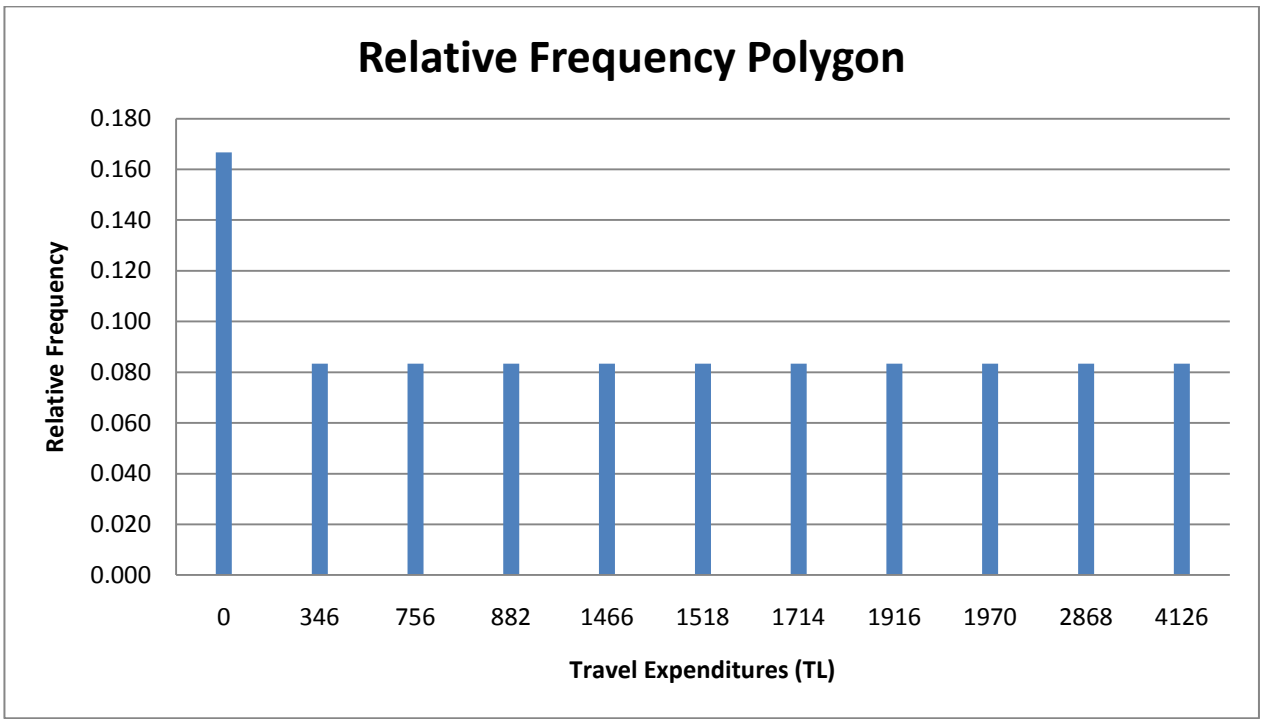


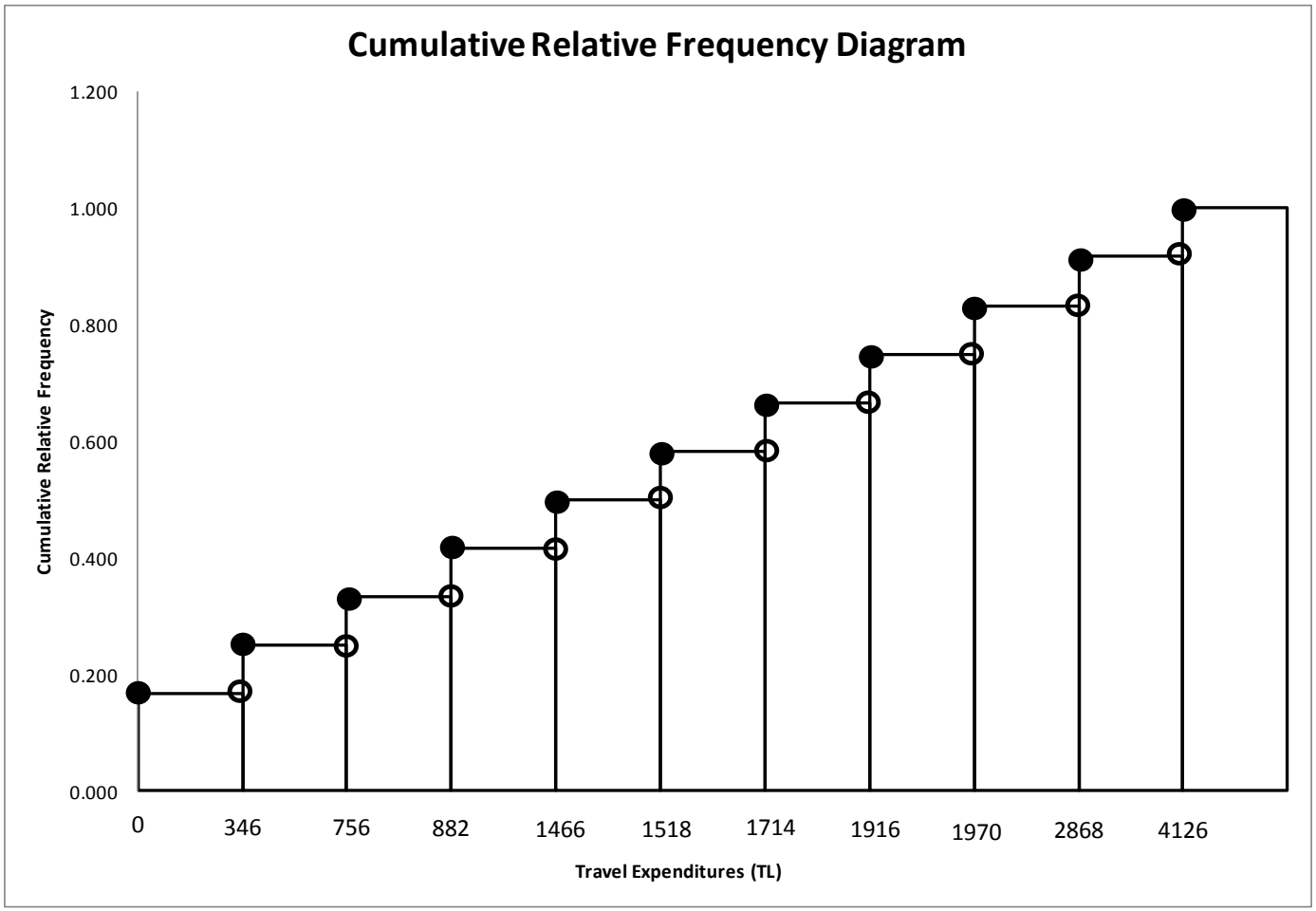

2. The data for the travel expenditures by the 12 members of a university's civil engineering department at a certain region is analyzed as follows:

a. The frequencies, total frequencies, relative frequencies and the cumulative relative frequencies are tabulated below.

| Data | Frequency | Total Frequency | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|---|---|
| 0 | 2 | 2 | 0.167 | 0.167 |
| 346 | 1 | 3 | 0.083 | 0.250 |
| 756 | 1 | 4 | 0.083 | 0.333 |
| 882 | 1 | 5 | 0.083 | 0.417 |
| 1466 | 1 | 6 | 0.083 | 0.500 |
| 1518 | 1 | 7 | 0.083 | 0.583 |
| 1714 | 1 | 8 | 0.083 | 0.667 |
| 1916 | 1 | 9 | 0.083 | 0.750 |
| 1970 | 1 | 10 | 0.083 | 0.833 |
| 2868 | 1 | 11 | 0.083 | 0.917 |
| 4126 | 1 | 12 | 0.083 | 1.000 |

The relative frequency polygon and the cumulative relative frequency polygon are drawn in Excel and shown in the following page.




b. The mean, median, mode, variance and standard deviation of the given data is found by using the below table and the formulas.

| Data | Frequency | $f_i x_i$ | $f_i x_i^2$ | $f_i*(x_i-x_{mean})^2$ |
|---|---|---|---|---|
| 0 | 2 | 0 | 0 | 114242 |
| 346 | 1 | 346 | 119716 | 11449 |
| 756 | 1 | 756 | 571536 | 267289 |
| 882 | 1 | 882 | 777924 | 413449 |
| 1466 | 1 | 1466 | 2149156 | 1505529 |
| 1518 | 1 | 1518 | 2304324 | 1635841 |
| 1714 | 1 | 1714 | 2937796 | 2175625 |
| 1916 | 1 | 1916 | 3671056 | 2812329 |
| 1970 | 1 | 1970 | 3880900 | 2996361 |
| 2868 | 1 | 2868 | 8225424 | 6911641 |
| 4126 | 1 | 4126 | 17023876 | 15108769 |
| | | 17562 | 41661708 | 33952524 |

$$\overline{x} = \frac{\sum x_i}{n} \qquad s^2 = \frac{\sum f_i (x_i - \overline{x})^2}{n-1}$$

$$s = \sqrt{s^2} \qquad v = \frac{s}{\overline{x}}$$

where $\overline{x}$ is the mean, $s^2$ is the variance, s is the standard deviation and v is the coefficient of variation.

| | |
|---|---|
| ***Mean*** | 1463.5 |
| ***Median*** | 1492 |
| ***Mode*** | 0 |
| ***Variance*** | 1450883.727 |
| ***Standard Deviation*** | 1204.526 |
| ***Coefficient of Variation*** | 0.823 |

It can be inferred that there is a high spread from the mean as the coefficient of variation is high.

c. There is only one outlier but no extreme outliers in the data set.

$Q_1$=551 and $Q_3$=1943 ⇒ iqr=$Q_3$-$Q_1$=1392

The smallest data = 0 ≥ $Q_1$-1.5iqr = 551-1.5*1392 = -1537
Check

The largest data = 4126 > $Q_3$+1.5iqr = 1943+1.5*1392 = 4031
< $Q_3$+3iqr = 1943+3*1392 = 6119
No Good

Therefore, 4126 is an outlier.

3. For the data set A, the range is

$Range_A$=Highest $Data_A$-Lowest $Data_A$=10-6=4

For the data set B, the range is

$Range_B$=Highest $Data_B$-Lowest $Data_B$=11-5=6

| Data A | Frequency | $f_iA_i$ | $f_iA_i^2$ | $f_i*(A_i-A_{mean})^2$ |
|---|---|---|---|---|
| 6 | 2 | 12 | 72 | 11.281 |
| 7 | 1 | 7 | 49 | 1.891 |
| 9 | 2 | 18 | 162 | 0.781 |
| 10 | 3 | 30 | 300 | 7.922 |
| | | 67 | 583 | 21.875 |

The standard deviation for the data set A is

$$n_A = 8$$

$$\bar{A} = \frac{\sum f_i A_i}{n_A} = \frac{67}{8} = 8.375$$

$$s_A = \sqrt{\frac{\sum f_i \left(A_i - \overline{A}\right)^2}{n_A - 1}} = \sqrt{\frac{21.875}{8-1}} = 1.768$$

| Data B | Frequency | $f_iB_i$ | $f_iB_i^2$ | $f_i*(B_i-B_{mean})^2$ |
|---|---|---|---|---|
| 5 | 1 | 5 | 25 | 11.391 |
| 8 | 3 | 24 | 192 | 0.422 |
| 9 | 3 | 27 | 243 | 1.172 |
| 11 | 1 | 11 | 121 | 6.891 |
| | | 67 | 581 | 19.875 |

The standard deviation for the data set B is

$$n_B = 8$$

$$\bar{B} = \frac{\sum f_i B_i}{n_B} = \frac{67}{8} = 8.375$$

$$s_B = \sqrt{\frac{\sum f_i \left(B_i - \overline{B}\right)^2}{n_B - 1}} = \sqrt{\frac{19.875}{8-1}} = 1.685$$

It can be inferred that the variability of the data set A is higher as far as the range is concerned. However, a totally different comment can be made by taking the standard deviation into consideration as data set A has a higher standard deviation. (This comment can be more meaningful if the coefficient of variation is used but in our case, the standard deviations can also be utilized as the mean of the two data sets are equal to each other.)

Although the range can be used to compare the variability of data sets, it is not a good indicator of the variability because having a single high value and a single low value means a high range but this does not guarantee that the data between these limiting values have also large spread. (Data set B)

The standard deviation also has some weaknesses as far as its indication of variability is concerned because it has a unit, which makes the classification of largeness very difficult. Therefore, the variability should be determined by referring to its proportion to the mean value, which helps to decide whether the value of the standard deviation is low. Therefore, the coefficient of variation is used for pointing out the variability.