

1. Consider the following frequency distribution of CGPA of 100 UIU students.

CGPA	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4
No of students	7	18	35	27	10	3

- Sketch histogram and frequency polygon of the frequency distribution in the same graph. Also, sketch the pie chart.
  - Sketch histogram and find the mode from it. Check your result.
  - Draw the cumulative frequency curve (OGIVE) and find median, 3<sup>rd</sup> quartile, 7<sup>th</sup> Decile, and 55<sup>th</sup> Percentile from the OGIVE. Also, verify by the analytical method.
  - Find arithmetic mean and standard deviation of the frequency distribution.
  - Find the statistical index for the value  $x = 2.6$ .
  - Estimate median and then mean deviation from median. Also, find the coefficient of the mean deviation.
  - Find first four raw moments about  $A = 1.75$  and convert them to the central moments.
2. Each member of an athletics club was asked to monitor the distance run in training during a particular week. The table below summarizes the results.

Distance to nearest Km	30-40	40-50	50-60	60-70	70-80	80-90
Number of athletes'	2	4	7	12	9	6

- Find the cumulative frequency polygon.
  - Estimate  $Q_1$ ,  $Me$ ,  $D_6$  &  $P_{85}$  from the cumulative frequency polygon.
  - Estimate the mean distance and the standard deviation of the distribution.
  - Find the interquartile range of the given distribution.
3. The frequency distribution table of CGPA of the some UIU students is given below.

CGPA	1 – 1.5	1.5 – 2	2 – 2.5	2.5 – 3	3 – 3.5	3.5 – 4
Frequency	4	18	$a$	25	9	5
Cumulative frequency	4	22	51	$b$	85	90

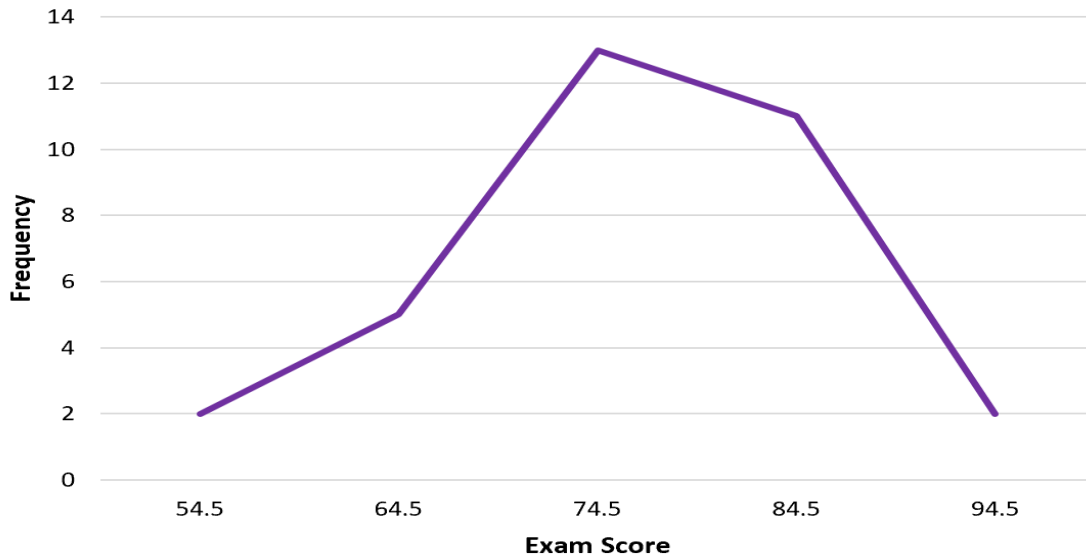
- Find the values of  $a$  &  $b$  to complete the frequency distribution table.
  - Find the median class and hence find the percentage of the frequency of that class.
4. The following table shows the results of a survey to find the average daily time, in minutes, that a group of children spent in internet chat rooms, where the mean time was estimated to be 27.5 minutes. From an equation involving  $f$  and hence show that the total number of children in the survey was 26.

Time per day	0-10	10-20	20-40	40-80
No of children	2	$f$	11	4

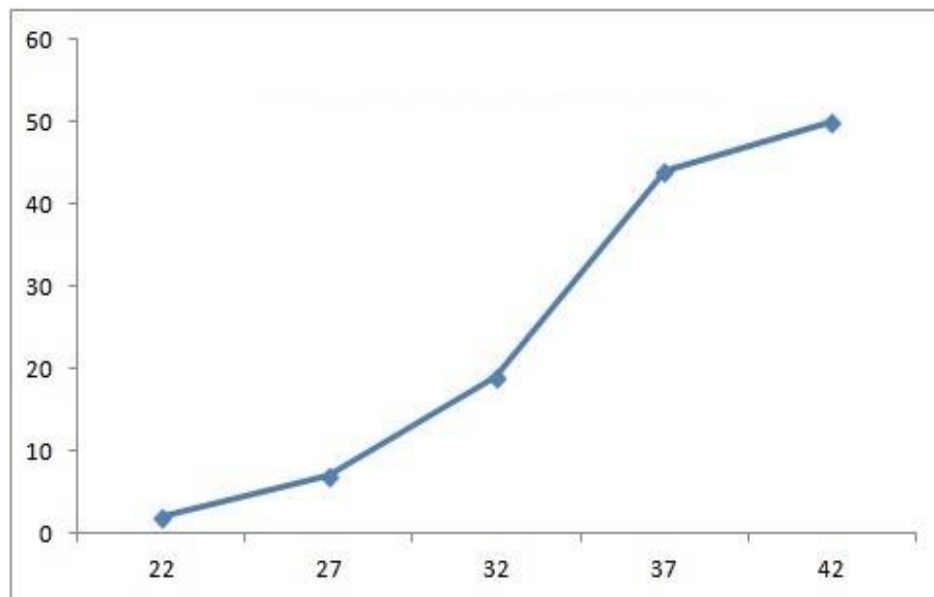
5. If the mode of the following frequency table is 2.34, find the unknown frequency  $f$ . Hence, find the statistical index (standardized value) of  $x = 2.75$ .

CGPA	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4
No of students	7	18	35	$f$	10	3

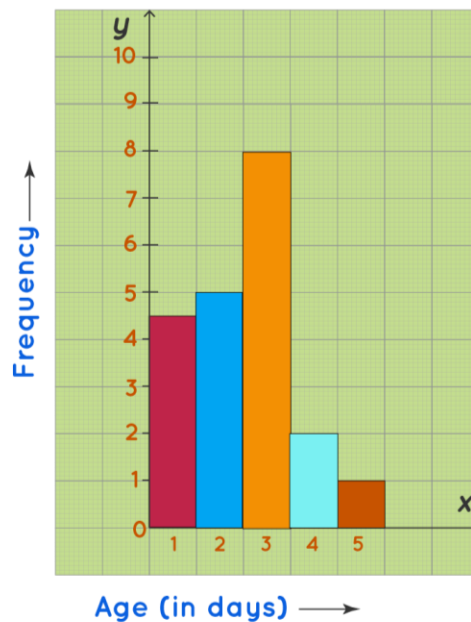
6. From the following frequency polygon construct the corresponding frequency table and hence find the standardized value for  $x = 60$ .



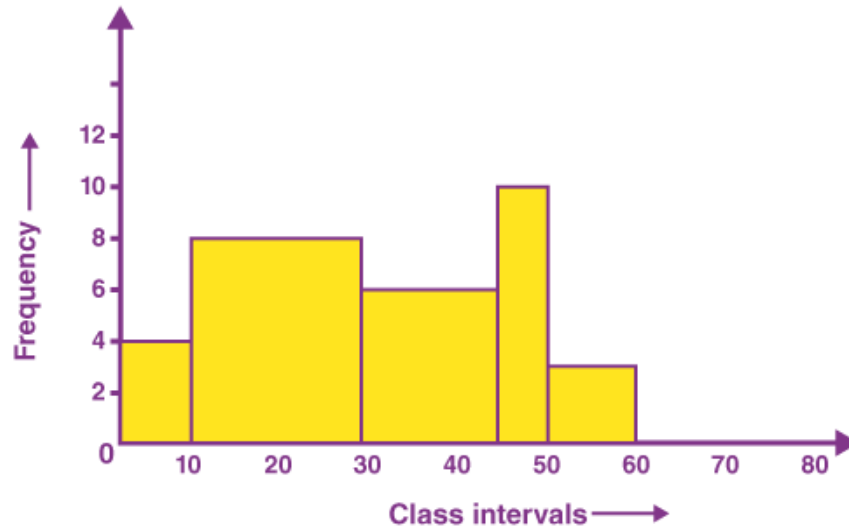
7. From the following cumulative frequency polygon construct the corresponding frequency table and hence find the geometric and harmonic mean of the distribution.



8. From the following histogram construct the corresponding frequency table. Find the median and corresponding mean deviation of the distribution.

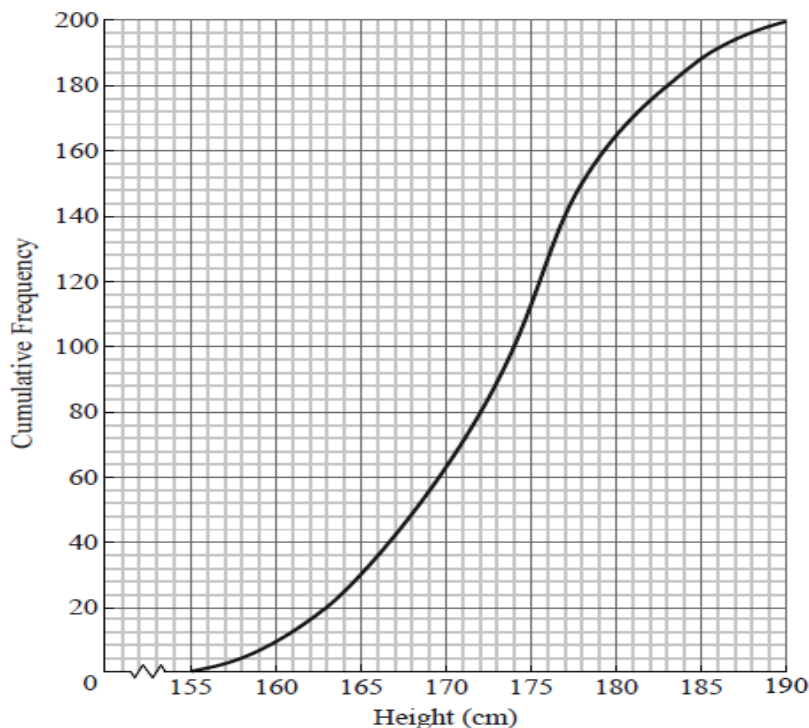


9. Explain the information depicted by the following histogram.



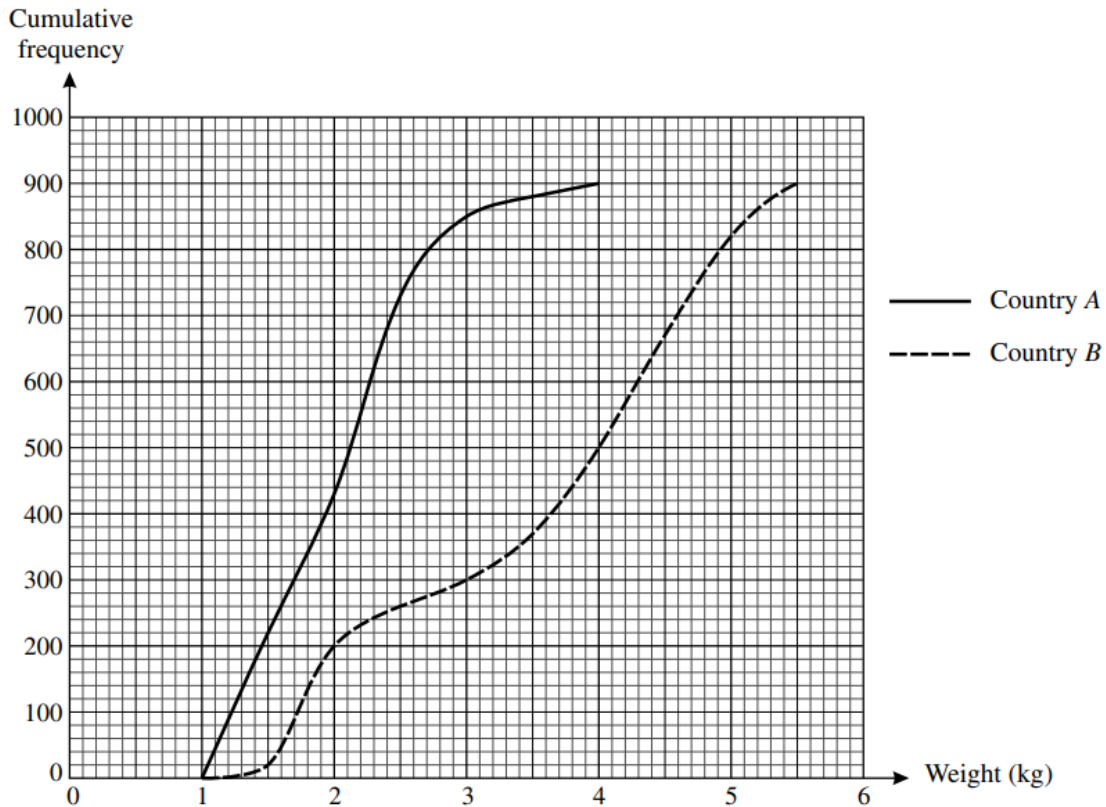
10. In a class of 35 students the CGPA of the class are summarized as the sums  $\sum x_i = 112$  and  $\sum x_i^2 = 562$ . Find their mean and standard deviation.
11. A summary of 24 observations of  $x$  gave that  $\sum(x - a) = -73.2$  and  $\sum(x - a)^2 = 2215$ , where the mean of these observations is 8.95.
- Find the value of the constant  $a$ .
  - Find the standard deviation of  $x$ .

12. A group of 10 married couples and 3 single men found that the mean age  $\bar{x}_w$  of the 10 women was 41.2 years and the standard deviation of women's ages was 15.1 years. For the 13 men, the mean age  $\bar{x}_m$  was 46.3 years and the standard deviation was 12.7 years.
- Find the mean age of whole group of 23 people.
  - The individual women's ages are denoted by  $x_w$  and individual men's ages are denoted by  $x_m$ . By first finding  $\sum x_w^2$  and  $\sum x_m^2$ , find the standard deviation of the whole group.
13. In a certain factory there are four working groups and they need 3, 4, 5 and 2 hours per product to make. What is the approximate average time required to make a product by those groups?
14. Let the class marks of a certain population table are 17, 22, 27, 32 & 37 and the corresponding frequencies are 9, 13, 8, 10 & 15.
- Determine size of the classes and hence construct the original classes.
  - Find the median class and hence find the percentage of the frequency of that class.
  - Find the sample size.
15. The cumulative frequency graph illustrates the height of 200 students in a community.



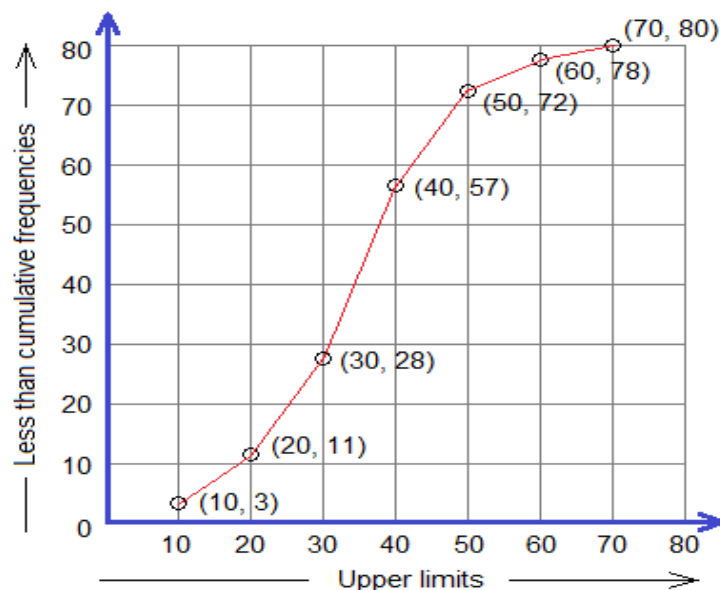
- State the range of the data.
- Construct a box and whisker plot to illustrate the data.
- What percentage of students have height more than 170 cm.?
- Find the position of the height of 40<sup>th</sup> percentile and 70<sup>th</sup> percentile.
- Find the outlier if there exists any.

16. The birth weights of random samples of 900 babies born in country A and 900 babies born in country B are illustrated in the cumulative frequency graphs. Use suitable data from these graphs to compare the central tendency and spread of the birth weights of the two sets of babies.



17. There are three groups of people have average earnings \$150, \$170 & \$ 140 respectively. If the first two groups contain 17 & 12 people respectively and the combined average of earnings is \$153.25, find the number of people in the last group.
18. If the mode of a certain frequency table is 65.5 and the lower limit of the modal class is 60.5 with the class size 10, find the frequency of the modal class. Here frequency difference of the modal class and pre-modal class is 7 and frequency of post-modal class is 14.
19. If the standard deviation of a frequency table is 3.6 and coefficient of standard deviation is 6.55%, find the arithmetic mean of that table.
20. Suppose the first four raw moments of a population are -3.7, 94, -547.2 & 1200 respectively.
- Find the first four central moments.
  - Estimate the coefficient of skewness and kurtosis.
  - Show your result graphically and comment about your findings.

21. For the following Cumulative Frequency Polygon construct a Box-Whisker plot.



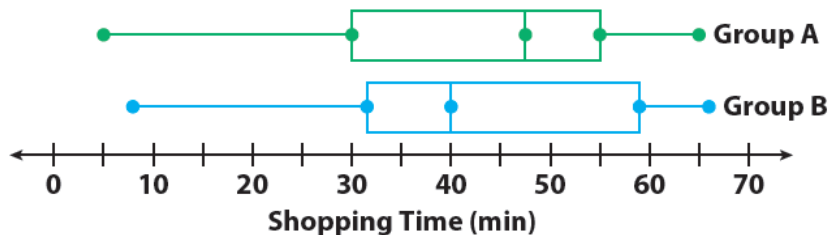
22. The following data represent the daily ticket sales at a park during three weeks.

52, 73, 34, 85, 62, 79, 89, 50, 45, 83, 84, 91, 85, 84, 87, 44, 86, 41, 35, 73, 86

- Construct a Stem-Leaf diagram to illustrate the data.
- Use your diagram to find the median and mode of the data.
- Find the interquartile range.
- Comment about the skewness of the given data.

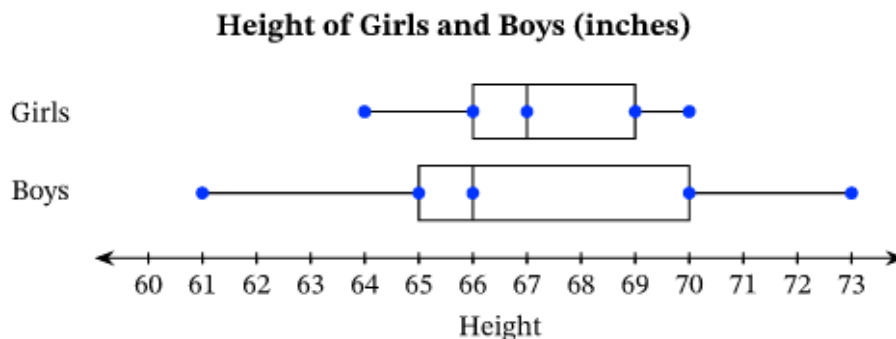
23. You are given that the lower quartile, median, and upper quartile of the weight of a county are 64, 74, 78, respectively. Draw the Box-Whisker plot of the given data. State the nature of the distribution of the given data

24. Distributions of the shopping time (in minutes) for two groups of people are given in the following Box-Whisker plot.

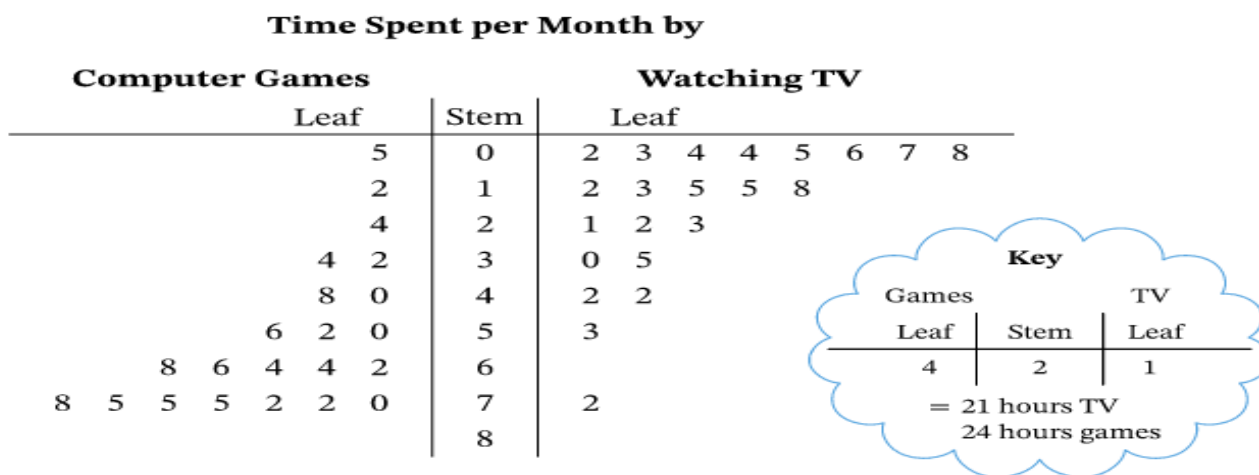


- Find inter-quartile range (IQR) and hence investigate the consistency of the shopping time for the target groups.
- Describe the nature (skewness) of the distributions.

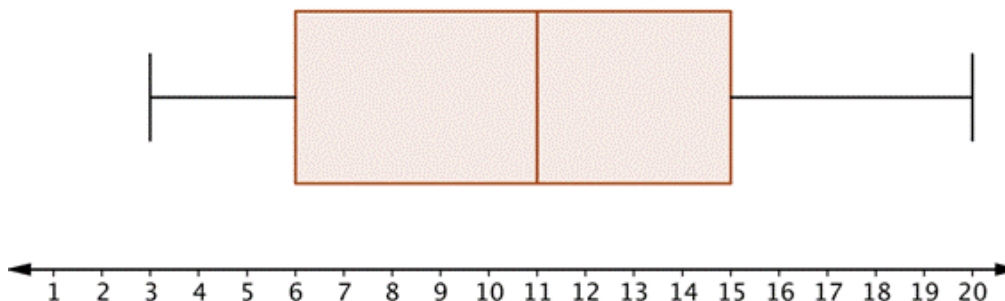
25. Analyze the following dual Box-Whisker plot. You may comment about the central tendency, consistency, skewness, and kurtosis.



26. A survey was done to collect data on the number of hours per month that students in a class spent doing two activities: (i) playing computer games, (ii) watching TV. The data was put into back-to-back Stem-Leaf plot, as shown below.



- (a). If the target is to minimize the number of hours, which activity is better?
- (b). Which one of the activities is consistent?
- (c). Construct a Box-Whisker plot according to the given Stem-Leaf plot.
27. According to the following Box-Whisker plot write the relation among the Mode, Median, and Mean.



28. For the following sets of data design, a two-leaf Stem-Leaf diagram.

Set A	33, 45, 21, 35, 41, 36, 37, 39, 35, 46, 47, 40, 20, 22, 35, 47, 34, 49, 23
Set B	34, 30, 31, 38, 42, 36, 27, 34, 37, 38, 19, 36, 36, 20, 32, 20, 30, 37, 38, 34, 41, 19

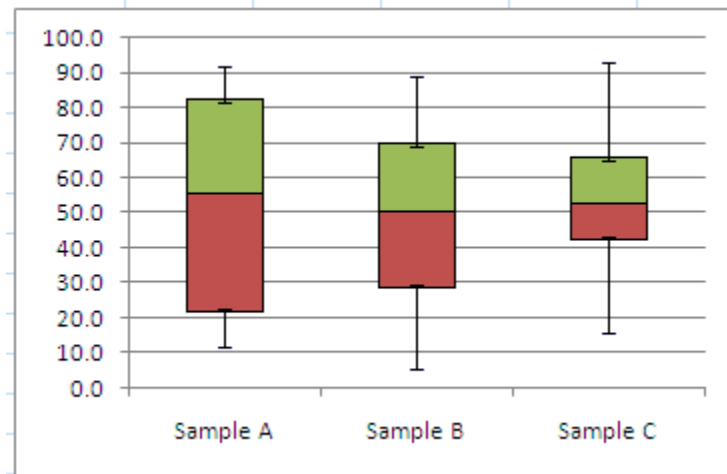
- Find the average of the given sets.
  - Find the half way position of both of the data set and investigate the consistency.
  - Convert your Stem-Leaf diagram to the Box-Whisker plot.
  - Does your diagram have any outliers?
  - Could you find the modes of the data sets?
29. The back-to-back stem and leaf plot below shows the LDL cholesterol levels (in milligram per deciliter mg/dL) of two groups of people, smokers and non-smokers. The digits in the stem represent the hundreds and tens and the digit in the leaf represent the ones. So, for example  $11|8 = 118$  and so on.

smokers		non smokers
	10	0 8
1	11	1 5 8 9
3 0	12	3 6 6 8 8 9 9
4 2 0	13	0 1 7 7 8 8 9
6 5 3 1	14	0 1 5 6 9 1
9 7 6 5 0	15	2 4 2
2 9 7 6 4 1	16	0 1 8
9 8 7 4 3	17	4 6
6 5 3 0	18	1
7 5 1	19	
5 0	20	

- People with a cholesterol level of 129 or less are said to have a near ideal level of cholesterol. How many people, in each group, have a near ideal level of cholesterol?
- People with a cholesterol level between 130 and 159 inclusive are said to be in the border high. How many people, in each group, are in the border high?
- People with a cholesterol level between 160 and 189 inclusive are said to have a high level of cholesterol. How many people, in each group, have a high level of cholesterol?
- People with a cholesterol level of 190 or above are said to have a very high level of cholesterol. How many people, in each group, have a very high level of cholesterol?
- Comparing the two groups, which group has more people with a higher level of cholesterol?



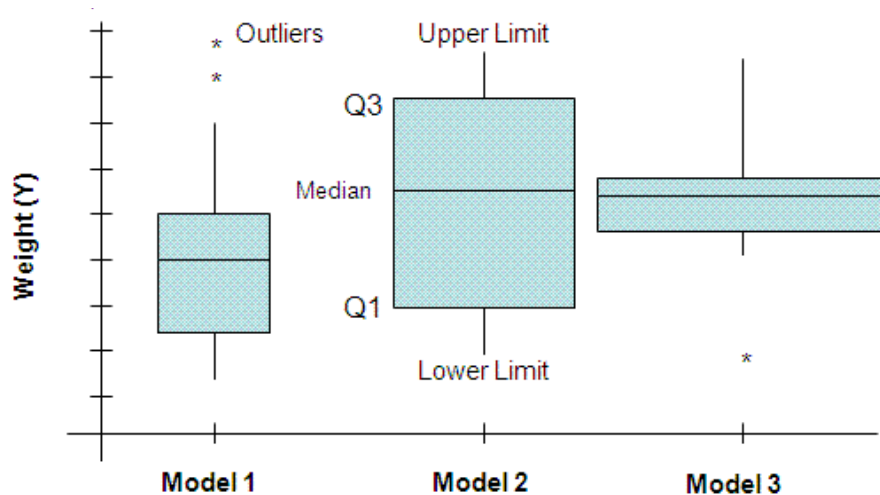
30. For the following Box-Whisker plot find the appropriate central tendency and check the consistency of the given samples.



31. Sketch the Box-Whisker plots for the Frequency Polygons given below.



32. Compare the data sets represented through the Box-Whisker plots given below.



33. Discuss the strength of correlation from the following Pearson' correlation coefficient.  
 (i)  $r = -0.15$       (ii)  $r = -1$       (iii)  $r = 0.65$       (iv)  $r = 1$       (v)  $r = 0$

34. If the correlation coefficient of two variables is 0.72 and regression coefficient of  $x$  on  $y$  is 1.08. If  $\bar{x} = 29.2$  and  $\bar{y} = 37.5$  find the regression line of  $y$  on  $x$ . Also, find the value of  $y$  when  $x = 44$  graphically.

35. If the correlation coefficient of  $x$  &  $y$  is 0.75 and the corresponding standard deviations 1.25 & 1.75. Find the regression coefficient of  $y$  on  $x$  and  $x$  on  $y$ .

36. If the correlation coefficient of two variables is 0.65 and regression coefficient of  $y$  on  $x$  is 1.68. Also,  $\bar{x} = 32.3$  and  $\bar{y} = 45.6$ .

- (a). Find the regression coefficient of  $x$  on  $y$ .  
 (b). Find and sketch the regression line  $x$  on  $y$ .  
 (c). Predict the value of  $x$  when  $y$  is 52. Also, verify your result graphically.

37. For the following data find the correlation co-efficient. How much  $y$  depends on  $x$ ? Also, find the corresponding regression line.

$x$	5	12	18	23	27	30	26	22
$y$	18	16	13	11	9	7	10	13

38. Fit a least-squares line to the following set of data by using  $x$  as the dependent variable. Also, from the graph of least-squares line predict  $x$  for  $y = 11$ .

$x$	1	3	4	6	8	9	11	14
$y$	1	2	4	4	5	7	8	9

39. Find the rank correlation co-efficient between obtained places of 8 students in Mathematics and Physics.

Serial	1	2	3	4	5	6	7	8
Mathematics	3	1	6	5	7	4	8	2
Physics	8	3	1	2	6	5	4	7

40. Ten candidates were ranked as follows by two independent examiners, according to the score they obtained in an interview. Calculate the Spearman's rank correlation Coefficient and interpret the result.

Candidate Number	1	2	3	4	5	6	7	8	9	10
Ranked by Ex. 1	7	9	1	3	8	4	10	5	6	2
Ranked by Ex. 2	9	5	1	4	6	7	8	2	10	3

41. A department store has the following statistics of sales (Y) for a period of 2 years of 10 salespersons who have varying years of experience (X) in sales promotion.

Experience (X) in Years	1	3	4	4	6	8	10	10	11	13
Average Annual sales (Y) in thousand	80	97	92	102	103	111	119	123	117	136

- (a). Using the above set of data calculate the value of  $r$  (coefficient of correlation) and interpret the result.
- (b). Find the regression line of Y on X in the form  $Y = a + bX$ .
- (c). Sketch a scatter diagram in Years (X) vs Average Annual sales (Y).
- (d). Verify your model found in question (b) with the tabular value for 6 years' experience.
- (e). Predict the annual sales volume of persons what have 12- and 15-years' experience.
- (f). Indicate why it may not be appropriate to use your question to predict the average annual sales at 30 years of experience.