Test 1 Statistics for Data Science, 3rd Semester Course,
RVCE Bangalore Marks 50

1. You know that mean, median and mode are measures of central tendency. Why is the median unaffected by outliers? ***Your answer must not be more than 2 sentences.*** **4 marks**

2. Two students, Spiderman and Batman, from the I3E department of RVCE, want to prove that they know more statistics than you do. (*They probably felt that you did not learn Statistics well from Prof. A*). Both of them visit your class and make the following statements. Justify with a ***logical reason of not more than 2 sentences*** as to which among the two fellows above is wrong. (*A chance to prove that you have actually learnt stats from Prof. A.*) **4 marks**

Spiderman: The number of classes you bunk and hence the attendance percentage have positive correlation.

***Batman*:** The number of classes you attend and hence the attendance percentage have positive correlation.

3. You are told that the average marks scored in a test by a class is 30 on 50, with the marks clustered between 24 to 36. List exactly 2 conclusions that you can draw from the given information. **4 marks**

4. For a data sample containing only negative numbers, is it correct to say that the standard deviation will still be positive? Give a **logical reason** supporting your answer. **4 marks**

5. The following are the scores of Steve Jobless in 5 tests in the Stats for Data Science course at RVCE. Each test is out of 100 marks.

60, 58, 30, 62, 90

Your task is to compute the mean, 20% trimmed mean and weighted mean with the following weights:

Weighted Mean = 0.6*Best score + 0.1*worst score + 0.3*(Best score among the remaining 3 scores).

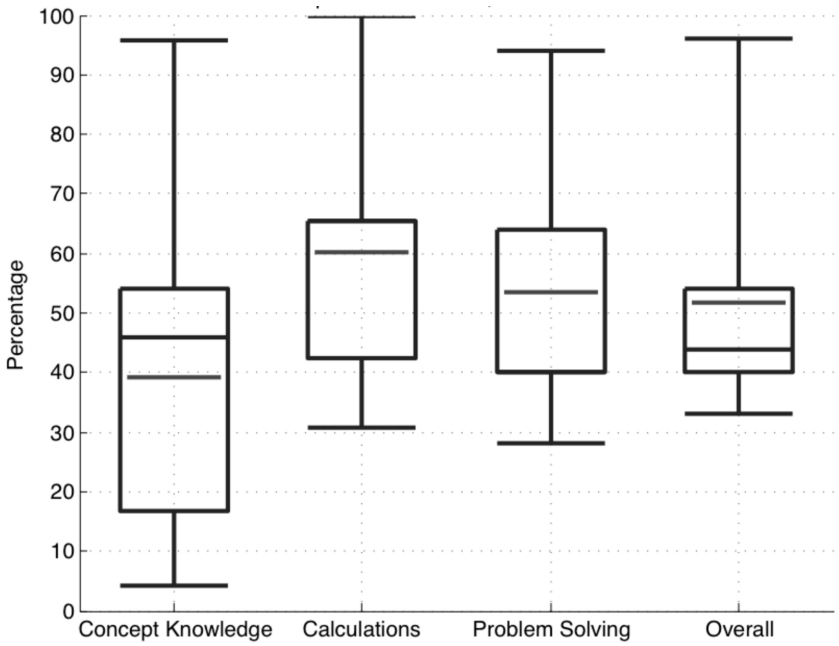
Which of the three values, the mean, 20% trimmed mean and the weighted mean, that you calculated above will benefit Steve Jobless and help him get a better grade in the course? **6 marks**

6. In five attempts, it took a person 11, 15, 12, 8, and 14 minutes to change a tire on a car. Find the mean, variance, and standard deviation. *For variance calculation please use N* and not *N*-1. **6 marks**

7. List exactly 3 reasons (one corresponding to each) as to why mean, median or mode are not useful metrics to describe data. **6 marks**

8. Give a real-life example for each of the following scenario **6 marks**
   (a) left skewed data distribution
   (b) Pair of positively correlated variables (X, Y)
   (c) Data which can be visualized using histogram.

9. The boxplot below corresponds to assessment of students in a course on Statistics. The assessment involves 3 components namely Concept knowledge, calculations and problem solving. *Note: The end to end lines inside the box correspond to median and the other line inside the box, which do not meet the opposite sides are the averages for each. The median for Calculations is the same as 75th percentile and that for problem solving is the same as 25th percentile.* **10 marks**



Answer the following questions with respect to the above boxplot. Please justify your answer in ***one sentence*** in each case.

(a) What can you conclude about the overall performance of students?
(b) Is it correct to conclude that students do much better when it comes to computations and problem solving than concepts?
(c) There is a belief that if you are conceptually strong, you can solve problems easily. Does the box plot reflect that?
(d) What reason can you attribute to the students doing better in problem solving and calculation based questions than conceptual ones?