Quiz 2 Statistics for Data Science, 4th Semester Course,
RVCE Bangalore Marks 10

1. For any two events A and B, with P(A) > 0 and P(B) > 0

Statement 1: $P(A \cup B) = P(A) + P(B)$
Statement 2: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Statement 3: $P(A \cap B) = P(A)\,P(B)$

Choose the correct option:

(a) Statement 1 is always TRUE
(b) Statement 2 is always TRUE
(c) Statement 3 is always TRUE
(d) All 3 statements are ALWAYS TRUE

2. It is well known that in every IPL season, RCB team plays with the slogan *ee saala cup namde* (ಈ ಸಲ ಕಪ್ ನಮ್ದೆ). Consider a hypothetical situation in which RCB gives up playing in IPL after it wins its first IPL Championship. Having enrolled in the course on Statistics for Data Science, your professor for the course tells you to model using a random variable X, the number of IPL seasons that RCB plays before winning the trophy for the first time. What is your choice among the following distributions for describing the random variable X?

(a) Geometric distribution
(b) Binomial distribution
(c) Equally likely distribution
(d) Poisson distribution

3. If X and Y are two random variables such that their correlation coefficient $r = 0$, then

(a) X and Y are independent
(b) X and Y are not linearly correlated
(c) we cannot comment on X and Y, unless we know the exact data
(d) it ALWAYS indicates significant error in measurement

4. Assume X is the data you have collected across 100 samples. For some reason, you decide to add a constant $k \neq 0$ to all the values. Then,

(a) the new average is *old average* + k

(b) the new average and the old average are the same

(c) the new standard deviation is *old standard deviation* + $\sqrt{k}$

(d) the new variance is old variance + k^2

5. A list has 10 entries and each of the entries can either be 1, 2 or 3. With this information, choose that option which is FALSE.

(a) The average of the list can be 1.5

(b) The average of the list can be 2.9

(c) The average of the list can be 3.2

(d) The average of the list can be 2.2

6. A list has 10 entries all of which are 10. With this information, choose the option that is TRUE.

(a) The variance of the list is $\sqrt{10}$

(b) The variance of the list is 100

(c) The variance of the list is 10

(d) The variance of the list is 0

7. Let E_1 and E_2 be two events of a sample space such that $E_1 \subset E_2$ (*Read that as* E_1 *is a subset of* E_2), then $P(E_2|E_1) =$

(a) 0

(b) 0.25

(c) 0.5

(d) 1

8. The levels of frustration that one hits and the amount of time that one spends in Bangalore traffic are negatively correlated random variables.

TRUE / FALSE

9. Let X be the random variable describing the number of sides of an answer booklet that you use in your Statistics for Data Science exam and Y be the random variable describing the number of sides that you leave unused in the same booklet in which you write your Statistics for Data Science exam. The correlation coefficient r between X and Y is

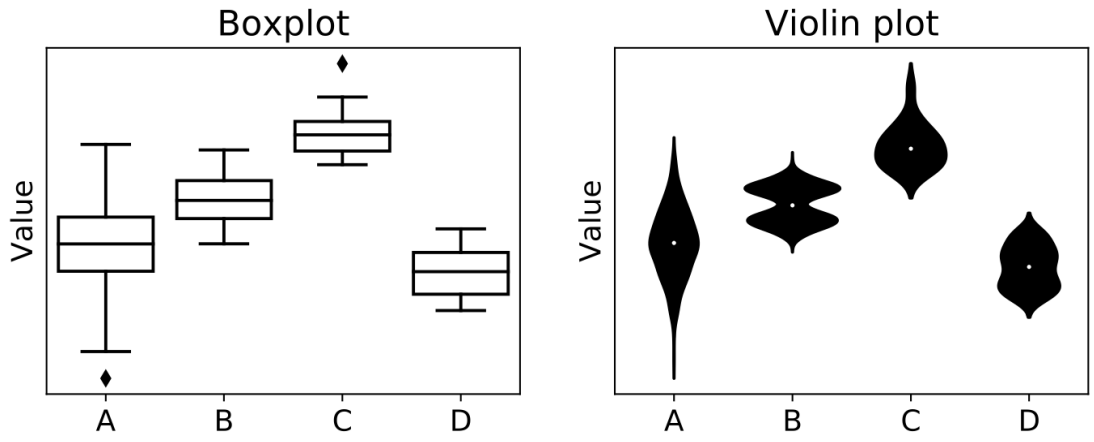
(a) -1

(b) 1

(c) 0

(d) indeterminate as no information about covariance or standard deviation values are given.

10. Identify the additional information that the violin plot on the right figure provides which the boxplot on the left figure does not provide:



(a) The violin plot displays the number of observations that are not found in the boxplot

(b) The violin plot shows the number of missing values

(c) The violin plot does not provide any additional information

(d) The violin plot shows the underlying distribution for each group.