PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-1 - Introduction to Data Science

QUESTION BANK

Summary Statistics

Exercises for Section 1.2

[Text Book Exercise – Section 1.2 – Q. No. [1 – 16] – Pg. No. [23 – 25]]

1. True or false: For any list of numbers, half of them will be below the mean.

2. Is the sample mean always the most frequently occurring value? If so, explain why. If not, give an example.

3. Is the sample mean always equal to one of the values in the sample? If so, explain why. If not, give an example.

4. Is the sample median always equal to one of the values in the sample? If so, explain why. If not, give an example.

5. Find a sample size for which the median will always equal one of the values in the sample.

6. For a list of positive numbers, is it possible for the standard deviation to be greater than the mean? If so, give an example. If not, explain why not.

7. Is it possible for the standard deviation of a list of numbers to equal 0? If so, give an example. If not, explain why not.

8. In a certain company, every worker received a $50-per-week raise. How does this affect the mean salary? The standard deviation of the salaries?

9. In another company, every worker received a5%raise. How does this affect the mean salary? The standard deviation of the salaries?
10. A sample of 100 adult women was taken, and each was asked how many children she had. The results were as follows:

Children	0	1	2	3	4	5
Number of Women	27	22	30	12	7	2

a. Find the sample mean number of children.
b. Find the sample standard deviation of the number of children.
c. Find the sample median of the number of children.
d. What is the first quartile of the number of children?
e. What proportion of the women had more than the mean number of children?
f. For what proportion of the women was the number of children more than one standard deviation greater than the mean?
g. For what proportion of the women was the number of children within one standard deviation of the mean?

11. In a sample of 20 men, the mean height was 178 cm. In a sample of 30 women, the mean height was 164 cm. What was the mean height for both groups put together?

12. Each of 16 students measured the circumference of a tennis ball by four different methods, which were:
Method A: Estimate the circumference by eye.
Method B: Measure the diameter with a ruler, and then compute the circumference.
Method C: Measure the circumference with a ruler and string.
Method D: Measure the circumference by rolling the ball along a ruler.
The results (in cm) are as follows, in increasing order for each method:

Method A: 18.0, 18.0, 18.0, 20.0, 22.0, 22.0, 22.5, 23.0, 24.0, 24.0, 25.0, 25.0, 25.0, 25.0, 26.0, 26.4.
Method B: 18.8, 18.9, 18.9, 19.6, 20.1, 20.4, 20.4, 20.4, 20.4, 20.5, 21.2, 22.0, 22.0, 22.0, 22.0, 23.6.
Method C: 20.2, 20.5, 20.5, 20.7, 20.8, 20.9, 21.0, 21.0, 21.0, 21.0, 21.0, 21.5, 21.5, 21.5, 21.5, 21.6.
Method D: 20.0, 20.0, 20.0, 20.0, 20.2, 20.5, 20.5, 20.7, 20.7, 20.7, 21.0, 21.1, 21.5, 21.6, 22.1, 22.3.

a. Compute the mean measurement for each method.
b. Compute the median measurement for each method.
c. Compute the 20% trimmed mean measurement for each method.
d. Compute the first and third quartiles for each method.
e. Compute the standard deviation of the measurements for each method.
f. For which method is the standard deviation the largest? Why should one expect this method to have the largest standard deviation?
g. Other things being equal, is it better for a measurement method to have a smaller standard deviation or a larger standard deviation? Or doesn't it matter? Explain.

13. Refer to Exercise 12.
    a. If the measurements for one of the methods were converted to inches (1 inch=2.54 cm), how would this affect the mean? The median? The quartiles? The standard deviation?
    b. If the students remeasured the ball, using a ruler marked in inches, would the effects on the mean, median, quartiles, and standard deviation be the same as in part (a)? Explain.

14. There are 10 employees in a particular division of a company. Their salaries have a mean of $70,000, a median of $55,000, and a standard deviation of $20,000. The largest number on the list is $100,000. By accident, this number is changed to $1,000,000.
    a. What is the value of the mean after the change?
    b. What is the value of the median after the change?
    c. What is the value of the standard deviation after the change?

15. Quartiles divide a sample into four nearly equal pieces. In general, a sample of size n can be broken into k nearly equal pieces by using the cutpoints (i/k)(n+1) for i = 1, . . . , k − 1. Consider the following ordered sample:

    2 18 23 41 44 46 49 61 62 74 76 79 82 89 92 95

    a. Tertiles divide a sample into thirds. Find the tertiles of this sample.
    b. Quintiles divide a sample into fifths. Find the quintiles of this sample.

16. In each of the following data sets, tell whether the outlier seems certain to be due to an error, or whether it could conceivably be correct.

    a. The length of a rod is measured five times. The readings in centimeters are 48.5, 47.2, 4.91, 49.5, 46.3.
    b. The prices of five cars on a dealer's lot are $25,000, $30,000, $42,000, $110,000, $31,000.