PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit-1 - Introduction to Data Science

QUESTION BANK – SOLVED

Summary Statistics

Exercises for Section 1.2 [Text Book Exercise– Pg. No. [23 – 25]]

1. **A sample of 100 adult women was taken, and each was asked how many children she had. The results were as follow**

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?**

[Text Book Exercise – Section 1.2 – Q. No.10 – Pg. No. 24]

Solution:

a. **Find the sample mean number of children.**

Sample Size (n) = 6

$$\sum x = 27(0) + 22(1) + 30(2) + 12(3) + 7(4) + 25=156$$

$$\therefore \bar{x} = \sum x/n = 156/100 = 1.56$$

b. Find the sample standard deviation of the number of children.

$$\sum x^2 = 27(0^2) + 22(1^2) + 30(2^2) + 12(3^2) +$$
$$742+252=412$$
$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)$$
$$= \frac{1}{99}[412 - 100(1.56^2]$$
$$= 1.7034$$
$$s = \sqrt{s^2} = \sqrt{1.7034} = 1.3052$$

c. Find the sample median of the number of children.

Position for median=(n+1)(.5)=(101)(.5)=50.5
The sample median is the average of the 50th and 51st value.

$$x_{50} = 2;\ x_{51} = 2;\ \therefore Median = 2$$

d. What is the first quartile of the number of children?

Position for 1st quartile=(n+1)(.25)=(101)(.25)=25.25
The sample median is the average of the 25th and 26th value.

$$x_{25} = 0;\ x_{26} = 0;\ \therefore \text{1st quartile} = 0$$

e. What proportion of the women had more than the mean number of children?

(30+12+7+2) women had more than the mean number of children.
$\therefore proportion = 51/100 = 51\%$

f. For what proportion of the women was the number of children more than one standard deviation greater than the mean?

$\bar{x} + s = 1.56 + 1.3052 = 2.8652$
(12+7+2) women had more than 2.86 children.
$\therefore proportion = 21/100 = 21\%$

g. For what proportion of the women was the number of children within one standard deviation of the mean?

$\bar{x} - s = 1.56 - 1.3052 = 0.2548$
(22+30) women had between 0.2548 and more than 2.86 children. $\therefore proportion = 52/100 = 52\%$

2. **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?**

   [Text Book Exercise – Section 1.2 – Q. No.11 – Pg. No. 24]

   Solution:
   The total height of the 20 men is 20 × 178 = 3560.

   The total height of the 30 women is 30 × 164 = 4920.

   The total height of all 50 people is 3560 + 4920 = 8480.

   There are 20 + 30 = 50 people in total.

   Therefore the mean height for both groups put together is 8480/50 = 169.6 cm.

3. **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

   [Text Book Exercise – Section 1.2 – Q. No.15 – Pg. No. 24]

   a. **Tertiles divide a sample into thirds. Find the tertiles of this sample.**

   Solution:
   The tertiles are 45 and 77.5.

   b. **Quintiles divide a sample into fifths. Find the quin-tiles of this sample.**

   Solution:
   The quintiles are 32, 47.5, 75, and 85.5.

4. **Find a sample size for which the median will always equal one of the values in the sample.**
   [Text Book Exercise – Section 1.2 – Q. No.4 – Pg. No. 23]

Solution:

The sample size can be any odd number

5. **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.**
   [Text Book Exercise – Section 1.2 – Q. No.7 – Pg. No. 23]

   Solution:
   Yes. If all the numbers on the list are the same, the standard deviation will equal 0.