Linear Statistical Models

Week-1: Graded Assignment

Objective Assignment: (Auto-grading) **Max. Marks: 10**

1. Heights (in inches) of all members of a family are 65, 66, 67, 67, 68, 69, 70 and 72. Based on the given information, answer the following questions.

   (a) Which statement of code in R-software can be used to read the data as a *vector* and find the number of members in the family. [1 Mark]

   (i) $> x \leftarrow$ c(65, 66, 67, 67, 68, 69, 70, 72)
   $> n \leftarrow count(x)$

   (ii) $> x \leftarrow$ c(65, 66, 67, 67, 68, 69, 70, 72)
   $> n \leftarrow total(x)$

   (iii) $> x \leftarrow$ c(65, 66, 67, 67, 68, 69, 70, 72)
   $> n \leftarrow size(x)$

   (iv) $> x \leftarrow$ c(65, 66, 67, 67, 68, 69, 70, 72)
   $> n \leftarrow length(x)$

   Answer: iv

   Solution:

   $length(x)$ is an in-built function in R - software to find the number of observation in a vector. Thus, option (iv) is the correct.

   (b) The output obtained by executing the command to compute the number of members in the family is: [1 Mark]

   (i) 8
   (ii) [1] 8
   (iii) [1] 7
   (iv) 7

   Answer : ii

   Solution:

   Any data is stored as a vector in R. [1] represents the position of that element in the vector.
   Thus, by executing the correct command, i.e., command in option (iv) of part(a), we get the output as [1] 8.
   Thus, option (ii) is correct.

   (c) Which of the following code can be used in R-software to find the sum of the height (in inches) of family members. [1 Mark]

(i) $> total \leftarrow add(x)$

(ii) $> total \leftarrow sum(x)$

(iii) $> total \leftarrow total(x)$

(iv) $> total \leftarrow summation(x)$

Answer: ii

Solution:

$sum()$ is an in-built function in R - software to find the sum of all observations in a vector.

Now, on applying the function $sum(x)$, we will get the sum of the height of family members.

Thus, option (ii) is the correct.

(d) The output obtained by executing the command to compute the sum of the heights (in inches) of family members is: [1 Mark]

(i) [1] 472

(ii) 544

(iii) [1] 544

(iv) 472

Answer : iii

Solution:

Any data is stored as a vector in R. [1] represents the position of that element in the vector.

Thus, by executing the correct command, i.e., command in option (ii) of part(c), we get the output as [1] 544, which is the sum of heights in inches) of family members

Thus, option (iii) is correct.

(e) Which of the following code(s) can be used in R-software to compute the average height (in inches) of family members. [1 Mark]

(i) $> xbar \leftarrow average(x)$
$> \text{print}(xbar)$

(ii) $> xbar \leftarrow avg(x)$
$> \text{print}(xbar)$

(iii) $> xbar \leftarrow mean(x)$
$> \text{print}(xbar)$

(iv) $> xbar \leftarrow sum(x)/length(x)$
$> \text{print}(xbar)$

Answer : iii, iv

Solution:

$mean()$ is an in-built function in R - software to find the average of all observations

in a vector.
Now, on executing the code $xbar \leftarrow mean(x)$, the average of the height of family members will be computed and stored in $xbar$. And, the command $\text{print}(xbar)$ will print the output.
Thus, option (iii) is the correct.

Since we can compute the average by formula $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ which can be done by using the command $xbar \leftarrow sum(x)/length(x)$ and output can be printed by the command $\text{print}(xbar)$.
Thus, option (iv) is also correct.

(f) The output obtained by executing the command to compute the average heights (in inches) of family members is: (Enter only the numerical value obtained by executing the command correct to 1 decimal place) [1 Mark]
Answer : 68, Range: 67.9 to 68.1

Solution:
By executing the codes of options (iii) or (iv) of the above part, we will get the output as 68 which is average heights (in inches) of family members.

(g) Which of the following code can be used in R-software to compute the sample variance of heights (in inches^2) of family members. [1 Mark]

(i) $> svar \leftarrow var.s(x)$
$> \text{print}(svar)$

(ii) $> svar \leftarrow s.var(x)$
$> \text{print}(svar)$

(iii) $> svar \leftarrow svariance(x)$
$> \text{print}(svar)$

(iv) $> svar \leftarrow var(x)$
$> \text{print}(svar)$

Answer : iv

Solution:
$var()$ is an in-built function in R - software to find the sample variance of all observations in a vector.
Now, on executing the code $svar \leftarrow var(x)$, the sample variance of heights of family members will be computed and stored in $svar$. And, the command $\text{print}(svar)$ will print the output.
Thus, option (iv) is the correct.

(h) The output obtained by executing the command to compute the sample variance of heights (in inches^2) of family members is: (Enter only the numerical value obtained by executing the command correct to 2 decimal places) [1 Mark]
Answer : 5.14, Range: 5.11 to 5.17

Solution:

By executing the code of option (iv) of the above part, we will get the output as 5.142857 which is the sample variance of heights (in inches^2) of family members.

(i) Which of the following code can be used in R-software, compute the population variance of heights (in inches^2) of family members. [1 Mark]

(i) $> pvar \leftarrow var.p(x)$
$>$ print$(pvar)$

(ii) $> pvar \leftarrow p.var(x)$
$>$ print$(pvar)$

(iii) $> pvar \leftarrow svar * n/(n-1)$
$>$ print$(pvar)$

(iv) $> pvar \leftarrow svar * (n-1)/n$
$>$ print$(pvar)$

Answer : iv

Solution

Since, there is not any inbuilt function in R - software to compute the population variance of the observations. Therefore, we can use the relationship formula between sample variance and population variance which is given as:

$$(n-1) \times s^2 = n \times \sigma^2$$

$$\implies \sigma^2 = \frac{s^2 \times (n-1)}{n} \qquad ...(*)$$

Where, s^2 and σ^2 denote sample variance and population variance respectively.
Since variable *svar* has stored the value of sample variance of heights of family members in the above part. Therefore, by the equation (*), it is clear that the command $pvar \leftarrow svar * (n-1)/n$ will compute the value of population variance and stored in variable *pvar*.
And, the command print$(pvar)$ will print the output.
Hence, option (iv) correct.

(j) The output obtained by executing the command to compute the population variance of heights (in inches^2) of family members is: (Enter only the numerical value obtained by executing the command correct to 2 decimal places) [1 Mark]
Answer : 4.50, Range: 4.47 to 4.53

Solution:

By executing the code of option (iv) of the above part, we will get the output as 4.5 which is the population variance of heights (in inches^2) of family members.