

Module 7 - Summary statistics of a continuous variable

Anik Chakraborty, (Intern) R Team, FOSSEE, Department of Mathematics, IIT Bombay

December 27, 2022

Steps to find summary measures

This module illustrates the procedure to find summary statistics of a continuous variable in R.

Step 1. Select a continuous data column using the methods described in the Modules 1, 2, and 3, and save it in the variable **smv**. In this module, we shall use the data column 'SMV'.

```
smv = Indian_Agriculture_Data$SMV
```

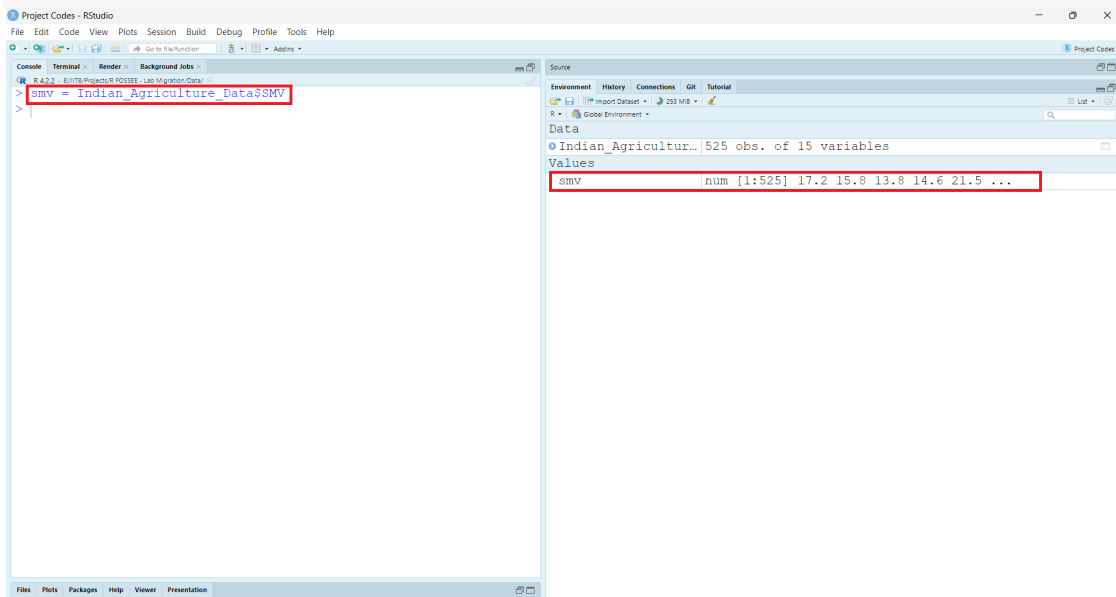


Figure 1: Select and store a continuous data column in the variable **smv**.

Step 2. Execute the following command to find the *mean* by using the **mean()** function of R.

```
mean(smv)
```

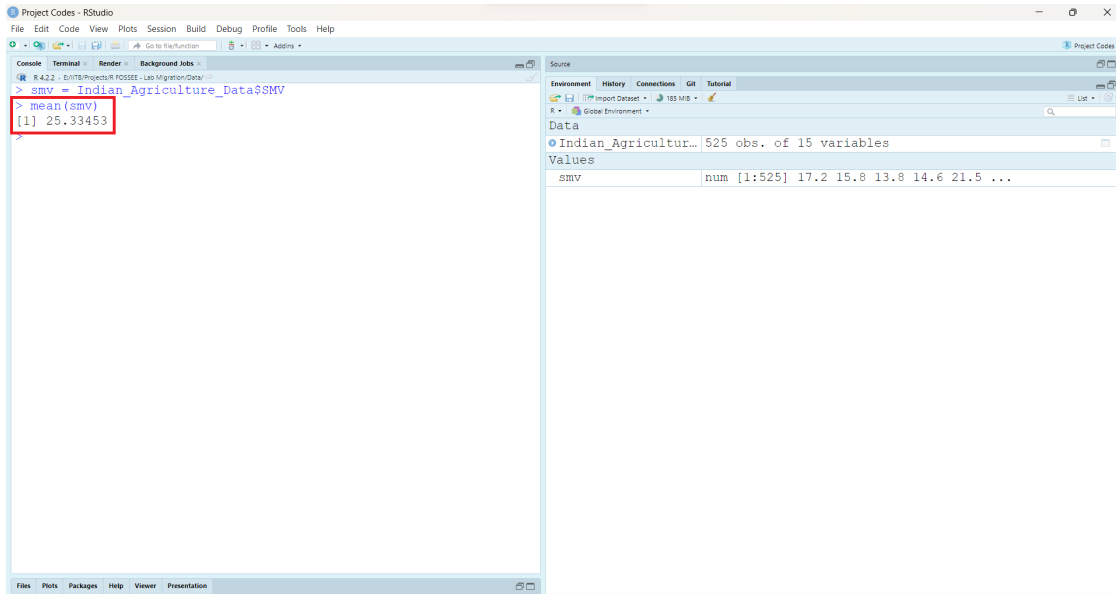


Figure 2: Mean of **SMV** column data

Step 3. Execute the following command to find the *median* by using the **median()** function of R.

```
median(smv)
```

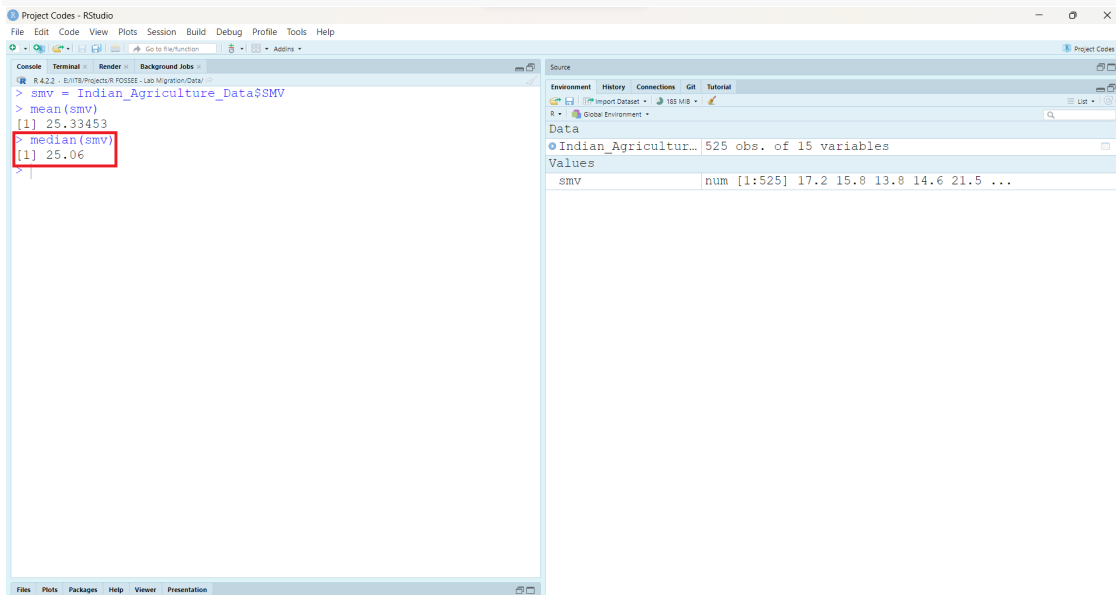


Figure 3: Median of **SMV** column data

Step 4. Execute the following command to find the *minimum* and *maximum* by using the **min()** and **max()** function of R, respectively. There is also a function **range()** to get the *minimum* and *maximum* simultaneously.

```
min(smv)
max(smv)
range(smv)
```

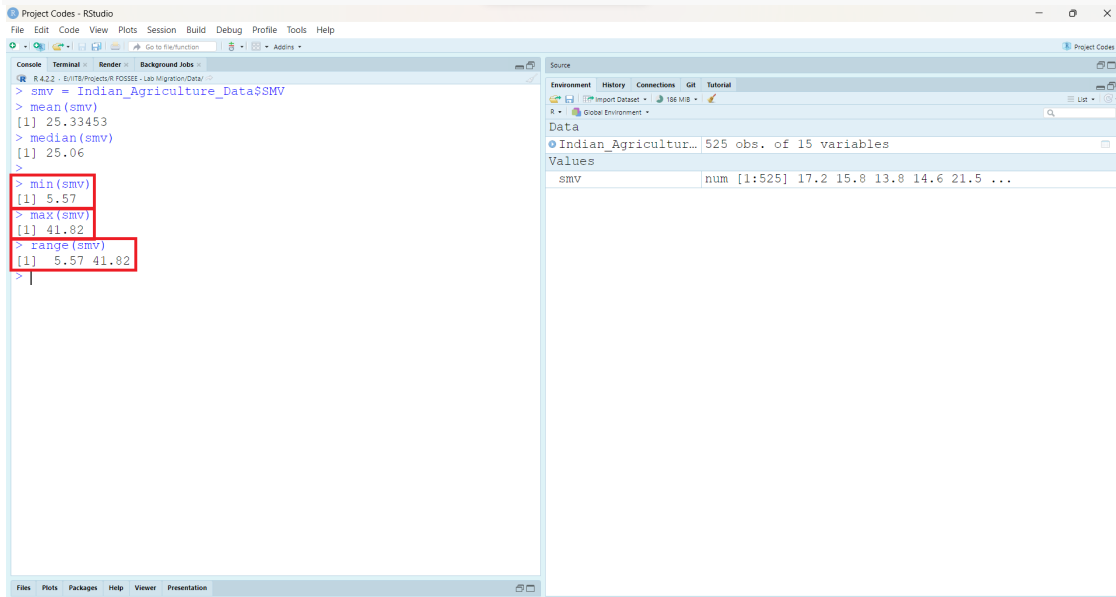


Figure 4: Minimum and Maximum of **SMV** column data

Step 5. Execute the following command to find the *variance* by using the **var()** function of R.

Formula of variance:
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

```
var(smv)
```

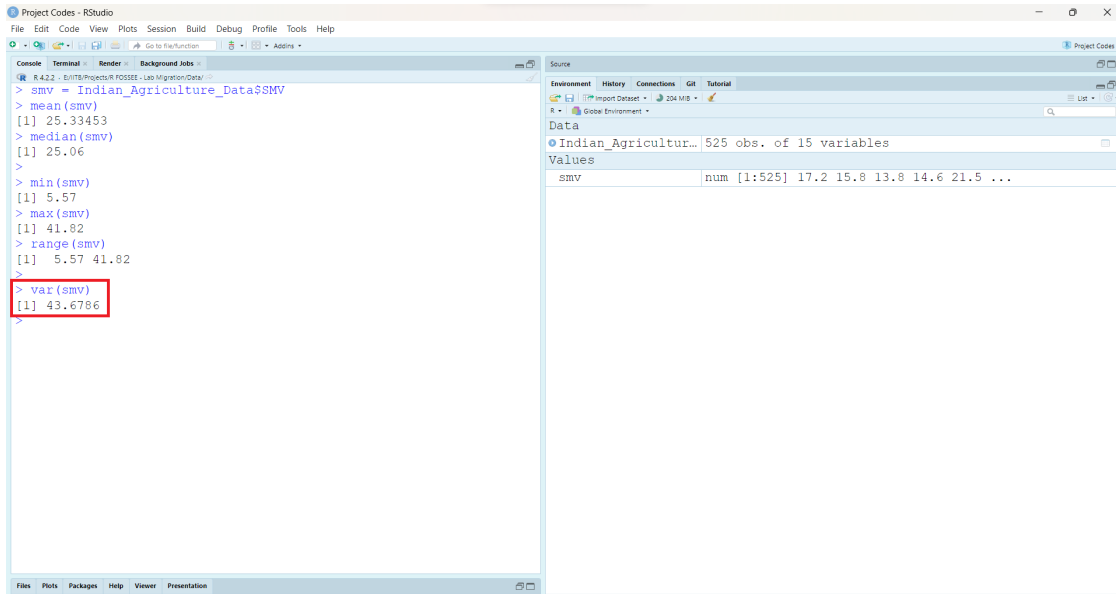


Figure 5: Variance of *SMV* column data

Step 6. Execute the following command to find the *standard deviation* by using the `sd()` function of R. It is the square-root of the variance.

```
sd(smv)
sqrt(var(smv))
```

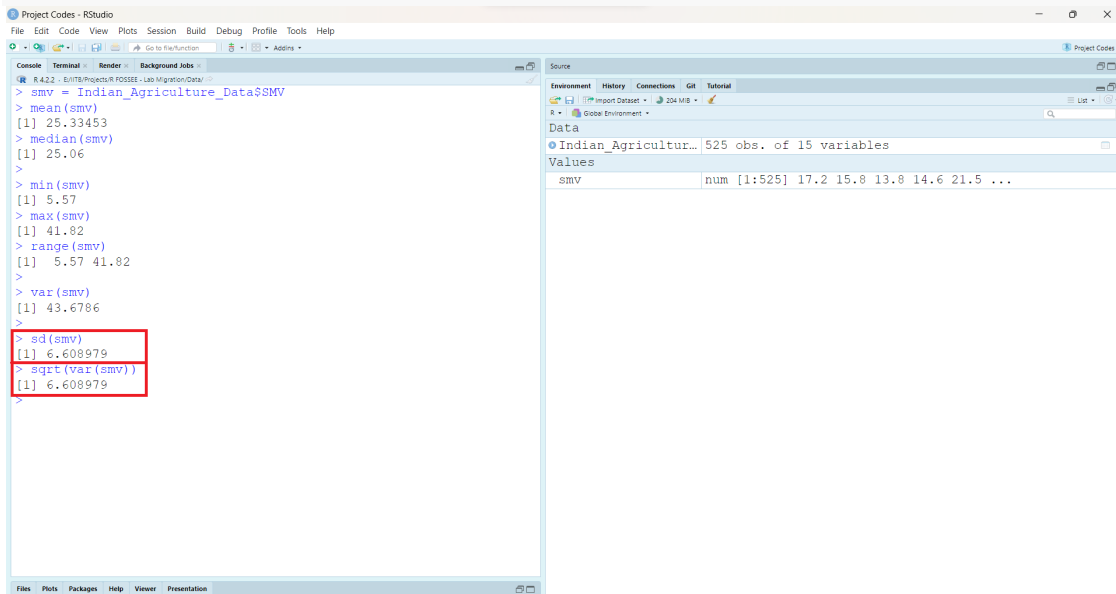


Figure 6: Standard Deviation of *SMV* column data

Step 7. Execute the following command to find the *skewness* and *kurtosis* by using the `Skew()` and `Kurt()` function of the *DescTools* package of R.

Moment measures of skewness and kurtosis

$$\text{Skewness: } \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}, \text{ Kurtosis: } \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3$$

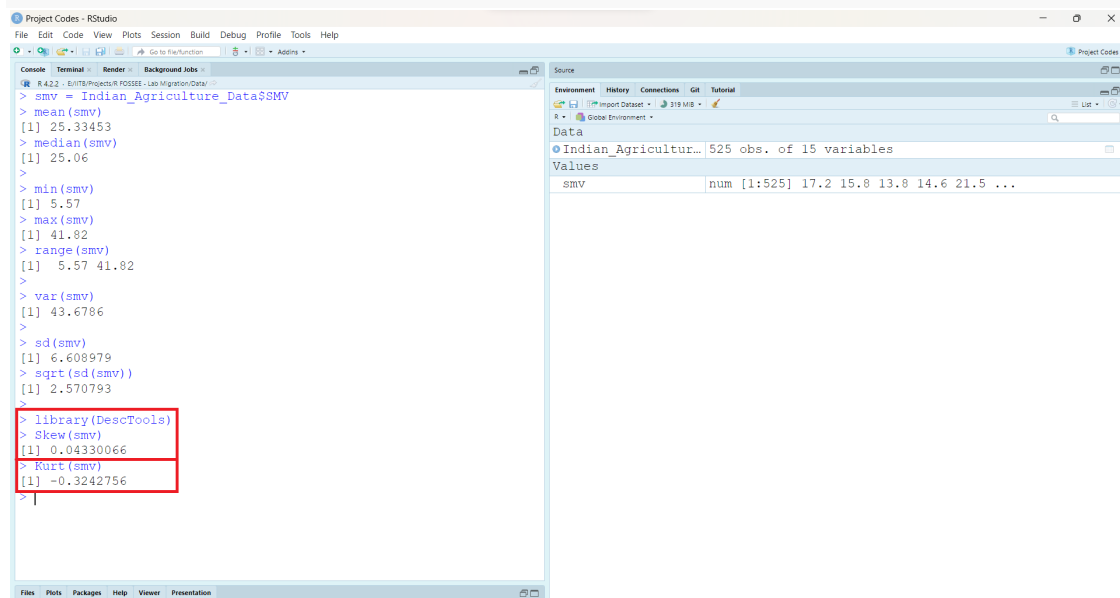
```
# Loading 'DescTools' Library to use its 'Skew()' and 'Kurt()' function
library(DescTools)
```

```
# Skewness
```

```
Skew(smv)
```

```
# Kurtosis
```

```
Kurt(smv)
```



```
Project Codes - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
R 4.2.2 - 5015/Projects/R_FOSSIE - Lab Migration/Data/...
Source
Environment History Connections Git Tutorial
Data
Indian Agricultur... 525 obs. of 15 variables
Values
smv num [1:525] 17.2 15.8 13.8 14.6 21.5 ...

> smv = Indian_Agriculture_Data$SMV
> mean(smv)
[1] 25.33453
> median(smv)
[1] 25.06
>
> min(smv)
[1] 5.57
> max(smv)
[1] 41.82
> range(smv)
[1] 5.57 41.82
>
> var(smv)
[1] 43.6786
>
> sd(smv)
[1] 6.608979
> sqrt(sd(smv))
[1] 2.570793
>
> library(DescTools)
> Skew(smv)
[1] 0.04330066
> Kurt(smv)
[1] -0.3242756
>
```

Figure 7: Skewness and Kurtosis of **SMV** column data

Step 8. Execute the following command to find the *summary statistics* by using the **summary()** function of R. It gives the *minimum-maximum, three quartiles and mean* as output.

```
summary(smv)
```

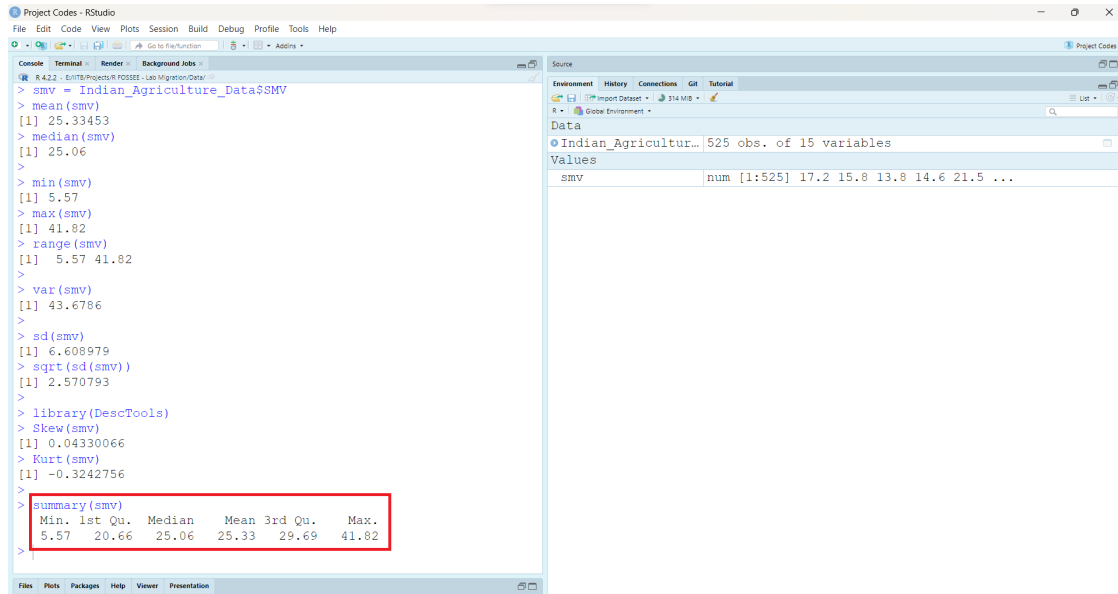


Figure 8: Summary statistics of **SMV** column data

Spoken Tutorials

For more details, refer to the [Plotting Histograms and Pie Chart Spoken Tutorial](#) video.