

## MOOC Course - Introduction to R Software

July 2021

### Assignment 7

1. Suppose a die is rolled 10 times and following numbers on the upper face are recorded: 4,5,5,1,2,1,1,6,3,1. Which one of the following is the correct command to obtain the absolute frequencies of this data in R?

- a. `table(c(4,5,5,1,2,1,1,6,3,1))`
- b. `table(4,5,5,1,2,1,1,6,3,1)`
- c. `table(c(4,5,5,1,2,1,1,6,3,1))/length(c(4,5,5,1,2,1,1,6,3,1))`
- d. `table(4,5,5,1,2,1,1,6,3,1)/length(4,5,5,1,2,1,1,6,3,1)`

Solution:

```
R Console
> table(c(4,5,5,1,2,1,1,6,3,1))
 1  2  3  4  5  6
4  1  1  1  2  1
> |
```

2. Suppose a die is rolled 20 times and following numbers on the upper face are recorded: 5,4,6,5,3,3,5,1,4,4,2,1,5,5,6,3,1,1,2,1,1. Which one of the following is the correct outcome of the following command in R?

`table(c(5,4,6,5,3,3,5,1,4,4,2,1,5,5,6,3,1,1,2,1,1))` ?

a.

```

  1    2    3    4    5    6
0.6 0.2 0.3 0.3 0.5 0.2

```

b.

```

0.6 0.2 0.3 0.3 0.5 0.2
  1    2    3    4    5    6

```

c.

```

1 2 3 4 5 6
6 2 3 3 5 2

```

d.

```

6 2 3 3 5 2
1 2 3 4 5 6

```

Solution:

```

R Console
> table(c(5,4,6,5,3,3,5,1,4,4,2,1,5,5,6,3,1,1,2,1,1))
 1 2 3 4 5 6
6 2 3 3 5 2
> |

```

3. Which one of the following is the correct command to obtain the 3<sup>rd</sup>, 5<sup>th</sup> and 7<sup>th</sup> deciles of a data vector **x** in R?

- a. `decile(x, probs=c(30%,70%,90%))`
- b. `decile(x, probs=(0.30,0.70,0.90))`
- c. `quantile(x, probs=(30%,70%,90%))`
- d. `quantile(x, probs=c(0.30,0.50,0.70))`

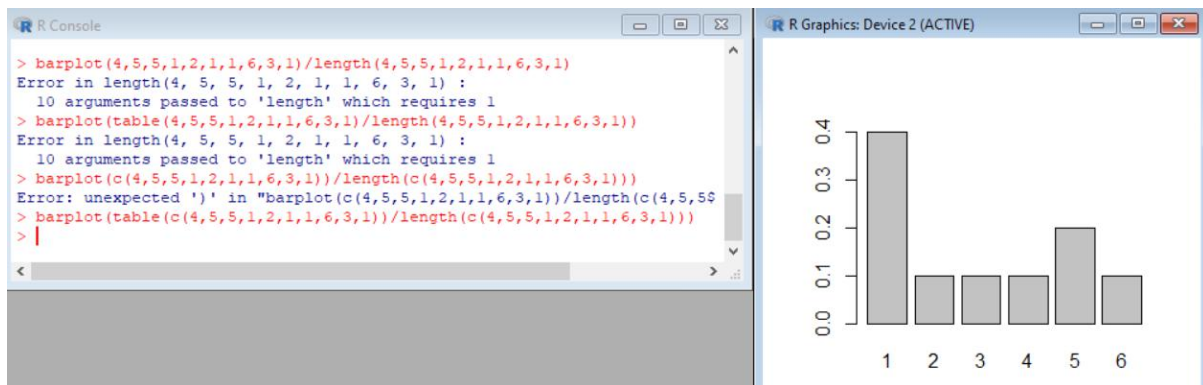
4. Which one of the following correctly specifies the quantiles of a data vector **marks** obtained as an outcome of the command `quantile(marks, probs = seq(0, 1, 0.20))` in R?

- a. 0<sup>th</sup>, 20<sup>th</sup>, 100<sup>th</sup>
- b. 0<sup>th</sup>, 20<sup>th</sup>, 40<sup>th</sup>, 60<sup>th</sup>, 80<sup>th</sup>, 100<sup>th</sup>
- c. 0<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 100<sup>th</sup>
- d. None of these

5. Suppose a die is rolled 10 times and following numbers on the upper face are recorded: 4,5,5,1,2,1,1,6,3,1. Which of the following command in R is used to obtain the bar plot of this data based on relative frequencies?

- a. `barplot(4,5,5,1,2,1,1,6,3,1)/length(4,5,5,1,2,1,1,6,3,1)`
- b. `barplot(table(4,5,5,1,2,1,1,6,3,1)/length(4,5,5,1,2,1,1,6,3,1))`
- c. `barplot(c(4,5,5,1,2,1,1,6,3,1))/length(c(4,5,5,1,2,1,1,6,3,1))`
- d. `barplot(table(c(4,5,5,1,2,1,1,6,3,1))/length(c(4,5,5,1,2,1,1,6,3,1)))`

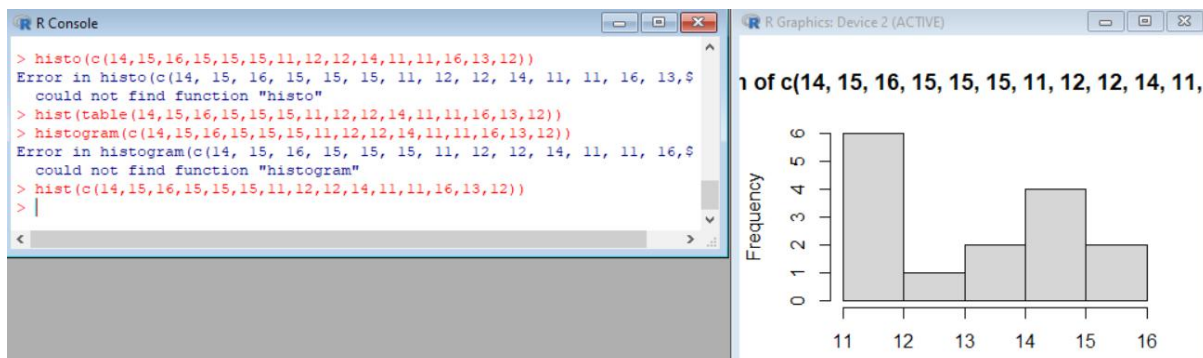
Solution:



6. Suppose the ages of 15 children (in completed years) are recorded as follows: 14,15,16,15,15,15,11,12,12,14,11,11,16,13,12. Which one of the following command is used to obtain the histogram of this data in R based on absolute frequencies?

- a. `histo(c(14,15,16,15,15,15,11,12,12,14,11,11,16,13,12))`
- b. `hist(c(14,15,16,15,15,15,11,12,12,14,11,11,16,13,12))`
- c. `hist(table(14,15,16,15,15,15,11,12,12,14,11,11,16,13,12))`
- d. `histogram(c(14,15,16,15,15,15,11,12,12,14,11,11,16,13,12))`

Solution:



7. Suppose the oil produced every day by 20 different refineries (in thousand liters) is recorded as follows: 820, 184, 921, 488, 721, 614, 801, 396, 864, 845, 404, 781, 457, 1029, 1047, 552, 718, 495, 382, 345. Which one of the following is the correct command to obtain the histogram of this data in R based on relative frequencies?

a.

```
hist(c(820,184,921,488,721,614,801,396,864,845,404,781,457,1029,1047,552,718,495,382,345), freq=FALSE)
```

b.

```
hist(c(820,184,921,488,721,614,801,396,864,845,404,781,457,1029,1047,552,718,495,382,345), freq=TRUE)
```

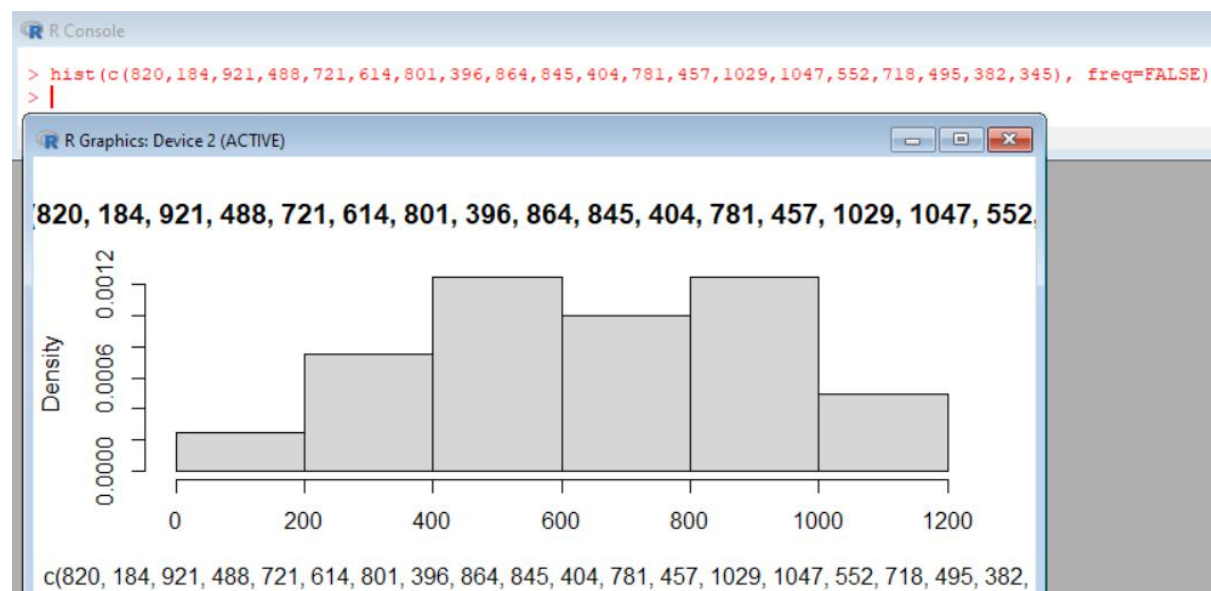
c.

```
hist((820,184,921,488,721,614,801,396,864,845,404,781,457,1029,1047,552,718,495,382,345), freq=FALSE)
```

d.

```
hist(table(820,184,921,488,721,614,801,396,864,845,404,781,457,1029,1047,552,718,495,382,345), freq=FALSE)
```

Solution:



8. Suppose the ages (in completed years) of 12 persons applying for health insurance are recorded and ages of two persons are missing. The collected observations are as follows: 24,35,25,51,NA,18,41,16,43,32,NA,34. Which one of the following is the correct command to obtain the arithmetic mean of this data in R?

- a. `mean(c(24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=TRUE)`
- b. `mean(c(24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=FALSE)`
- c. `mean((24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=TRUE)`
- d. `mean((24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=FALSE)`

Solution:

```
R Console
> mean(c(24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=TRUE)
[1] 31.9
> mean(c(24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=FALSE)
[1] NA
> mean((24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=TRUE)
Error: unexpected ',' in "mean((24,"
> mean((24,35,25,51,NA,18,41,16,43,32,NA,34), na.rm=FALSE)
Error: unexpected ',' in "mean((24,"
> |
```

9. Suppose the number of members in 5 households is recorded but one observation gets missing as follows: 4,5,8,6,NA. Which one of the following is the correct command to obtain the geometric mean of this data in R?

- a. `prod(c(4,5,8,6,NA), na.rm = TRUE)^(1/length(c(4,5,8,6,NA), na.rm = TRUE))`
- b. `prod(c(4,5,8,6,NA), na.rm=TRUE)^(1/length(c(4,5,8,6,NA)))`
- c. `prod(c(4,5,8,6))^(1/length(c(4,5,8,6,NA), na.rm=TRUE))`
- d. None of these

Solution:

```

R Console
> (4*5*8*6)^(1/4)
[1] 5.566315
> prod(c(4,5,8,6,NA), na.rm = TRUE)^(1/length(c(4,5,8,6,NA), na.rm = TRUE))
Error in length(c(4, 5, 8, 6, NA), na.rm = TRUE) :
  2 arguments passed to 'length' which requires 1
> prod(c(4,5,8,6,NA),na.rm=TRUE)^(1/length(c(4,5,8,6,NA)))
[1] 3.948701
> prod(c(4,5,8,6))^(1/length(c(4,5,8,6,NA),na.rm=TRUE))
Error in length(c(4, 5, 8, 6, NA), na.rm = TRUE) :
  2 arguments passed to 'length' which requires 1
> |

```

10. Suppose the daily room rents of 10 hotels in rupees are recorded and two observations are missing as follows: 380,683,NA,456,559,878,NA,897,893,765. Which one of the following is the correct command to obtain the product of this data in R?

- `prod(c(380,683,NA,456,559,878,NA,897,893,765),na.rm=true)`
- `prod((380,683,NA,456,559,878,NA,897,893,765),na.rm=TRUE)`
- `prod(c(380,683,NA,456,559,878,NA,897,893,765), na.rm=TRUE)`
- `product(c(380,683,NA,456,559,878,NA,897,893,765),na.rm=T)`

Solution:

```

R Console
> prod(c(380,683,NA,456,559,878,NA,897,893,765),na.rm=true)
Error: object 'true' not found
> prod((380,683,NA,456,559,878,NA,897,893,765),na.rm=TRUE)
Error: unexpected ',' in "prod((380,"
> product(c(380,683,NA,456,559,878,NA,897,893,765),na.rm=T)
Error in product(c(380, 683, NA, 456, 559, 878, NA, 897, 893, 765), na.rm = T) :
  could not find function "product"
> prod(c(380,683,NA,456,559,878,NA,897,893,765), na.rm=TRUE)
[1] 3.559433e+22
> |

```

11. Suppose the number of students present in 5 class rooms are recorded as follows: 24,15,35,51,42. Which one of the following is the correct command to obtain the harmonic mean of this data in R?

- a. `length(c(24,15,35,51,42))/mean(1/c(24,15,35,51,42))`
- b. `length(c(24,15,35,51,42))/(1/mean(c(24,15,35,51,42)))`
- c. `1/mean(1/(24,15,35,51,42))`
- d. `1/mean(1/c(24,15,35,51,42))`

Solution:

```
R Console
> length(c(24,15,35,51,42))/mean(1/c(24,15,35,51,42))
[1] 138.6408
> length(c(24,15,35,51,42))/(1/mean(c(24,15,35,51,42)))
[1] 167
> 1/mean(1/(24,15,35,51,42))
Error: unexpected ',' in "1/mean(1/(24,"
> 1/mean(1/c(24,15,35,51,42))
[1] 27.72816
> |
```

12. Suppose the per hour salary of salesperson in 10 different stores in a city is recorded and two observations get missing as follows: 45,75,NA,68,69,57,61,NA,52,81. Which one of the following is the correct command to obtain the median of this data in R?

- a. `median(c(45,75,NA,68,69,57,61,NA,52,81),na.rm=true)`
- b. `median(c(45,75,NA,68,69,57,61,NA,52,81), na.rm=TRUE)`
- c. `median((45,75,NA,68,69,57,61,NA,52,81),na.rm=TRUE)`
- d. `median((45,75,NA,68,69,57,61,NA,52,81),na.rm=T)`

Solution:



```

R Console
> median(c(45,75,NA,68,69,57,61,NA,52,81),na.rm=true)
Error in median.default(c(45, 75, NA, 68, 69, 57, 61, NA,
  object 'true' not found
> median((45,75,NA,68,69,57,61,NA,52,81),na.rm=TRUE)
Error: unexpected ',' in "median((45,"
> median((45,75,NA,68,69,57,61,NA,52,81),na.rm=T)
Error: unexpected ',' in "median((45,"
> median(c(45,75,NA,68,69,57,61,NA,52,81), na.rm=TRUE)
[1] 64.5
> |

```

13. Suppose the per day income of 12 persons (in rupees) are recorded and two observations are missing as follows: 14514, 185, 415, 711, 212, NA, 771, 151, 616, NA, 913, 312. Which one of the following is the correct command to obtain the standard deviation of this data in R?

a.

```
sqrt(variance(c(14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=TRUE))
```

b.

```
sqrt(var((14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=TRUE))
```

c.

```
sqrt(var(c(14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=TRUE))
```

d. `sqrt(var(c(14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=true))`

Solution:

```

R Console
> sqrt(variance(c(14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=TRUE))
Error in variance(c(14514, 185, 415, 711, 212, NA, 771, 151, 616, NA, :
could not find function "variance"
> sqrt(var((14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=TRUE))
Error: unexpected ',' in "sqrt(var((14514,"
> sqrt(var(c(14514,185,415,711,212,NA,771,151,616,NA,913,312),na.rm=true))
Error in var(c(14514, 185, 415, 711, 212, NA, 771, 151, 616, NA, 913, :
object 'true' not found
> sqrt(var(c(14514,185,415,711,212,NA,771,151,616,NA,913,312), na.rm=TRUE))
[1] 4447.18
> |

```

14. Suppose the cost of servicing of 5 brands of cars (in thousand rupees) is recorded but one value gets missing as follows: 4,3,6,7,NA. Which one of the following is the correct command to obtain the range of this data in R:

- `range(c(4,3,6,7,NA), na.rm=TRUE)`
- `maximum(c(4,3,6,7,NA), na.rm=TRUE)-minimum(c(4,3,6,7,NA), na.rm=TRUE)`
- `max(c(4,3,6,7,NA), na.rm=TRUE)-min(c(4,3,6,7,NA), na.rm=TRUE)`
- `max[c(4,3,6,7,NA), na.rm=TRUE]-min[c(4,3,6,7,NA), na.rm=TRUE]`

Solution:

```

R Console
> range(c(4,3,6,7,NA), na.rm=TRUE)
[1] 3 7
> maximum(c(4,3,6,7,NA), na.rm=TRUE)-minimum(c(4,3,6,7,NA), na.rm=TRUE)
Error in maximum(c(4, 3, 6, 7, NA), na.rm = TRUE) :
could not find function "maximum"
> max[c(4,3,6,7,NA), na.rm=TRUE]-min[c(4,3,6,7,NA), na.rm=TRUE]
Error in max[c(4, 3, 6, 7, NA), na.rm = TRUE] :
object of type 'builtin' is not subsettable
> max(c(4,3,6,7,NA), na.rm=TRUE)-min(c(4,3,6,7,NA), na.rm=TRUE)
[1] 4
> |

```

15. Suppose the number of songs played on 10 channels of a radio in 3 hours time are recorded and two observations are missing as follows: 14,15,NA,11,12,11,11,16,13,NA. Which one of the following is the correct command to obtain the interquartile range of this data in R?

- a. `IQRRange(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)`
- b. `QR(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)`
- c. `IQR[c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE]`
- d. `IQR(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)`

Solution:

```
R Console
> IQRRange(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)
Error in IQRRange(c(14, 15, NA, 11, 12, 11, 11, 16, 13, NA)
  could not find function "IQRRange"
> QR(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)
Error in QR(c(14, 15, NA, 11, 12, 11, 11, 16, 13, NA), na.
  could not find function "QR"
> IQR[c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE]
Error in IQR[c(14, 15, NA, 11, 12, 11, 11, 16, 13, NA), na
  object of type 'closure' is not subsettable
> IQR(c(14,15,NA,11,12,11,11,16,13,NA),na.rm=TRUE)
[1] 3.25
> |
```

16. Suppose the ages of 5 cars in completed years are recorded as follows: 14,15,15,11,12. Which one of the following is the correct command to obtain the

mean absolute deviation ( $MD(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$ ) of this data in R?

- a. `sum(abs((14,15,15,11,12)) - mean((14,15,15,11,12)))/length((14,15,15,11,12))`
- b. `sum(abs(c(14,15,15,11,12)) - mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))`
- c. `sum(absolute(c(14,15,15,11,12)) - mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))`

d. `mean(abs(c(14,15,15,11,12))-  
mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))`

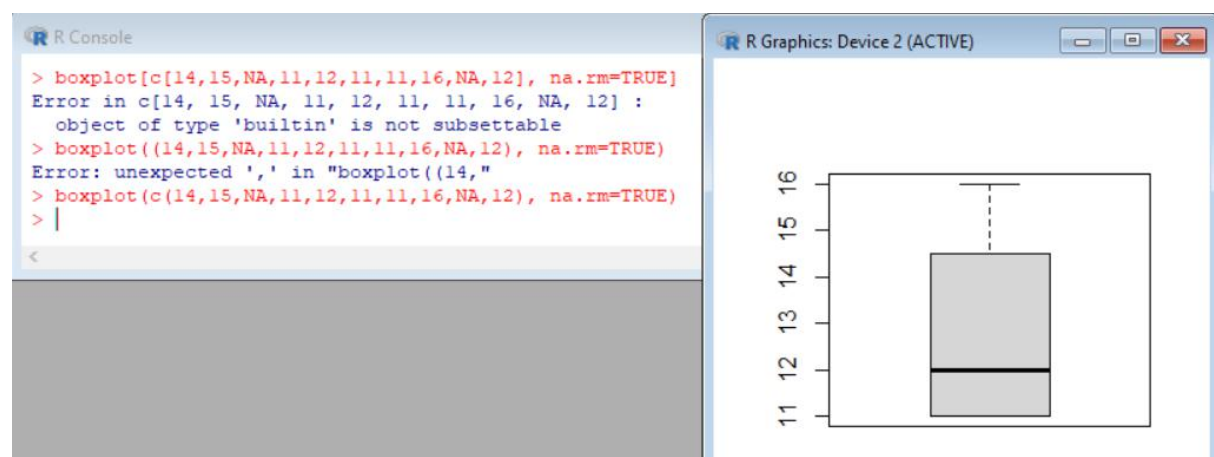
Solution:

```
R Console
> sum(abs((14,15,15,11,12))-mean((14,15,15,11,12)))/length((14,15,15,11,12))
Error: unexpected ',' in "sum(abs((14,"
> sum(absolute(c(14,15,15,11,12))-mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))
Error in absolute(c(14, 15, 15, 11, 12) - mean(c(14, 15, 15, 11, 12))) :
  could not find function "absolute"
> mean(abs(c(14,15,15,11,12))-mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))
[1] 0.304
> sum(abs(c(14,15,15,11,12))-mean(c(14,15,15,11,12)))/length(c(14,15,15,11,12))
[1] 1.52
> |
```

17. Suppose the number of songs played on 10 channels of a radio in 3 hours time are recorded and two observations are missing as follows: 14,15,NA,11,12,11,11,16,NA,12. Which one of the following is the correct command to obtain the boxplot of this data in R?

- a. `boxplot[c[14,15,NA,11,12,11,11,16,NA,12], na.rm=TRUE]`
- b. `boxplot((14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)`
- c. `boxplot(c(14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)`
- d. None of these

Solution:



18. Suppose the number of songs played on 10 channels of a radio in 3 hours time are counted and two observations are missing as follows: 14,15,NA,11,12,11,11,16,NA,12. Which one of the following is the correct command to obtain the coefficient of skewness of this data in R?

- a. `skewness[c(14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE]`
- b. `skewness(c(14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)`
- c. `skewness((14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)`
- d. None of these

Solution:

```
R Console
> library(moments)
> skewness[c(14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE]
Error in c[14, 15, NA, 11, 12, 11, 11, 16, NA, 12] :
  object of type 'builtin' is not subsettable
> skewness((14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)
Error: unexpected ',' in "skewness((14,"
> skewness(c(14,15,NA,11,12,11,11,16,NA,12), na.rm=TRUE)
[1] 0.6031024
> |
```

19. Suppose the share rates (in rupees) of 10 companies on a day in a stock exchange are recorded as follows: 514,615,185,191,152,311,116,168,163,512. Which one of the following is the correct command to obtain the coefficient of kurtosis in R?

- a. `kurtosis(c(514,615,185,191,152,311,116,168,163,512))`
- b. `kurt(c(514,615,185,191,152,311,116,168,163,512))`
- c. `kurt(514,615,185,191,152,311,116,168,163,512)`
- d. `kurtosis[c(514,615,185,191,152,311,116,168,163,512)]`

Solution:

```
R Console
> kurt(c(514,615,185,191,152,311,116,168,163,512))
Error in kurt(c(514, 615, 185, 191, 152, 311, 116, 168, 163, 512)) :
  could not find function "kurt"
> kurt(514,615,185,191,152,311,116,168,163,512)
Error in kurt(514, 615, 185, 191, 152, 311, 116, 168, 163, 512) :
  could not find function "kurt"
> kurtosis(c(514,615,185,191,152,311,116,168,163,512))
Error in kurtosis(c(514, 615, 185, 191, 152, 311, 116, 168, 163, 512)) :
  object of type 'closure' is not subsettable
> kurtosis(c(514,615,185,191,152,311,116,168,163,512))
[1] 1.875844
> |
```

20. Suppose the share rates (in rupees) of 10 companies on a day in a stock exchange are recorded and one observations gets missing as follows: 514,615,185,191,152,311,NA,168,163,512. Which one of the following correctly specify the command to obtain the coefficient of kurtosis of this data in R and its obtained value?

a. `kurtosis(c(514,615,185,191,152,311,NA,168,163,512), na.rm=FALSE)` , 1.668406

b. `kurtosis(c(514,615,185,191,152,311,NA,168,163,512), na.rm=TRUE)` , NA

c. `kurtosis(c(514,615,185,191,152,311,NA,168,163,512), na.rm=TRUE)` , 1.668406

d. None of these

Solution:

```
R Console
> kurtosis(c(514,615,185,191,152,311,NA,168,163,512) na.rm=FALSE)
Error: unexpected symbol in "kurtosis(c(514,615,185,191,152,311,NA,168,163,512) na.rm=FALSE)"
> kurtosis(c(514,615,185,191,152,311,NA,168,163,512) na.rm=TRUE)
Error: unexpected symbol in "kurtosis(c(514,615,185,191,152,311,NA,168,163,512) na.rm=TRUE)"
> kurtosis(c(514,615,185,191,152,311,NA,168,163,512) , na.rm=TRUE)
[1] 1.668406
> |
```

## **MOOC Course - Introduction to R Software**

### **Answers of Assignment 7**

1. a
2. c
3. d
4. b
5. d
6. b
7. a
8. a
9. d
10. c
11. d
12. b
13. c
14. c
15. d
16. b
17. c
18. b
19. a
20. c