

ENGINEERING DATA ANALYSIS DOCUMENTATION

TOPIC: DATA MANIPULATION (BONUS ASSESSMENT 4)	COURSE AND SECTION: BSCOE 1-1
PROFESSOR: PERLYN DILLA	DATE: MAY 7, 2022

GROUP MEMBERS:

**BUHAIN, JASON
CABATE, REIVEN
DELFIN, DANIELLA
RAMOS, ARVELYN
VIERNES, MICHAEL**

RStudio interface showing the R script for the programming exercises. The script is as follows:

```

1 library(readxl)
2 SAMPLE_DATA <- read_excel("SAMPLE_DATA.xlsx")
3 library(dplyr)
4
5 # I. From SAMPLE_DATA, create a data with file name NEWDATA with the following conditions:
6 # 1. Randomly take 30% of the number of observations.
7
8 NEWDATA <- SAMPLE_DATA %>% sample_frac(0.3, replace=F) %>%
9
10 # 2. Include only those students whose favorite color are either red, blue, or yellow.
11 filter(FAVORITE_COLOR %in% c('RED','YELLOW','BLUE')) %>%
12
13 # 3. Select only of the following columns:
14 # Student Number
15 # Favorite Subject
16 # Grade in 4 subjects
17 select(-c(GENDER,AGE,HEIGHT,WEIGHT,FAVORITE_COLOR)) %>%
18
19 # 4. Rename all the column names so that the first letter is capitalized.
20 rename(Respondent_Number="RESPONDENT_NUMBER",
21        Favorite_Subject="FAVORITE_SUBJECT",
22        Grade_in_Math="GRADE_IN_MATH",
23        Grade_in_Statistics_and_Probability="GRADE_IN_STATISTICS_AND_PROBABILITY",
24        Grade_in_Science="GRADE_IN_SCIENCE",
25        Grade_in_Media_and_Information_Literacy="GRADE_IN_MEDIA_AND_INFORMATION_LITERACY") %>%
26
27 # 5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects.
28 mutate(Average=(Grade_in_Math + Grade_in_Statistics_and_Probability + Grade_in_Science
29               + Grade_in_Media_and_Information_Literacy)/4)
30 na.omit(NEWDATA)
31 # II. Summarize the data to find out the mean and standard deviation of the 'Average' column.
32 # Summarize the data according to favorite subject.
33 summary1 <- NEWDATA %>% group_by(Favorite_Subject) %>% summarise(mean(Average, na.rm=T),
34                                                                    sd(Average, na.rm=T))
35
36
37

```

The console output shows the following data:

Respondent_Number	Favorite_Subject	Grade_in_Math	Grade_in_Statistics_and_Probability	Grade_in_Science	Grade_in_Media_and_Information_Literacy	Average
6	SCIENCE	90	94	93	98	93.8
19	MATH	94	91	92	92	92.2
22	ENGLISH	85		95	90	90.8

The RStudio interface also shows the Environment pane with the following data:

Object	Class	Size
NEWDATA	data.frame	7 obs. of 7 ...
SAMPLE...	data.frame	22 obs. of 1...
summa...	data.frame	2 obs. of 3 ...

• SOURCE CODE FOR THE PROGRAMMING EXERCISES

Assessment:

- From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 - Randomly take 30% of the number of observations. (2 points)
 - Include only those students whose favorite color are either red, blue or yellow. (2 points)
 - Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 - Rename all the column names so that the first letter is capitalised. (2 points)
 - Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Bonus-Assessment-4.r NEWDATA summary1 SAMPLE_DATA

Filter

IT_NUMBER	GENDER	AGE	FAVORITE_SUBJECT	FAVORITE_COLOR	HEIGHT	WEIGHT	GRADE_IN_MATH	GRADE_IN_STATISTICS_AND_PROBABILITY	GRADE_IN_SCIENCE	GRADE_IN_MEDIA_AND_INFORMATION_LITERACY
1	MALE	19	SCIENCE	BLUE	171	73	89	89	90	92
2	FEMALE	19	SCIENCE	YELLOW	154	48	95	94	91	97
3	MALE	19	SCIENCE	BLUE	165	60	NA	93	90	95
4	MALE	19	SCIENCE	BLUE	165	45	94	97	94	96
5	MALE	19	SCIENCE	RED	166	51	95	89	94	91
6	MALE	19	SCIENCE	BLUE	160	53	90	94	93	98
7	MALE	20	ENGLISH	YELLOW	162	60	85	87	85	93
8	MALE	19	SCIENCE	YELLOW	167	45	89	89	88	89
9	MALE	18	ENGLISH	BLUE	169	48	83	87	85	85
10	FEMALE	19	ENGLISH	RED	162	54	93	94	94	93
11	MALE	19	SCIENCE	BLUE	175	65	88	91	93	97
12	MALE	18	ENGLISH	GREEN	161	65	85	87	89	96
13	MALE	19	SCIENCE	BLUE	172	91	94	93	92	90
14	MALE	18	ENGLISH	BLUE	175	68	96	95	96	95
15	MALE	20	SCIENCE	YELLOW	173	75	87	95	90	97
16	MALE	20	MATH	BLUE	174	49	92	94	93	92
17	FEMALE	19	SCIENCE	BLUE	152	60	94	94	92	94
18	MALE	18	ENGLISH	BLUE	175	68	96	95	96	95
19	MALE	19	MATH	YELLOW	175	57	94	91	92	92
20	MALE	18	MATH	BLUE	176	58	92	95	90	94
21	FEMALE	18	MATH	BLUE	158	49	83	91	86	89
22	FEMALE	18	ENGLISH	YELLOW	156	43	85	88	95	95

Showing 1 to 22 of 22 entries, 11 total columns

Console Terminal Jobs

```
R 4.1.3 ~ /
> source("D:/Codings/R/Bonus-Assessment-4.r")
Error: 'path' does not exist: 'SAMPLE_DATA.xlsx'
> source("D:/Codings/R/Bonus-Assessment-4.r")
> source("D:/Codings/R/Bonus-Assessment-4.r")
> |
```

meet.google.com is sharing your screen. Stop sharing Hide

Type here to search

34°C Partly sunny 1:14 PM 5/7/2022

THE SAMPLE DATA

Assessment:

I. From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:

1. Randomly take 30% of the number of observations. (2 points)
2. Include only those students whose favorite color are either red, blue or yellow. (2 points)
3. Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
4. Rename all the column names so that the first letter is capitalised. (2 points)
5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)

	RESPONDENT_NUMBER	GENDER	AGE	FAVORITE_SUBJECT	FAVORITE_COLOR	HEIGHT	WEIGHT	GRADE_IN_MATH
1	8	MALE	19	SCIENCE	YELLOW	167	45	89
2	2	FEMALE	19	SCIENCE	YELLOW	154	48	95
3	18	MALE	18	ENGLISH	BLUE	175	68	96
4	16	MALE	20	MATH	BLUE	174	49	92
5	10	FEMALE	19	ENGLISH	RED	162	54	93
6	4	MALE	19	SCIENCE	BLUE	165	45	94
7	20	MALE	18	MATH	BLUE	176	58	92

1. Randomly take 30% of the number of observations.

Assessment:

- I. From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 1. Randomly take 30% of the number of observations. (2 points)
 2. Include only those students whose favorite color are either red, blue or yellow. (2 points)
 3. Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 4. Rename all the column names so that the first letter is capitalised. (2 points)
 5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)



RGui (64-bit) - [Data: NEWDATA2]

File

	RESPONDENT_NUMBER	GENDER	AGE	FAVORITE_SUBJECT	FAVORITE_COLOR	HEIGHT	WEIGHT	GRADE_IN_MATH
1	8	MALE	19	SCIENCE	YELLOW	167	45	89
2	2	FEMALE	19	SCIENCE	YELLOW	154	48	95
3	18	MALE	18	ENGLISH	BLUE	175	68	96
4	16	MALE	20	MATH	BLUE	174	49	92
5	10	FEMALE	19	ENGLISH	RED	162	54	93
6	4	MALE	19	SCIENCE	BLUE	165	45	94
7	20	MALE	18	MATH	BLUE	176	58	92

Windows taskbar: 9:57 PM

2. Include only those students whose favorite color are either red, blue or yellow.

Assessment:

- I. From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 1. Randomly take 30% of the number of observations. (2 points)
 2. Include only those students whose favorite color are either red, blue or yellow. (2 points)
 3. Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 4. Rename all the column names so that the first letter is capitalised. (2 points)
 5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)



RGui (64-bit) - [Data: NEWDATA3]

File

	RESPONDENT_NUMBER	FAVORITE_SUBJECT	GRADE_IN_MATH	GRADE_IN_STATISTICS_AND_PROBABILITY	GRADE_IN_SCIENCE
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90

9:57 PM

RGui (64-bit) - [Data: NEWDATA4]

File

	Respondent_Number	Favorite_Subject	Grade_in_Math	Grade_in_Statistics_and_Probability	Grade_in_Science
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90

Windows taskbar: 9:58 PM

3. Select only the following columns:

Student Number

Favorite Subject

Grade in 4 subjects

Assessment:



- I. From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 1. Randomly take 30% of the number of observations. (2 points)
 2. Include only those students whose favorite color are either red, blue or yellow. (2 points)
 3. Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 4. Rename all the column names so that the first letter is capitalised. (2 points)
 5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)

RGui (64-bit) - [Data: NEWDATA5]

File

	Respondent_Number	Favorite_Subject	Grade_in_Math	Grade_in_Statistics_and_Probability	Grade_in_Science
1	8	SCIENCE	89	89	88
2	2	SCIENCE	95	94	91
3	18	ENGLISH	96	95	96
4	16	MATH	92	94	93
5	10	ENGLISH	93	94	94
6	4	SCIENCE	94	97	94
7	20	MATH	92	95	90

Windows taskbar: 9:58 PM

4. Rename all the column names so that the first letter is capitalized.

Assessment:

- I. From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 1. Randomly take 30% of the number of observations. (2 points)
 2. Include only those students whose favorite color are either red, blue or yellow. (2 points)
 3. Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 4. Rename all the column names so that the first letter is capitalised. (2 points)
 5. Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)



RGui (64-bit) - [Data: summary]

File

	Favorite_Subject	mean(Average, na.rm = T)	sd(Average, na.rm = T)
1	ENGLISH	94.50	1.414214
2	MATH	92.75	0.000000
3	SCIENCE	92.75	3.500000

9:58 PM

**Create a new column and name it as Average.
Calculate it as the average of the 4 subjects.**

Assessment:

- From SAMPLE DATA, create a data with file name NEWDATA with the following conditions:
 - Randomly take 30% of the number of observations. (2 points)
 - Include only those students whose favorite color are either red, blue or yellow. (2 points)
 - Select only the following columns: (2 points)
 - Student Number
 - Favorite Subject
 - Grade in 4 subjects
 - Rename all the column names so that the first letter is capitalised. (2 points)
 - Create a new column and name it as Average. Calculate it as the average of the 4 subjects. (2 points)