

MVLU COLLEGE PRACTICAL NO.3

AIM: Exploring data: View() or print() (R).

The top screenshot shows RStudio with a data frame named 'Cleaned_BMW_Sales_Data' containing 99 observations and 11 variables. The variables are: model, year, region, color, fuel_type, transmission, engine_size_l, mileage_km, price_usd, sales_volume, and sales_classification. The data is displayed in a table format.

The bottom screenshot shows an R script with the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5 my_data <- read.csv("C:/Users/itlab/OneDrive/Documents/S105/Cleaned_BMW_Sales_Data.csv")
6 head(my_data)
7 tail(my_data)
8 dim(my_data)
9 cat("Dimensions (Rows, columns): ", dim(my_data), "\n")
10 str(my_data)
11 summary(my_data)
12 names(my_data)
13 cat("Column Names: ", names(my_data), "\n")
14 describe(my_data)
15
```

The console shows an error message: "Error in file(file, "rt") : cannot open the connection". Below the error, a warning message states: "In addition: warning message: In file(file, "rt") : cannot open file 'C:/Users/itlab/OneDrive/Documents/S105/Cleaned_BMW_Sales_Data': no such file or directory".

The Environment pane on the right shows the following data objects:

- Cleaned_BMW_Sales_Data: 99 obs. of 11 variables
- College_Marks_Dataset: 1000 obs. of 8 variables
- my_data: 99 obs. of 11 variables

MVLU COLLEGE

PRACTICAL NO.3

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source Console Background Jobs
R - R 4.5.2 - ~/
> my_data <- read.csv("C:/Users/itlab/OneDrive/Documents/S105/Cleaned_BMW_Sales_Data.csv")
> head(my_data)
  model year region color fuel_type transmission engine_size_l mileage_km price_usd sales_volume
1 5 Series 2016 Asia Red Petrol Manual 3.5 151748 98740 8300
2 18 2013 North America Red Hybrid Automatic 1.6 121671 79219 3428
3 5 Series 2022 North America Blue Petrol Automatic 4.5 10991 113265 6994
4 X3 2024 Middle East Blue Petrol Automatic 1.7 27255 60971 4047
5 7 Series 2020 South America Black Diesel Manual 2.1 122131 49898 3080
6 5 Series 2017 Middle East Silver Diesel Manual 1.9 171362 42926 1232
sales_classification
1 High
2 Low
3 Low
4 Low
5 Low
6 Low
> tail(my_data)
  model year region color fuel_type transmission engine_size_l mileage_km price_usd sales_volume
94 13 2013 Africa Black Diesel Manual 1.6 192509 72385 455
95 13 2015 Asia Silver Electric Manual 3.5 68161 63696 7218
96 X6 2022 Asia white Electric Manual 1.9 182027 35951 3389
97 X5 2024 Asia Blue Hybrid Automatic 2.6 120597 95900 8581
98 18 2011 Europe Red Hybrid Automatic 3.8 184765 44846 1431
99 X5 2019 Africa Blue Electric Manual 2.1 161177 95908 2775
sales_classification
94 Low
95 High
96 Low
97 High
98 Low
99 Low
> dim(my_data)
[1] 99 11
> cat("Dimensions (Rows, Columns): ", dim(my_data), "\n")
Dimensions (Rows, Columns): 99 11
> str(my_data)
'data.frame': 99 obs. of 11 variables:
 $ model : chr "5 Series" "18" "5 Series" "X3" ...
 $ year : int 2016 2013 2022 2024 2020 2017 2022 2014 2016 2019 ...
 $ region : chr "Asia" "North America" "North America" "Middle East" ...

Environment History Connections Tutorial
R - Global Environment
Data
Cleaned_BMW_Sales_Data 99 obs. of 11 variables
college_Marks_Dataset 1000 obs. of 8 variables
my_data 99 obs. of 11 variables

Files Plots Packages Help Viewer Presentation
33°C Sunny
Search
ENG IN
12:29 24-11-2025
```

MVLU COLLEGE PRACTICAL NO.3

The image displays two screenshots of the RStudio interface, showing the analysis of a car dataset. The top screenshot shows the initial data structure and summary statistics, while the bottom screenshot shows the same data with additional kurtosis statistics calculated for each variable.

Top Screenshot: Initial Data Structure and Summary Statistics

```
> str(my_data)
'data.frame':   99 obs. of  11 variables:
 $ model       : chr  "5 Series" "i8" "5 Series" "X3" ...
 $ year        : int   2016 2013 2022 2024 2020 2017 2022 2014 2016 2019 ...
 $ region      : chr  "Asia" "North America" "North America" "Middle East" ...
 $ color       : chr  "Red" "Red" "Blue" "Blue" ...
 $ fuel_type   : chr  "Petrol" "Hybrid" "Petrol" "Petrol" ...
 $ transmission : chr  "Manual" "Automatic" "Automatic" "Automatic" ...
 $ engine_size_l : num  3.5 1.6 4.5 1.7 2.1 1.9 1.8 1.6 1.7 3 ...
 $ mileage_km  : int   151748 121671 10991 27255 122131 171362 196741 121156 48073 35700 ...
 $ price_usd   : int   98740 79219 113265 60971 49898 42926 55064 102778 116482 96257 ...
 $ sales_volume : int   8300 3428 6994 4047 3080 1232 7949 632 8944 4411 ...
 $ sales_classification : chr  "High" "Low" "Low" "Low" ...

> summary(my_data)
   model      year      region      color      fuel_type      transmission
Length:99   Min.   :2010   Length:99   Length:99   Length:99   Length:99
Class :character 1st Qu.:2014   Class :character 1st Qu.:2014   Class :character 1st Qu.:2014
Mode :character  Median:2017   Mode :character  Median:2017   Mode :character  Median:2017
                Mean  :2017                Mean  :2017                Mean  :2017
                3rd Qu.:2021                3rd Qu.:2021                3rd Qu.:2021
                Max.   :2024                Max.   :2024                Max.   :2024

 engine_size_l  mileage_km  price_usd  sales_volume  sales_classification
Min.   :1.500   Min.   :1428   Min.   :31815   Min.   :173   Length:99
1st Qu.:2.550   1st Qu.: 61221 1st Qu.: 40200   1st Qu.:2526   Class :character
Median :3.200   Median:105162 Median : 82675   Median:4823   Mode  :character
Mean   :3.195   Mean :103617  Mean : 79109   Mean :5029
3rd Qu.:3.800   3rd Qu.:142628 3rd Qu.:101498 3rd Qu.:7793
Max.   :4.900   Max. :196741  Max. :119927  Max. : 9797

> cat("Column Names: ", names(my_data), "\n")
Column Names: model year region color fuel_type transmission engine_size_l mileage_km price_usd sales_volume sales_classification

> describe(my_data)
  vars n  mean  sd median trimmed mad min  max range skew
model*  1 99   6.34   2.94    6.0    6.38   2.97  1.0  11.0  10.0 -0.06
year    2 99  2017.23  4.26  2017.0  2017.23  5.93 2010.0 2024.0 14.0 -0.01
region*  3 99   3.75   1.64    4.0    3.80   1.48  1.0   6.0   5.0 -0.16
color*   4 99   3.55   1.76    4.0    3.56   1.48  1.0   6.0   5.0 -0.11
fuel_type* 5 99   2.51   1.08    3.0    2.51   1.48  1.0   4.0   3.0 -0.04
transmission* 6 99   1.55   0.50    2.0    1.56   0.00  1.0   2.0   1.0 -0.18
engine_size_l 7 99   3.19   0.92    3.2    3.19   0.89  1.5   4.9   3.4 -0.05
mileage_km  8 99 103617.16 53893.14 105162.0 103746.31 57410.72 1428.0 196741.0 195313.0 -0.03
price_usd   9 99 79109.01 27034.89 82675.0 79609.12 29804.71 31815.0 119927.0 88112.0 -0.13
sales_volume 10 99 5028.74 2918.70 4823.0 5028.46 3955.58 173.0 9797.0 9624.0 0.00
sales_classification* 11 99   1.67   0.47    2.0    1.70   0.00  1.0   2.0   1.0 -0.70

model*      kurtosis  se
year        -1.10   0.30
region*     -1.28   0.43
color*      -1.23   0.16
fuel_type*  -1.28   0.18
transmission* -1.30   0.11
engine_size_l -0.91   0.09
mileage_km  -0.89 5416.46
price_usd   -1.27 2717.11
sales_volume -1.30 293.34
sales_classification* -1.53 0.05
```

Bottom Screenshot: Data Structure and Summary Statistics (Including Kurtosis)

```
> str(my_data)
'data.frame':   99 obs. of  11 variables:
 $ model       : chr  "5 Series" "i8" "5 Series" "X3" ...
 $ year        : int   2016 2013 2022 2024 2020 2017 2022 2014 2016 2019 ...
 $ region      : chr  "Asia" "North America" "North America" "Middle East" ...
 $ color       : chr  "Red" "Red" "Blue" "Blue" ...
 $ fuel_type   : chr  "Petrol" "Hybrid" "Petrol" "Petrol" ...
 $ transmission : chr  "Manual" "Automatic" "Automatic" "Automatic" ...
 $ engine_size_l : num  3.5 1.6 4.5 1.7 2.1 1.9 1.8 1.6 1.7 3 ...
 $ mileage_km  : int   151748 121671 10991 27255 122131 171362 196741 121156 48073 35700 ...
 $ price_usd   : int   98740 79219 113265 60971 49898 42926 55064 102778 116482 96257 ...
 $ sales_volume : int   8300 3428 6994 4047 3080 1232 7949 632 8944 4411 ...
 $ sales_classification : chr  "High" "Low" "Low" "Low" ...

> summary(my_data)
   model      year      region      color      fuel_type      transmission
Length:99   Min.   :2010   Length:99   Length:99   Length:99   Length:99
Class :character 1st Qu.:2014   Class :character 1st Qu.:2014   Class :character 1st Qu.:2014
Mode :character  Median:2017   Mode :character  Median:2017   Mode :character  Median:2017
                Mean  :2017                Mean  :2017                Mean  :2017
                3rd Qu.:2021                3rd Qu.:2021                3rd Qu.:2021
                Max.   :2024                Max.   :2024                Max.   :2024

 engine_size_l  mileage_km  price_usd  sales_volume  sales_classification
Min.   :1.500   Min.   :1428   Min.   :31815   Min.   :173   Length:99
1st Qu.:2.550   1st Qu.: 61221 1st Qu.: 40200   1st Qu.:2526   Class :character
Median :3.200   Median:105162 Median : 82675   Median:4823   Mode  :character
Mean   :3.195   Mean :103617  Mean : 79109   Mean :5029
3rd Qu.:3.800   3rd Qu.:142628 3rd Qu.:101498 3rd Qu.:7793
Max.   :4.900   Max. :196741  Max. :119927  Max. : 9797

> cat("Column Names: ", names(my_data), "\n")
Column Names: model year region color fuel_type transmission engine_size_l mileage_km price_usd sales_volume sales_classification

> describe(my_data)
  vars n  mean  sd median trimmed mad min  max range skew
model*  1 99   6.34   2.94    6.0    6.38   2.97  1.0  11.0  10.0 -0.06
year    2 99  2017.23  4.26  2017.0  2017.23  5.93 2010.0 2024.0 14.0 -0.01
region*  3 99   3.75   1.64    4.0    3.80   1.48  1.0   6.0   5.0 -0.16
color*   4 99   3.55   1.76    4.0    3.56   1.48  1.0   6.0   5.0 -0.11
fuel_type* 5 99   2.51   1.08    3.0    2.51   1.48  1.0   4.0   3.0 -0.04
transmission* 6 99   1.55   0.50    2.0    1.56   0.00  1.0   2.0   1.0 -0.18
engine_size_l 7 99   3.19   0.92    3.2    3.19   0.89  1.5   4.9   3.4 -0.05
mileage_km  8 99 103617.16 53893.14 105162.0 103746.31 57410.72 1428.0 196741.0 195313.0 -0.03
price_usd   9 99 79109.01 27034.89 82675.0 79609.12 29804.71 31815.0 119927.0 88112.0 -0.13
sales_volume 10 99 5028.74 2918.70 4823.0 5028.46 3955.58 173.0 9797.0 9624.0 0.00
sales_classification* 11 99   1.67   0.47    2.0    1.70   0.00  1.0   2.0   1.0 -0.70

model*      kurtosis  se
year        -1.10   0.30
region*     -1.28   0.43
color*      -1.23   0.16
fuel_type*  -1.28   0.18
transmission* -1.30   0.11
engine_size_l -0.91   0.09
mileage_km  -0.89 5416.46
price_usd   -1.27 2717.11
sales_volume -1.30 293.34
sales_classification* -1.53 0.05
```