

RWorksheet_BIBIT#3b

John Benedict Bibit

2024-10-13

1. Create a data frame using the table below.

a. Write the codes.

```
respondents_data <- data.frame(  
  Respondents = 1:20,  
  Sex = c(2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2),  
  Fathers_Occupation = c(1, 3, 3, 3, 1, 2, 3, 1, 1, 1, 3, 2, 1, 3, 3, 1, 3, 1, 2, 1),  
  Persons_at_Home = c(5, 7, 3, 8, 5, 9, 6, 7, 8, 4, 7, 5, 4, 7, 8, 8, 3, 11, 7, 6),  
  Siblings_at_School = c(6, 4, 4, 1, 2, 1, 5, 3, 1, 2, 3, 2, 5, 5, 2, 1, 2, 5, 3, 2),  
  Types_of_Houses = c(1, 2, 3, 1, 1, 3, 3, 1, 2, 3, 2, 3, 2, 2, 3, 3, 3, 3, 3, 2)  
)
```

b. Describe the data. Get the structure or the summary of the data

```
str(respondents_data)
```

```
## 'data.frame':    20 obs. of  6 variables:  
## $ Respondents      : int  1 2 3 4 5 6 7 8 9 10 ...  
## $ Sex              : num  2 2 1 2 2 2 2 2 2 2 ...  
## $ Fathers_Occupation: num  1 3 3 3 1 2 3 1 1 1 ...  
## $ Persons_at_Home   : num  5 7 3 8 5 9 6 7 8 4 ...  
## $ Siblings_at_School: num  6 4 4 1 2 1 5 3 1 2 ...  
## $ Types_of_Houses   : num  1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(respondents_data)
```

```
##   Respondents      Sex      Fathers_Occupation Persons_at_Home  
## Min.   : 1.00   Min.   :1.00   Min.   :1.00   Min.   : 3.0  
## 1st Qu.: 5.75   1st Qu.:2.00   1st Qu.:1.00   1st Qu.: 5.0  
## Median :10.50   Median :2.00   Median :2.00   Median : 7.0  
## Mean   :10.50   Mean   :1.85   Mean   :1.95   Mean   : 6.4  
## 3rd Qu.:15.25   3rd Qu.:2.00   3rd Qu.:3.00   3rd Qu.: 8.0  
## Max.   :20.00   Max.   :2.00   Max.   :3.00   Max.   :11.0  
## Siblings_at_School Types_of_Houses  
## Min.   :1.00   Min.   :1.0  
## 1st Qu.:2.00   1st Qu.:2.0  
## Median :2.50   Median :2.5  
## Mean   :2.95   Mean   :2.3  
## 3rd Qu.:4.25   3rd Qu.:3.0  
## Max.   :6.00   Max.   :3.0
```

c. Is the mean number of siblings attending is 5?

```
mean_siblings <- mean(respondents_data$Siblings_at_School)
print(paste("Mean number of siblings attending school:", mean_siblings))
```

```
## [1] "Mean number of siblings attending school: 2.95"
```

```
print(paste("Is the mean number of siblings attending 5?", mean_siblings == 5))
```

```
## [1] "Is the mean number of siblings attending 5? FALSE"
```

d. Extract the 1st two rows and then all the columns using the subsetting functions. Write the codes and its output.

```
first_two_rows <- respondents_data[1:2, ]
print(first_two_rows)
```

```
## Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1          1 2          1          5          6
## 2          2 2          3          7          4
## Types_of_Houses
## 1          1
## 2          2
```

e. Extract 3rd and 5th row with 2nd and 4th column. Write the codes and its result.

```
selected_rows_cols <- respondents_data[c(3,5), c(2,4)]
print(selected_rows_cols)
```

```
## Sex Persons_at_Home
## 3 1          3
## 5 2          5
```

f. Select the variable types of houses then store the vector that results as types_houses. Write the codes.

```
types_houses <- respondents_data$Types_of_Houses
print(types_houses)
```

```
## [1] 1 2 3 1 1 3 3 1 2 3 2 3 2 2 3 3 3 3 3 2
```

g. Select only all Males respondent that their father occupation was farmer. Write the codes and its output.

```
male_farmer <- respondents_data[respondents_data$Sex == 1 & respondents_data$Fathers_Occupation == 1, ]
print(male_farmer)
```

```
## [1] Respondents Sex Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

- h. Select only all females respondent that have greater than or equal to 5 number of siblings attending school. Write the codes and its outputs.

```
female_5siblings <- respondents_data[respondents_data$Sex == 2 & respondents_data$Siblings_at_School >= 5]
print(female_5siblings)
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1   2                1             5             6
## 7             7   2                3             6             5
## 13            13   2                1             4             5
## 14            14   2                3             7             5
## 18            18   2                1            11             5
##      Types_of_Houses
## 1                 1
## 7                 3
## 13                2
## 14                2
## 18                3
```

2. Write a R program to create an empty data frame. Using the following codes: df = data.frame(Ints=integer(),

Doubles=double(), Characters=character(), Logicals=logical(), Factors=factor(), stringsAsFactors=FALSE)
print("Structure of the empty dataframe:") print(str(df)) a. Describe the results.

```
df = data.frame(Ints=integer(),
Doubles=double(), Characters=character(),
Logicals=logical(),
Factors=factor(),
stringsAsFactors=FALSE)

print("Structure of the empty dataframe:")
```

```
## [1] "Structure of the empty dataframe:"
print(str(df))
```

```
## 'data.frame':    0 obs. of  5 variables:
## $ Ints      : int
## $ Doubles   : num
## $ Characters: chr
## $ Logicals  : logi
## $ Factors   : Factor w/ 0 levels:
## NULL
```

3. Create a .csv file of this. Save it as HouseholdData.csv

```
write.csv(respondents_data, "HouseholdData.csv", row.names = FALSE)
```

- a. Import the csv file into the R environment. Write the codes.

```
imported_data <- read.csv("HouseholdData.csv")
```

- b. Convert the Sex into factor using factor() function and change it into integer.[Legend: Male = 1 and Female = 2]. Write the R codes and its output.

```
imported_data$Sex <- factor(imported_data$Sex, levels = c("Male", "Female"))
```

- c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2; Semi-Congrete = 3]. Write the R codes and its output.

```
imported_data$Types_of_Houses <- factor(imported_data$Types_of_Houses,
                                         levels = c("Wood", "Semi-Concrete", "Concrete"))
```

- d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3. What is the R code and its output?

```
imported_data$Fathers_Occupation <- factor(imported_data$Fathers_Occupation,
                                           levels = c("Farmer", "Driver", "Others"))
```

- e. Select only all females respondent that has a father whose occupation is driver. Write the codes and its output.

```
female_driver_fathers <- respondents_data[respondents_data$Sex == "Female" & respondents_data$Fathers_Occupation == "Driver", ]
print(female_driver_fathers)
```

```
## [1] Respondents      Sex      Fathers_Occupation Persons_at_Home
## [5] Siblings_at_School Types_of_Houses
## <0 rows> (or 0-length row.names)
```

- f. Select the respondents that have greater than or equal to 5 number of siblings attending school. Write the codes and its output.

```
five_or_more_siblings <- respondents_data[respondents_data$Siblings_at_School >= 5, ]
print(five_or_more_siblings)
```

```
##      Respondents Sex Fathers_Occupation Persons_at_Home Siblings_at_School
## 1             1  2             1             5             6
## 7             7  2             3             6             5
## 13            13  2             1             4             5
## 14            14  2             3             7             5
## 18            18  2             1            11             5
##      Types_of_Houses
## 1             1
## 7             3
## 13            2
## 14            2
## 18            3
```

4. Interpret the graph.

The sentiment analysis graph is a stacked bar chart that categorizes public sentiment across five topics: Technology, Food, Politics, Sports, and Entertainment. It shows that Technology has the highest positive sentiment, while Politics has the most negative sentiment, reflecting public frustration. Food has a balanced distribution of positive, negative, and significant neutral sentiments, indicating mixed feelings. Sports display slightly more positive than negative sentiments, and Entertainment leans towards positivity with nearly 50% positive reactions. Overall, the graph highlights varying sentiments across these categories, providing insights into public opinion that can inform marketing strategies and content creation.