

# RWorksheet\_MATIAS#3b

2023-10-17

*#1a. Write the codes*

```
respon <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
occu <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
personhome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sib <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
housetype <- c(1,2,3,1,1,3,3,1,2,3,2,3,2,2,3,3,3,3,3,2)
```

```
household_data <- data.frame(
  Respondents = respon,
  Sex = sex,
  FatherOccupation = occu,
  PersonAtHome = personhome,
  SiblingsAtSchool = sib,
  HouseType = housetype
)
household_data
```

##	Respondents	Sex	FatherOccupation	PersonAtHome	SiblingsAtSchool	HouseType
## 1	1	2	1	5	6	1
## 2	2	2	3	7	4	2
## 3	3	1	3	3	4	3
## 4	4	2	3	8	1	1
## 5	5	2	1	5	2	1
## 6	6	2	2	9	1	3
## 7	7	2	3	6	5	3
## 8	8	2	1	7	3	1
## 9	9	2	1	8	1	2
## 10	10	2	1	4	2	3
## 11	11	1	3	7	3	2
## 12	12	2	2	5	2	3
## 13	13	2	1	4	5	2
## 14	14	2	3	7	5	2
## 15	15	2	3	8	2	3
## 16	16	2	1	8	1	3
## 17	17	2	3	3	2	3
## 18	18	2	1	11	5	3
## 19	19	1	2	7	3	3
## 20	20	2	1	6	2	2

*#1b. Describe the data.*

```
str(household_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 ...
## $ FatherOccupation: num 1 3 3 3 1 2 3 1 1 ...
## $ PersonAtHome : num 5 7 3 8 5 9 6 7 8 4 ...
## $ SiblingsAtSchool: num 6 4 4 1 2 1 5 3 1 2 ...
## $ HouseType : num 1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(household_data)
```

```
## Respondents Sex FatherOccupation PersonAtHome
## 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
## SiblingsAtSchool HouseType
## 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
```

```
# the data frame consists of 20 observations(rows) and 6 variables (columns)
```

```
# the variables are:
```

```
# respondents - which contains a numeric identifier for each respondent
```

```
# sex - represents the gender of the respondent (1 for male, 2 for female)
```

```
# father's occupation - indicates the father's occupation (1 for farmer, 2 for driver, 3 for others)
```

```
# persons at home - represents the number of people at home
```

```
# siblings at school - indicates the number of siblings attending school
```

```
# type of house - describes the type of house (1 for wood, 2 for semi-concrete, 3 for concrete)
```

```
#1c. Is the mean number of siblings attending is 5?
```

```
sib_mean <- mean(household_data$SiblingsAtSchool)
```

```
sib_mean
```

```
## [1] 2.95
```

```
# the mean of the number of siblings at school is 2.95 , not 5.
```

```
#1d. Extract the 1st two rows
```

```
first2Rows <- household_data[1:2,]
```

```
first2Rows
```

```
## Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 1 1 2 1 5 6 1
## 2 2 2 3 7 4 2
```

```
#1e. Extract 3rd and 5th row with 2nd and 4th column.
```

```
thirdand5thRows <- household_data[c(3,5),c(2,4)]
```

```
thirdand5thRows
```

```
## Sex PersonAtHome
```

```
## 3 1 3
```

```
## 5 2 5
```

```
#1e. Select the variable types of houses
```

```
types_houses <- household_data$HouseType
```

```
types_houses
```

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

```
#1g. Select only all Males respondent that their father occupation was farmer.
```

```
farmer <- household_data[household_data$Sex == 1 & household_data$FatherOccupation == 1,]  
farmer
```

```
## [1] Respondents      Sex              FatherOccupation PersonAtHome
```

```
## [5] SiblingsAtSchool HouseType
```

```
## <0 rows> (or 0-length row.names)
```

```
# there is no observations
```

```
#1h. Select only all females respondent that have greater than or equal to 5 number of siblings attending
```

```
female_respon <- household_data[household_data$SiblingsAtSchool >= 5,]
```

```
female_respon
```

```
##      Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
```

```
## 1              1  2              1              5              6              1
```

```
## 7              7  2              3              6              5              3
```

```
## 13             13  2              1              4              5              2
```

```
## 14             14  2              3              7              5              2
```

```
## 18             18  2              1             11              5              3
```

```
# there are five observations
```

```
#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:")
```

```
## [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
```

```
#2a. Describe the results.
```

```
# df is an empty data frame created with 0 rows and 5 columns
```

```
# the columns has the following data type:
```

```
# ints = integer
```

```
# doubles = double
```

```
# characters = character
```

```
# logicals = logical
```

```

# factors = factor (0 levels which means empty)
# serves as a template and can be populated with data

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

new_respon <- c(1:10)
new_sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
new_occu <- c(1,2,3,3,1,2,2,3,1,3)
new_personHome <- c(5,7,3,8,6,4,4,2,11,6)
new_sib <- c(2,3,0,5,2,3,1,2,6,2)
new_housetype <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood", "S", "S", "S")

HouseholdData <- data.frame(
  Respondents = new_respon,
  Sex = new_sex,
  FatherOccupation = new_occu,
  PersonAtHome = new_personHome,
  SiblingsAtSchool = new_sib,
  HouseType = new_housetype
)
HouseholdData

##      Respondents      Sex FatherOccupation PersonAtHome SiblingsAtSchool
## 1             1    Male                1             5             2
## 2             2  Female                2             7             3
## 3             3  Female                3             3             0
## 4             4    Male                3             8             5
## 5             5    Male                1             6             2
## 6             6  Female                2             4             3
## 7             7  Female                2             4             1
## 8             8    Male                3             2             2
## 9             9  Female                1            11             6
## 10           10    Male                3             6             2
##      HouseType
## 1           Wood
## 2         Congrete
## 3         Congrete
## 4           Wood
## 5 Semi-concrete
## 6 Semi-concrete
## 7           Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10          Congrete

write.csv(HouseholdData, file = "HouseholdData.csv")

#3a. Import the csv file into the R environment.

import <- read.csv("HouseholdData.csv")
import

```

```
##      X Respondents      Sex FatherOccupation PersonAtHome SiblingsAtSchool
## 1    1              1   Male                1              5              2
## 2    2              2 Female                2              7              3
## 3    3              3 Female                3              3              0
## 4    4              4   Male                3              8              5
## 5    5              5   Male                1              6              2
## 6    6              6 Female                2              4              3
## 7    7              7 Female                2              4              1
## 8    8              8   Male                3              2              2
## 9    9              9 Female                1             11              6
## 10  10             10   Male                3              6              2
```

```
##      HouseType
## 1           Wood
## 2         Congrete
## 3         Congrete
## 4           Wood
## 5 Semi-concrete
## 6 Semi-concrete
## 7           Wood
## 8 Semi-concrete
## 9 Semi-concrete
## 10          Congrete
```

*#3b. Convert the Sex into factor using factor() function and change it into integer. [Legend: Male = 1 and Female = 2]*

```
import$Sex <- factor(import$Sex, levels = c("Male", "Female"))
import$Sex <- as.integer(import$Sex)
import$Sex
```

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

*#3c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2 and Semi-concrete = 3]*

```
import$HouseType <- factor(import$HouseType, levels = c("Wood", "Congrete", "Semi-concrete"))
import$HouseType <- as.integer(import$HouseType)
import$HouseType
```

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

*#3d. On father's occupation, factor it as Farmer = 1; Driver = 2; and Others = 3.*

```
import$FatherOccupation <- factor(import$FatherOccupation, levels = c(1,2,3), labels = c("Farmer", "Driver", "Others"))
import$FatherOccupation
```

```
## [1] Farmer Driver Others Others Farmer Driver Driver Others Farmer Others
## Levels: Farmer Driver Others
```

*#3e. Select only all females respondent that has a father whose occupation is driver.*

```
femaledriver <- import[import$Sex == 2 & import$FatherOccupation == "Driver",]
femaledriver
```

```
##      X Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType
## 2 2              2  2             Driver              7              3              2
## 6 6              6  2             Driver              4              3              3
## 7 7              7  2             Driver              4              1              1
```

*#3f. Select the respondents that have greater than or equal to 5 number of siblings attending school.*

```
greater5 <- import[import$SiblingsAtSchool >= 5,]  
greater5
```

```
##      X Respondents Sex FatherOccupation PersonAtHome SiblingsAtSchool HouseType  
## 4 4          4      1          Others           8           5           1  
## 9 9          9      2          Farmer          11           6           3
```

*#4. Interpret the graph.*

*# In comparison to the other sentiments, there were more negative sentiments on July 14. This may be a*

*# On July 15, all opinions grew, with the negative sentiment reaching its maximum level. This may indic*

*# The negative sentiments remained strong on July 17 and 18, while the neutral and positive feelings re*

*# Even though all sentiments reached their lowest point on July 20, there were still more negative feel*

*# On July 21, all sentiments increased, with the negative mood reaching its greatest level. This can al*

*# We may infer from this data that public sentiment is sensitive to outside causes and that it changes*