

RWorksheets_Bagilidad#3b

2023-10-11

#1.

#a. Write the codes.

```
respondents <- c(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)

household_data <- data.frame(Respondents =respondents,
                             Sex = sex,
                             FathersOccupation = fathers_occupation,
                             PersonsAtHome = persons_at_home,
                             SiblingsAtSchool = siblings_at_school,
                             TypeOfHouses = types_of_houses)

household_data
```

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

```
## 5      1
## 6      3
## 7      3
## 8      1
## 9      2
## 10     3
## 11     2
## 12     3
## 13     2
## 14     2
## 15     3
## 16     3
## 17     3
## 18     3
## 19     3
## 20     2
```

#b. Describe the data. Get the structure or the summary of 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 2 ...
## $ FathersOccupation: num 1 3 3 3 1 2 3 1 1 1 ...
## $ PersonsAtHome : num 5 7 3 8 5 9 6 7 8 4 ...
## $ SiblingsAtSchool : num 6 4 4 1 2 1 5 3 1 2 ...
## $ TypeOfHouses : num 1 2 3 1 1 3 3 1 2 3 ...
```

```
summary(household_data)
```

```
## Respondents Sex FathersOccupation PersonsAtHome
## 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 TypeOfHouses
## 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 data frame consists of 6 variables (columns) with 20 observations(rows)

respondents - which lists each respondent's unique numeric identifier.

sex - represents the gender of the respondent (1 for Male and 2 for Female).

father's occupation - indicates the occupation of the father (1 for farmer, 2 for driver, and 3 for other)

persons at home - consists of number of people at home.

siblings at school - consists of number of siblings at the school.

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

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

```
means_siblings <- mean(siblings_at_school)
means_siblings
```

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

The mean of siblings attending in school is not 5 but 2.95.

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

```
firstTwo_rows <- head(household_data, 2)
firstTwo_rows
```

```
## Respondents Sex FathersOccupation PersongsAtHome SiblingsAtSchool
## 1          1 2              1              5              6
## 2          2 2              3              7              4
## TypeOfHouses
## 1          1
## 2          2
```

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

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

```
## Sex PersongsAtHome
## 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 and its result.

```
types_houses <- household_data$types_of_houses
types_houses
```

```
## NULL
```

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

```
male_father_farmer= household_data[household_data$sex == 1 & household_data$fathers_occupation == 1,]
male_father_farmer
```

```
## [1] Respondents Sex FathersOccupation PersongsAtHome
## [5] SiblingsAtSchool TypeOfHouses
## <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 in school. Write the codes and its result.

```
females_greaterThan5 <- household_data[household_data$siblings_at_school >= 5,]
females_greaterThan5
```

```
## [1] Respondents Sex FathersOccupation PersongsAtHome
## [5] SiblingsAtSchool TypeOfHouses
## <0 rows> (or 0-length row.names)
```

```
#-----
```

#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
```

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

```
# df have no data frame with 0 rows and 5 columns
```

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

```
# ints means integer
```

```
# doubles is a data type that is used to store high-precision floating-point data or numbers.
```

```
# characters is a single display unit of information equivalent to one alphabetic symbol, digit, or let
```

```
# logicals is a programming, database and knowledge-representation and reasoning paradigm which is base
```

```
# factors data structures that are implemented to categorize the data or represent categorical data and
```

```
# 0 levels which means empty
```

```
# can be serve as a template that can be populated with data
```

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

```
New_Respondent <- c(1:10)
```

```
New_Sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
```

```
New_Fathers_Occupation <- c(1,2,3,3,1,2,2,3,1,3)
```

```
New_PersonsAt_Home <- c(5,7,3,8,6,4,4,2,11,6)
```

```
New_SiblingsAt_School <- c(2,3,0,5,2,3,1,2,6,2)
```

```
New_Types_Houses <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-concrete", "Semi-concrete", "Wood",
```

```
HouseholdData <- data.frame(
  Respondents = New_Respondent,
  Sex = New_Sex,
  FatherOccupation = New_Fathers_Occupation,
  PersonsAtHome=New_PersonsAt_Home,
  SiblingsAtSchool=New_SiblingsAt_School,
  TypesOfHouses=New_Types_Houses)
HouseholdData
```

```
##   Respondents   Sex FatherOccupation PersonsAtHome 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
```

```
##      TypesOfHouses
## 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")
```

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

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

```
##      X Respondents      Sex FatherOccupation PersonsAtHome 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
```

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

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

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

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

c. Convert the Type of Houses into factor and change it into integer. [Legend: Wood = 1; Congrete = 2

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

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

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

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

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

```
## Levels: Farmer Driver Others
```

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

```
femaleDriver <- imported[imported$Sex == 2 & imported$FatherOccupation == "Driver",]
femaleDriver
```

```
## X Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool
```

```
## 2 2 2 Driver 7 3
```

```
## 6 6 2 Driver 4 3
```

```
## 7 7 2 Driver 4 1
```

```
## TypesOfHouses HouseType
```

```
## 2 Congrete 2
```

```
## 6 Semi-concrete 3
```

```
## 7 Wood 1
```

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

```
upper_five <- imported[imported$SiblingsAtSchool >= 5,]
upper_five
```

```
## X Respondents Sex FatherOccupation PersonsAtHome SiblingsAtSchool
```

```
## 4 4 4 1 Others 8 5
```

```
## 9 9 9 2 Farmer 11 6
```

```
## TypesOfHouses HouseType
```

```
## 4 Wood 1
```

```
## 9 Semi-concrete 3
```

#4

#The majority of the other sentiments on this day, July 14, are negative. This indicates that some subjects

#Even if all of the attitudes increased on this day, July 15, the negative sentiment is still at its greatest

#Negative attitude is still strong on July 17 and 18, while both neutral and positive sentiment is still low

#There were still more negative attitudes among the others on July 20 even though all sentiments reached

#On this day, July 21, all emotions are higher, with the negative continuing dominating. This could imply

#This information could lead us to the conclusion that public opinion is subject to outside influences