RWorksheet_Suero#3b

2023-10-11

```
#1.a
respondents_no <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2)
focc <- c(1,3,3,3,1,2,3,1,1,1,3,2,1,3,3,1,3,1,2,1)
persAtHome <- c(5,7,3,8,5,9,6,7,8,4,7,5,4,7,8,8,3,11,7,6)
sibsschool <- c(6,4,4,1,2,1,5,3,1,2,3,2,5,5,2,1,2,5,3,2)
type_house<- 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_no,
  Sex = sex,
  FatherOccupation = focc,
  PersonAtHome = persAtHome,
  SiblingsAtSchool = sibsschool,
  HouseType = type_house
)
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
```

```
#1.b

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 ...
##  $ FatherOccupation: num  1 3 3 3 1 2 3 1 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 data frame consists of 6 variables (colums) with 20 observations(rows)


# respondents - which contains a numeric identifier for each respondent
# sex -(1 for male, 2 for female)it represents the gender of the respondent
# father's occupation - indicates the occupation of father (1 for farmer, 2 for driver, 3 for others)
# persons at home - shows the number of people at home
# siblings at school - indicates the number of siblings at the school
# type of house - describes the house type (1 for wood, 2 for semi-concrete, 3 for concrete)

#1.c
sibsschool_mean <- mean(household_data$SiblingsAtSchool)
sibsschool_mean
```

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

```
# no it is not 5 because the mean of siblings at school is 2.95

#1.d
firstTwo_rows <- household_data[1:2,]
firstTwo_rows
```

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

```
#1.e
third_and_fifth_rows <- household_data[c(3,5),c(2,4)]
third_and_fifth_rows
```

```
##   Sex PersonAtHome
## 3   1            3
```

```
## 5   2            5
```

```
#1.f
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
```

```
#1.g
maleFarmer <- household_data[household_data$Sex == 1 & household_data$FatherOccupation == 1,]

maleFarmer
```

```
## [1] Respondents      Sex              FatherOccupation PersonAtHome
## [5] SiblingsAtSchool HouseType
## <0 rows> (or 0-length row.names)
```

```
# no observations

#1.h

femaleResp <- household_data[household_data$SiblingsAtSchool >= 5,]

femaleResp
```

```
##    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
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
```

```r
# df have a empty data frame 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)

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

# ------------------------------------------------

# 3

newResp <- c(1:10)
newSex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "Male")
newOcc <- c(1,2,3,3,1,2,2,3,1,3)
newPersonsAtHome <- c(5,7,3,8,6,4,4,2,11,6)
newSibs <- c(2,3,0,5,2,3,1,2,6,2)
newType <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete", "Semi-congrete", "Wood", "Semi-co

Household_Data <- data.frame(
  Respondents = newResp,
  Sex = newSex,
  FatherOccupation = newOcc,
  PersonAtHome = newPersonsAtHome,
  SiblingsAtSchool = newSibs,
  HouseType = newType
)

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

#3a

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

```
##     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-congrete
## 6  Semi-congrete
## 7           Wood
## 8  Semi-congrete
## 9  Semi-congrete
## 10      Congrete
```

```
#3b

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
```

```
#3c

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

imported$HouseType
```

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

```
#3d

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

imported$FatherOccupation
```

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

```
#3e

femaleDriver <- imported[imported$Sex == 2 & imported$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

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

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

```
#4
# On this day, July 14 the negative sentiments has the most among the other sentiments. This means that

# On this day, July 15 the negative sentiment is still at the highest even if all the sentiments increa

# On these days, July 17 and July 18 negative sentiments remains high and both neutral and positive sen
```

```
# On the day, July 20 all sentiments got to their lowest even so there were still more negative sentime
# On this day July 21, all sentiments increases, still the negative being at the top. This could mean t
# From this data, we could come to the conclusion that public sentiment is responsive to external facto
```