RWorksheet_Salinas#3b

Reysha Marie S. Salinas

2023-10-13

```
#1
#A. Create a data frame using the table below.

        respondent <- c(1:20)
        sex <- c(2,2,1,2,2,2,2,2,2,2,2,1,2,2,2,2,2,2,2,1,2)
        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)
        type_of_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 = respondent,
          Sex = sex,
          Father_Occupation = fathers_occupation,
          Person_At_Home = persons_at_home,
          Siblings_At_School = siblings_at_school,
          House_Type = type_of_house
        )
        household_data
```

```
##    Respondents Sex Father_Occupation Person_At_Home Siblings_At_School
## 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
##    House_Type
```

```
## 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.

```r
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 ...
##  $ Father_Occupation : num  1 3 3 3 1 2 3 1 1 1 ...
##  $ Person_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 ...
##  $ House_Type        : num  1 2 3 1 1 3 3 1 2 3 ...
```

```r
summary(household_data)
```

```
##   Respondents          Sex        Father_Occupation Person_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   House_Type
##  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 and 6 variables
        #Respondents: A variable is type of integer and there are 20 respondents in dataset.
        #Sex: A variable is type of numeric with values 1 and 2 represent the gender of male and fema
        #Father_Occupation: A variable is type of numeric with values 1, 2, and 3 , maybe it represen
        #Person_At_Home: A variable is type of numeric, and it represent the number of people at home
        #Siblings_At_School: A variable is type of numeric and represent the number of siblings that
```

```r
        #House_Type: A variable is numeric with values 1, 2, and 3, it represent kind of houses.


#C Is the mean number of siblings attending is 5?
        siblings_mean <- mean(household_data$Siblings_At_School)

        siblings_mean
```

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

```r
        # No, The mean number of siblings attending is 2.95.

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

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

```
##   Respondents Sex Father_Occupation Person_At_Home Siblings_At_School
## 1           1   2                 1              5                  6
## 2           2   2                 3              7                  4
##   House_Type
## 1          1
## 2          2
```

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

#E Extract 3rd and 5th row with 2nd and 4th column.
#Write the codes and its result.
        third_and_fifthrows <- household_data[c(3,5),c(2,4)]
        third_and_fifthrows
```

```
##   Sex Person_At_Home
## 3   1              3
## 5   2              5
```

```r
#OUTPUT
#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 <- household_data$House_Type
        types_houses
```

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

```r
#G Select only all Males respondent that their father occupation was farmer.
#Write the codes and its output.

        household_data[household_data$Sex == 1 & household_data$Father_Occupation == "farmer", ]
```

```
## [1] Respondents         Sex              Father_Occupation  Person_At_Home
## [5] Siblings_At_School House_Type
## <0 rows> (or 0-length row.names)
```

household_data$Father_Occupation

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

#OUPUT <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
#Write the codes and its outputs.

female <- household_data[household_data$Siblings_At_School >= 5,]
female

```
##    Respondents Sex Father_Occupation Person_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
##    House_Type
## 1           1
## 7           3
## 13          2
## 14          2
## 18          3
```

 # OUTPUT There are five observations

#Respondents	Sex	FatherOccupation	PersonAtHome	SiblingsAtSchool	HouseType
#1	2	1	5	6	1
#7	2	3	6	5	3
#13	2	1	4	5	2
#14	2	3	7	5	2
#18	2	1	11	5	3

#2. Write a R program to create an empty data frame. Using the following codes:

```r
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:"
```

```r
print(str(df))
```

```
## 'data.frame':    0 obs. of  5 variables:
##  $ Ints      : int
```

```
##  $ Doubles   : num
##  $ Characters: chr
##  $ Logicals  : logi
##  $ Factors   : Factor w/ 0 levels:
## NULL
```

```
#Data frame is empty, this part of the output indicates that you have a data frame with 0 observations
# Ints,column integer data type.
# Doubles, column is numeric (double) data type.
# Characters,  column is character data type.
# Logicals, column is logical (boolean) data type.
# Factors, column a factor variable with 0 levels, currently has no unique levels.


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

        ot_respondent <- c(1:10)
        ot_sex <- c("Male", "Female", "Female", "Male", "Male", "Female", "Female", "Male", "Female", "
        ot_fathers_occupation <- c(1,2,3,3,1,2,2,3,1,3)
        ot_persons_at_home <- c(5,7,3,8,6,4,4,2,11,6)
        ot_siblings_at_school <- c(2,3,0,5,2,3,1,2,6,2)
        ot_type_of_house <- c("Wood", "Congrete", "Congrete", "Wood", "Semi-congrete", "Semi-congrete",

        HouseholdData <- data.frame(
          Respondents = ot_respondent,
          Sex = ot_sex,
          FatherOccupation = ot_fathers_occupution,
          PersonAtHome =ot_persons_at_home,
          SiblingsAtSchool = ot_siblings_at_school,
          HouseType = ot_type_of_house
        )

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

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

        householddata <- read.csv("HouseholdData.csv")
        head(householddata)
```

```
##   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
##       HouseType
## 1          Wood
## 2      Congrete
## 3      Congrete
## 4          Wood
## 5 Semi-congrete
## 6 Semi-congrete
```

```
#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.


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

        householddata$Sex
```

[1] 1 2 2 1 1 2 2 1 2 1

```
        #OUTPUT 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; Semi-Congrete = 3]. Write the R codes and its output.

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

        householddata$HouseType
```

[1] 1 2 2 1 3 3 1 3 3 2

```
        #OUTPUT 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 its output?

        householddata$FatherOccupation <- factor(householddata$FatherOccupation, levels = c(1,2,3), lab
        householddata$FatherOccupation
```

[1] Farmer, Driver, Others, Others, Farmer, Driver, Driver, Others, Farmer,
[10] Others,
Levels: Farmer, Driver, Others,

```
        #OUTPUT
        #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 its output.

        female <- householddata[householddata$Sex == 2 & householddata$FatherOccupation == "Driver",]
        female
```

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

```
        #OUTPUT
        #[1] X                Respondents     Sex             FatherOccupation PersonAtHome
        #[6] SiblingsAtSchool HouseType
        #<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 <- householddata[householddata$SiblingsAtSchool >= 5,]
    five
```

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

```
#OUTOUT 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.

```
#2022
# On July 14, the sentiments expressed in tweets on that day consisted of 2500 negative, 1500 n
# On July 15, the number of negative tweets increased from 2500 to 4000 compared to the previou
# On July 17, the number of negative sentiments decreased to a range between 3000 and 3500, whi
# On July 18, the number of negative sentiments remained the same as the previous day, with a r
# On July 20, the number of negative sentiments in tweets decreased from the previous day, rang
# On July 21, On July 21, the number of negative sentiments in tweets increased again from the

#The specific reasons for these sentiment changes would require additional context and analysis
```