

# RWorksheet\_Salinas#3b

Reysa Marie S. Salinas

2023-10-13

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

```
respondent <- c(1:20)
sex <- c(2,2,1,2,2,2,2,2,2,2,1,2,2,2,2,2,2,1,2)
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)
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
```

#B Describe the data. Get the structure or the summary of the data.

```
str(household_data) summary(household_data)
```

```
# 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 female.
#Father_Occupation: A variable is type of numeric with values 1, 2, and 3 , maybe it represent the
#Person_At_Home: A variable is type of numeric, and it represent the number of people at home for
#Siblings_At_School: A variable is type of numeric and represent the number of siblings that the
#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
```

```
# 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
```

```
#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 #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
```

```
#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", ]
household_data$Father_Occupation
```

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

```
# OUTPUT There are five observations
```

	#Respondents	Sex	FatherOccupation	PersonAtHome	SiblingsAtSchool	HouseType
#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

```
#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))
```

```
#Data frame is empty, this part of the output indicates that you have a data frame with 0 observations
(rows) and 5 variables (columns). # 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", "Male")
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", "Wo
```

```
HouseholdData <- data.frame(
  Respondents = ot_respondent,
```

```

    Sex = ot_sex,
    FatherOccupation = ot_fathers_occupation,
    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)

#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

#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-congrete"))
householddata$HouseType <- as.integer(householddata$HouseType)

householddata$HouseType

#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), labels = c("Farmer", "Driver", "Others"))
householddata$FatherOccupation

#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

#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
```

#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 neutral, and a range of 1500 to 2000 positive sentiments. # On July 15, the number of negative tweets increased from 2500 to 4000 compared to the previous day. The number of neutral tweets also increased, ranging from 2500 to 3000, and positive tweets saw an increase ranging from 3000 to 3500. # On July 17, the number of negative sentiments decreased to a range between 3000 and 3500, while the number of neutral sentiments dropped to nearly 2000, and positive sentiments decreased to almost 2500. # On July 18, the number of negative sentiments remained the same as the previous day, with a range between 3000 and 3500. The number of neutral sentiments increased to 2000, and positive sentiments increased to 2500. # On July 20, the number of negative sentiments in tweets decreased from the previous day, ranging to almost 2500. Neutral sentiments decreased to almost 1500, and positive sentiments decreased to a range between 1500 and 2000. # On July 21, On July 21, the number of negative sentiments in tweets increased again from the previous day, ranging to around 4100. Neutral sentiments increased from the previous day, ranging between 2500 and 3000, and positive sentiments also increased from the previous day, almost reaching 3500.

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