Worksheet-7a

Wendy Nalaza BSIT-2A

2022-12-10

```
install.packages("Hmisc")
install.packages("pastecs")
```

#1. Create a data frame for the table below

```
Student <- seq(1:10)
PreTest <- c(55,54,47,57,51,61,57,54,63,58)
PostTest <- c(61,60,56,63,56,63,59,56,62,61)

Data_F <- data.frame(Student,PreTest,PostTest)
Data_F
```

#a. Compute the descriptive statistics using different packages (Hmisc and pastecs).

#Write the codes and its result.

```
library(Hmisc)
library(pastecs)

describe(Data_F)

stat.desc(Data_F)
```

#2. The Department of Agriculture was studying the effects of several levels of a #fertilizer on the growth of a plant. For some analyses, it might be useful to convert #the fertilizer levels to an ordered factor.

```
DoA <- c(10,10,10,20,20,50,10,20,10,50,20,50,20,10)
```

#a. Write the codes and describe the result.

```
In_Ord <- sort(DoA, decreasing = FALSE)
In_Ord
```

#3. Abdul Hassan, president of Floor Coverings Unlimited, has asked you to study #the exercise levels undertaken by 10 subjects were “l”, “n”, “n”, “i”, “l” , #“l”, “n”, “n”, “i”, “l” ; n=none, l=light, i=intense

```
Subjects <- c("l","n","n","i","l","l","n","n","i","l")
```

#a. What is the best way to represent this in R?

#DATAFRAME

```
output <- data.frame(Subjects)
output
```

#4.Sample of 30 tax accountants from all the states and territories of Australia and #their individual state of origin is specified by a character vector of state mnemonics #as:

```
  state <- c("tas", "sa", "qld", "nsw", "nsw", "nt", "wa", "wa", "qld",
             "vic", "nsw", "vic", "qld", "qld", "sa", "tas", "sa", "nt",
             "wa", "vic", "qld", "nsw", "nsw", "wa", "sa", "act", "nsw",
             "vic", "vic", "act")
state
```

#a. Apply the factor function and factor level. Describe the results.

```
alter <- function(state)
alter
```

#5. From #4 - continuation:

• Suppose we have the incomes of the same tax accountants in another vector (in incomes <- c(60, 49, 40, 61, 64, 60, 59, 54, 62, 69, 70, 42, 56, 61, 61, 61, 58, 51, 48, 65, 49, 49, 41, 48, 52, 46, 59, 46, 58, 43)

#a. Calculate the sample mean income for each state we can now use the special #function tapply():

```
 Calc <- tapply(state, incomes, mean)
 Calc
```

#b. Copy the results and interpret.

40 41 42 43 46 48 49 51 52 54 56 58 59 60 61 62 64 65 69 70

#NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA

#6.Calculate the standard errors of the state income means (refer again to number 3)

```
Calc_ST.n <- length(Calc)
Calc_1.sd <- sd(Calc)
Calc_Final.se <- Calc_1.sd/sqrt(Calc_ST.n)
Calc_Final.se
```

#a. What is the standard error? Write the codes. #NA #b. Interpret the result. #the result is not available because some variables are character type so it won't able to get the standard error. #7. Use the titanic dataset.

```
data("Titanic")
```

```
head<- data.frame(Titanic)
```

#a. subset the titatic dataset of those who survived and not survived. Show the #codes and its result.

```
head_subset <- subset(head, select = "Survived")
head_subset
```

#8. The data sets are about the breast cancer Wisconsin. The samples arrive periodically as Dr. Wolberg reports his clinical cases. The database therefore reflects this #chronological grouping of the data. You can create this dataset in Microsoft Excel.

#a. describe what is the dataset all about. #The dataset s all about Breast Cancer.

#b. Import the data from MS Excel. Copy the codes.

```
library("readxl")

DATA <- read_excel("C:\\Users\\User\\Worksheet#7//Breast_Cancer.xlsx")
DATA
```

#c. Compute the descriptive statistics using different packages. Find the values of:

#c.1 Standard error of the mean for clump thickness.

```
Clump <- length(DATA$`CL. thickness`)
Clump_A <- sd(DATA$`CL. thickness`)
Clump_B <- Clump_A/sqrt(DATA$`CL. thickness`)
Clump_B
```

#c.2 Coefficient of variability for Marginal Adhesion.

```
COV <- sd(DATA$`Marg. Adhesion`) / mean(DATA$`Marg. Adhesion`)* 100
COV
```

#c.3 Number of null values of Bare Nuclei.

```
Null_Values <- subset(DATA,`Bare. Nuclei` == "NA")
```

#c.4 Mean and standard deviation for Bland Chromatin

```
mean(DATA$`Bl. Cromatin`)
sd(DATA$`Bl. Cromatin`)
```

#c.5 Confidence interval of the mean for Uniformity of Cell Shape

#Calculate the mean

```
Calc_Mean <- mean(DATA$`Cell Shape`)
Calc_Mean
```

#Calculate the standard error of the mean

```
SE_M <- length(DATA$`Cell Shape`)
SD_B <- sd(DATA$`Cell Shape`)
Ans_1 <- SD_B/sqrt(SE_M)
Ans_1
```

#Find the t-score that corresponds to the confidence level

```
D = 0.05
numE = SE_M - 1
numF = qt(p = D/ 2, df = numE,lower.tail = F)
numF
```

#Constructing the confidence interval

```
numG <- numF * numC
```

#Lower

```
numH <- Calc_Mean - numG
```

#Upper

```
numI <- Calc_Mean + numG
```

```
c(numH,numI)
```

#d. How many attributes?

```
attributes(DATA)
```

#e. Find the percentage of respondents who are malignant. Interpret the results.

```
P_R <- subset(DATA, Class == "maligant")
P_R
```

#There 17 respondents who are malignant. #And there are total of 49 respondent.

#Getting the percentage

```
17  / 49 * 100
```

#9. Export the data abalone to the Microsoft excel file. Copy the codes.

```
install.packages("AppliedPredictiveModeling")
```

```
library("AppliedPredictiveModeling")
data("abalone")
View(abalone)
head(abalone)
summary(abalone)
```

#Exporting the data abalone to the Microsoft excel file install.packages("xlsxjars")

```
library(xlsx)
write.xlsx("abalone","C:\\Users\\User\\Desktop\\Worksheets\\Worksheet#7\\abalone.xlsx")
```