## RWorksheet\_#7a

## Jessa Mae Mirabuena

2022-12-11

```
library(Hmisc)
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
## format.pval, units
library(pastecs)
```

#1. Create a data frame for the table below

```
Student <- seq(1:10)</pre>
PreTest <- c(55,54,47,57,51,61,57,54,63,58)
PostTest <- c(61,60,56,63,56,63,59,56,62,61)
a <- data.frame(Student,PreTest,PostTest)</pre>
а
##
      Student PreTest PostTest
## 1
             1
                     55
                               61
## 2
             2
                     54
                               60
             3
                     47
                               56
## 3
## 4
             4
                     57
                               63
             5
                     51
## 5
                               56
             6
                     61
                               63
## 6
                               59
## 7
             7
                     57
             8
                     54
## 8
                               56
## 9
             9
                     63
                               62
                     58
                               61
```

#a. Compute the descriptive statistics using different packages (Hmisc and pastecs). #Write the codes and its result.

```
describe(a)
```

```
## a
##
## 3 Variables 10 Observations
## -----
-----
## Student
     n missing distinct
                          Info
                                 Mean
                                         Gmd
                                               .05
                                                      .10
##
                                 5.5
      10
          0
                    10
                           1
                                       3.667
                                               1.45
                                                      1.90
            .50
                    .75
                                 .95
##
      .25
                           .90
                   7.75
##
     3.25
            5.50
                          9.10
                                 9.55
##
## lowest: 1 2 3 4 5, highest: 6 7 8 9 10
##
## Value
           1 2 3 4 5 6 7
## Frequency
          1 1 1
                     1
                        1
                           1
                               1
                                  1
                                     1
## ------
## PreTest
      n missing distinct
                         Info
                                 Mean
                                         Gmd
##
      10
             0
                    8
                         0.988
                                55.7
                                       5.444
##
## lowest : 47 51 54 55 57, highest: 55 57 58 61 63
##
## Value
        47 51 54 55 57 58 61 63
           1 1 2 1
## Frequency
                        2
                           1
## Proportion 0.1 0.1 0.2 0.1 0.2 0.1 0.1 0.1
## -----
## PostTest
       n missing distinct
                          Info
                                 Mean
                                         Gmd
      10
         0
                         0.964
                                 59.7
                  6
                                       3.311
##
## lowest : 56 59 60 61 62, highest: 59 60 61 62 63
## Value
           56 59 60 61 62 63
          3 1 1
                     2
## Frequency
                        1
## Proportion 0.3 0.1 0.1 0.2 0.1 0.2
-----
stat.desc(a)
##
              Student
                        PreTest
                                  PostTest
## nbr.val
           10.0000000 10.00000000 10.00000000
## nbr.null
           0.0000000
                     0.00000000
                               0.00000000
## nbr.na
             0.0000000
                      0.00000000
                                0.00000000
## min
            1.0000000 47.00000000 56.00000000
## max
            10.0000000 63.00000000 63.00000000
## range 9.0000000 16.00000000 7.00000000
```

```
## sum
               55.0000000 557.00000000 597.00000000
## median
                5.5000000 56.00000000 60.50000000
## mean
                5.5000000 55.70000000 59.70000000
## SE.mean
                0.9574271
                            1.46855938
                                         0.89504811
## CI.mean.0.95 2.1658506
                            3.32211213
                                         2.02473948
## var
                9.1666667 21.56666667
                                         8.01111111
## std.dev
                3.0276504
                            4.64399254
                                         2.83039063
## coef.var
                0.5504819
                            0.08337509
                                         0.04741023
```

#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. #The data were 10,10,10, 20,20,50,10,20,10,50,20,50,20,10.

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

#a. Write the codes and describe the result.

```
order <- factor(Fertilizer_Levels, ordered = TRUE)
order

## [1] 10 10 10 20 20 50 10 20 10 50 20 50 20 10
## Levels: 10 < 20 < 50</pre>
```

#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","n","n","i","l")
```

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

```
Df <- data.frame(Subjects)</pre>
Df
##
       Subjects
## 1
               1
## 2
               n
## 3
               n
## 4
               i
## 5
               1
## 6
               1
## 7
               n
## 8
               n
## 9
               i
## 10
               1
```

#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",
```

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

```
c <- factor(state)
c

## [1] tas sa qld nsw nsw nt wa wa qld vic nsw vic qld qld sa tas sa
nt wa
## [20] vic qld nsw nsw wa sa act nsw vic vic act
## Levels: act nsw nt qld sa tas vic wa
levels(c)
## [1] "act" "nsw" "nt" "qld" "sa" "tas" "vic" "wa"</pre>
```

#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(): #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:

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

```
c <- factor(state)
c</pre>
```

```
## [1] tas sa qld nsw nsw nt wa wa qld vic nsw vic qld qld sa tas sa
nt wa
## [20] vic qld nsw nsw wa sa act nsw vic vic act
## Levels: act nsw nt qld sa tas vic wa
levels(c)
## [1] "act" "nsw" "nt" "qld" "sa" "tas" "vic" "wa"
```

#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():

```
meanincome <- tapply(incomes, state, mean )</pre>
meanincome
##
                             nt
                                     ald
                                                sa
                                                         tas
                                                                   vic
                  nsw
## 44.50000 57.33333 55.50000 53.60000 55.00000 60.50000 56.00000 52.25000
meanincome <- tapply(incomes, state, mean )</pre>
meanincome
##
                                     ald
                                                                   vic
        act
                  nsw
                             nt
                                                sa
                                                         tas
                                                                             wa
## 44.50000 57.33333 55.50000 53.60000 55.00000 60.50000 56.00000 52.25000
```

#b. Copy the results and interpret. # act nsw nt #44.50000 57.33333 55.50000 # qld sa tas #53.60000 55.00000 60.50000 # vic wa #56.00000 52.25000 #56.00000 these are the mean of income

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

```
stdError <- function(x) sqrt(var(x)/length(x))</pre>
```

#a. What is the standard error? Write the codes.

```
incomestdr <- tapply(incomes, state, stdError)
incomestdr

## act nsw nt qld sa tas vic wa
## 1.500000 4.310195 4.500000 4.106093 2.738613 0.500000 5.244044 2.657536</pre>
```

#b. Interpret the result. #It tells the sample mean of income #7. Use the titanic Data set

```
data("Titanic")
titanic<- data.frame(Titanic)</pre>
```

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

#Survives <- subset(Titanic, Survived == "Yes") #Survives

#Died <- subset(Titanic, Survived == "No") #Died</pre>

#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.

```
getwd()
## [1] "/cloud/project"
breastcancer<- read.csv(file = "Breast_Cancer.csv", sep = "",stringsAsFactors</pre>
breastcancer
##
                                        ID.CL. thickness.Cell size.Cell
Shape.Marg.
                                                                       NA
## 1
            1000025,5,1,1,1,2,1,3,1,1,benign
                                                            NA
NA
           1002945,5,4,4,5,7,10,3,2,1,benign
## 2
                                                            NA
                                                                       NA
NA
## 3
            1015425,3,1,1,1,2,2,3,1,1,benign
                                                            NA
                                                                       NA
NA
            1016277,6,8,8,1,3,4,3,7,1,benign
                                                            NA
                                                                       NA
## 4
NA
            1017023,4,1,1,3,2,1,3,1,1,benign
## 5
                                                            NΑ
                                                                       NA
NA
## 6
      1017122,8,10,10,8,7,10,9,7,1,malignant
                                                            NA
                                                                       NA
NA
## 7
           1018099,1,1,1,1,2,10,3,1,1,benign
                                                            NA
                                                                       NA
NA
            1018561,2,1,2,1,2,1,3,1,1,benign
                                                                       NA
## 8
                                                            NA
NA
## 9
            1033078,2,1,1,1,2,1,1,1,5,benign
                                                            NA
                                                                       NA
NA
            1033078,4,2,1,1,2,1,2,1,1,benign
## 10
                                                            NA
                                                                       NA
NA
## 11
            1035283,1,1,1,1,1,1,3,1,1,benign
                                                                       NA
                                                            NA
NA
            1036172,2,1,1,1,2,1,2,1,1,benign
## 12
                                                            NA
                                                                       NA
NA
          1041801,5,3,3,3,2,3,4,4,2,maligant
## 13
                                                            NA
                                                                       NA
NA
## 14
            1043999,1,1,1,1,2,3,3,1,1,benign
                                                            NA
                                                                       NA
```

NA ##	15	1044572,8,7,5,10,7,9,5,5,4,maligant	NA	NA
NA ##	16	1047630,7,4,6,4,6,1,4,3,1,maligant	NA	NA
NA ##	17	1048672,4,1,1,1,2,1,2,1,1,benign	NA	NA
NA ##	18	1049815,4,1,1,1,2,1,3,1,1,benign	NA	NA
NA ##		1050670,10,7,7,6,4,10,4,1,2,maligant	NA	NA
NA		, , , , , , , , , , , ,		
## NA		1050718,6,1,1,1,2,1,3,1,1,benign	NA	NA
## NA	21	1054590,7,3,2,10,5,10,5,4,4,maligant	NA	NA
## NA	22	1054593,10,5,5,3,6,7,7,10,1,maligant	NA	NA
## NA	23	1056784,3,1,1,1,2,1,2,1,1,benign	NA	NA
## NA	24	1057013,8,4,5,1,2,NA,7,3,1,maligant	NA	NA
## NA	25	1059552,1,1,1,1,2,1,3,1,1,benign	NA	NA
##	26	1065726,5,2,3,4,2,7,3,6,1,maligant	NA	NA
NA ##	27	1066373,3,2,1,1,1,1,2,1,1,benign	NA	NA
NA ##	28	1066979,5,1,1,1,2,1,2,1,1,benign	NA	NA
NA ##	29	1067444,2,1,1,1,2,1,2,1,1,benign	NA	NA
NA ##	30	1070935,1,1,3,1,2,1,1,1,1,benign	NA	NA
NA ##	31	1070935,3,1,1,1,1,1,2,1,1,benign	NA	NA
NA ##	32	1071760,2,1,1,1,2,1,3,1,1,benign	NA	NA
NA ##	33	1072179,10,7,7,3,8,5,7,4,3,maligant	NA	NA
NA ##	34	1074610,2,1,1,2,2,1,3,1,1,benign	NA	NA
NA ##	35	1075123,3,1,2,1,2,1,2,1,1,benign	NA	NA
NA ##	36	1079304,2,1,1,1,2,1,2,1,1,benign	NA	NA
NA ##		1080185,10,10,10,8,6,1,8,9,1,maligant	NA NA	NA
NA				
## NA		1081791,6,2,1,1,1,7,1,1,benign	NA	NA
##	39	1084584,5,4,4,9,2,10,5,6,1,maligant	NA	NA

```
NA
## 40
           1091262,2,5,3,3,6,7,7,5,1,maligant
                                                                NA
                                                                           NA
NA
## 41
            1096800,6,6,6,9,6,NA,7,8,1,benign
                                                                           NA
                                                                NA
NA
## 42
          1099510,10,4,3,1,3,3,6,5,2,maligant
                                                                NA
                                                                           NA
NA
##
   43
       1100524,6,10,10,2,8,10,7,3,3,maligant
                                                                NA
                                                                           NA
NA
## 44
          1102573,5,6,5,6,10,1,3,1,1,maligant
                                                                           NA
                                                                NA
NA
   45 1103608,10,10,10,4,8,1,8,10,1,maligant
                                                                NA
                                                                           NA
NA
## 46
             1103722,1,1,1,1,2,1,2,1,2,benign
                                                                NA
                                                                           NA
NA
           1105257,3,7,7,4,4,9,4,8,1,maligant
## 47
                                                                NA
                                                                           NA
NA
## 48
             1105524,1,1,1,1,2,1,2,1,1,benign
                                                                           NA
                                                                NA
NA
## 49
             1106095,4,1,1,3,2,1,3,1,2,benign
                                                                           NA
                                                                NA
NA
##
      Adhesion.Epith. C.size.Bare. Nuclei.Bl.
                                                   Cromatin.Normal
## 1
                                    NA
                                                NA
                                                                  NA
## 2
                                                NA
                                                                  NA
                     NA
                                    NΑ
## 3
                     NΑ
                                                                  NA
                                    NA
                                                NA
## 4
                     NA
                                    NA
                                                NA
                                                                  NA
                     NA
                                    NΑ
                                                NA
                                                                  NA
## 5
                                                                  NA
## 6
                     NA
                                    NA
                                                NA
## 7
                     NA
                                    NA
                                                NA
                                                                  NA
## 8
                                    NA
                                                NA
                                                                  NA
                     NA
## 9
                     NA
                                    NA
                                                NA
                                                                  NA
                     NA
                                    NA
                                                                  NA
## 10
                                                NA
                                                                  NA
## 11
                     NA
                                    NA
                                                NA
## 12
                     NA
                                    NA
                                                NA
                                                                  NA
                     NA
                                                NA
                                                                  NA
## 13
                                    NA
## 14
                     NA
                                    NΑ
                                                NA
                                                                  NA
## 15
                     NA
                                    NA
                                                NA
                                                                  NA
## 16
                     NA
                                    NΑ
                                                NA
                                                                  NA
## 17
                     NA
                                    NA
                                                NA
                                                                  NA
## 18
                     NA
                                    NA
                                                NA
                                                                  NA
                                                                  NA
## 19
                     NA
                                    NA
                                                NΑ
## 20
                                                                  NA
                     NA
                                    NA
                                                NA
## 21
                                    NA
                                                                  NA
                     NA
                                                NA
                                                NA
                                                                  NA
## 22
                     NA
                                    NA
## 23
                     NA
                                    NA
                                                NA
                                                                  NA
                                                                  NA
## 24
                     NA
                                    NA
                                                NA
## 25
                     NA
                                    NΑ
                                                NA
                                                                  NA
## 26
                     NA
                                    NA
                                                NA
                                                                  NA
## 27
                     NA
                                    NA
                                                NA
                                                                  NA
## 28
                     NA
                                                NA
                                                                  NA
                                    NA
```

##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##		NA	NA	NA	NA	
##	49	NA	NA	NA	NA	
##		nucleoli.Mitoses.Class				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##		NA				
##	28	NA				

```
## 29
                             NA
## 30
                             NA
## 31
                             NA
## 32
                             NA
## 33
                             NA
## 34
                             NA
## 35
                             NA
## 36
                             NA
## 37
                             NA
## 38
                             NA
## 39
                             NA
## 40
                             NA
## 41
                             NA
## 42
                             NA
## 43
                             NA
## 44
                             NA
## 45
                             NA
## 46
                             NA
## 47
                             NA
## 48
                             NA
## 49
                             NA
```

#c. Compute the descriptive statistics using different packages. Find the values of: #c.1 Standard error of the mean for clump thickness.

```
num8c1.n <- length(breastcancer$`CL. thickness`)
num8c1.sd <- sd(breastcancer$`CL. thickness`)</pre>
```

#num8c1.se <- num8c1.sd/sqrt (breastcancer\$CL. thickness) #num8c1.se</pre>

#c.2 Coefficient of variability for Marginal Adhesion. Marginal\_Adhesion <- as.numeric(breastcancer\$V5) stat.desc(Marginal\_Adhesion)

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

```
Bare_Nuclei <- as.numeric(breastcancer$V7)</pre>
   stat.desc( Bare_Nuclei)
## Warning in min(x): no non-missing arguments to min; returning Inf
## Warning in max(x): no non-missing arguments to max; returning -Inf
## Warning in qt((0.5 + p/2), (Nbrval - 1)): NaNs produced
##
        nbr.val
                     nbr.null
                                     nbr.na
                                                      min
                                                                    max
range
##
              0
                            0
                                          0
                                                      Inf
                                                                   -Inf
Inf
                                                 SE.mean CI.mean.0.95
##
                       median
            sum
                                       mean
var
##
                           NA
                                                       NA
                                        NaN
                                                                    NaN
```

```
NA
## std.dev coef.var
## NA NA
```

#c.4 Mean and standard deviation for Bland Chromatin ``{r, include = TRUE} Bland\_Chromatin <- as.numeric(breastcancer\$V8)

mean(Bland\_Chromatin, na.rm = TRUE) sd(Bland\_Chromatin, na.rm = TRUE)
stat.desc( Bland\_Chromatin)

```
#c.5 Confidence interval of the mean for Uniformity of Cell Shape
``{r, include = TRUE}
cell_shape <- as.numeric(breastcancer$V4)
stat.desc(cell_shape )</pre>
```

#d. How many attributes? #e. Find the percentage of respondents who are malignant. Interpret the results.

```
describe(breastcancer$V11, na.rm =TRUE)
##
## NULL
```

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

```
library("AppliedPredictiveModeling")
data("abalone")
head(abalone)
     Type LongestShell Diameter Height WholeWeight ShuckedWeight
##
VisceraWeight
## 1
                 0.455
                          0.365 0.095
                                             0.5140
                                                            0.2245
        М
0.1010
## 2
                 0.350
                          0.265 0.090
                                             0.2255
                                                            0.0995
        Μ
0.0485
## 3
        F
                          0.420 0.135
                                             0.6770
                 0.530
                                                            0.2565
0.1415
## 4
                 0.440
                          0.365 0.125
                                             0.5160
                                                            0.2155
        Μ
0.1140
## 5
                 0.330
                          0.255 0.080
                                             0.2050
                                                            0.0895
        Ι
0.0395
        Ι
                 0.425
                          0.300 0.095
                                                            0.1410
## 6
                                             0.3515
0.0775
     ShellWeight Rings
## 1
           0.150
                    15
## 2
           0.070
                     7
## 3
           0.210
                     9
## 4
           0.155
                    10
           0.055
                     7
## 5
## 6
           0.120
                     8
```

```
summary(abalone)
              LongestShell
                                 Diameter
                                                   Height
                                                                  WholeWeight
##
    Type
##
    F:1307
             Min.
                    :0.075
                              Min.
                                     :0.0550
                                               Min.
                                                      :0.0000
                                                                 Min.
                                                                        :0.0020
   I:1342
##
             1st Qu.:0.450
                              1st Qu.:0.3500
                                               1st Qu.:0.1150
                                                                 1st Qu.:0.4415
##
   M:1528
             Median :0.545
                              Median :0.4250
                                               Median :0.1400
                                                                 Median :0.7995
##
             Mean
                    :0.524
                              Mean
                                     :0.4079
                                               Mean
                                                       :0.1395
                                                                 Mean
                                                                        :0.8287
##
                                                                 3rd Qu.:1.1530
             3rd Qu.:0.615
                              3rd Qu.:0.4800
                                               3rd Qu.:0.1650
##
                    :0.815
                                     :0.6500
                                                      :1.1300
                                                                 Max.
                                                                        :2.8255
             Max.
                              Max.
                                               Max.
##
    ShuckedWeight
                     VisceraWeight
                                        ShellWeight
                                                             Rings
##
    Min.
           :0.0010
                     Min.
                             :0.0005
                                       Min.
                                                                : 1.000
                                              :0.0015
                                                        Min.
##
    1st Qu.:0.1860
                     1st Qu.:0.0935
                                       1st Qu.:0.1300
                                                         1st Qu.: 8.000
   Median :0.3360
##
                     Median :0.1710
                                       Median :0.2340
                                                         Median : 9.000
##
    Mean
           :0.3594
                     Mean
                             :0.1806
                                       Mean
                                              :0.2388
                                                         Mean
                                                                : 9.934
##
    3rd Qu.:0.5020
                     3rd Qu.:0.2530
                                       3rd Qu.:0.3290
                                                         3rd Qu.:11.000
##
   Max.
           :1.4880
                     Max.
                             :0.7600
                                       Max.
                                              :1.0050
                                                         Max.
                                                                :29.000
library(xlsx)
write.xlsx("abalone","/cloud/project/Worksheet7a/abalone.xlsx")
```