DSCC_401_Fall_2021_Homework_5

R Markdown

Question 1

Part a:

```
X <- matrix(c(8, 1, 3, 5, -3, 11, -6, 7, 1), nrow = 3, ncol = 3, byrow = TRUE)
Y <- matrix(c(-1, 3, 8, 7, -6, 4, -8, 3, 1), nrow = 3, ncol = 3, byrow = TRUE)
trans_Y <- t(Y)
paste("Matrix algebra product")
```

```
## [1] "Matrix algebra product"
```

```
Z <- X %*% trans_Y
Z
```

```
##      [,1] [,2] [,3]
## [1,]   19   62  -58
## [2,]   74   97  -38
## [3,]   35  -80   70
```

Part b:

```
paste("Trace of multiplicative inverse of matrix Z:",sum(diag(solve(Z))))
```

```
## [1] "Trace of multiplicative inverse of matrix Z: 0.020995670995671"
```

Question 2

Part a:

```
sample <- c(117, 119, 114, 123, 132, 109, 113, 129, 124, 122, 124, 115, 138,
130, 135, 154, 118, 119, 134, 128, 116, 119, 132, 123, 120)
x_bar <- mean(sample)
mu <- 120
sd <- sd(sample)
n <- 25
crit_val <- 1.96

alpha <- 0.05
df <- n-1
t_val <- qt(alpha/2, df, lower.tail = FALSE)
```

```
## [1] "critical value: 1.96"
```

```
## [1] "t value: 2.06389856162803"
```

```
## [1] "t value > critical value"
```

```
## [1] "Null hypothesis is rejected."
```

Part b:

```
sample <- c(117, 119, 114, 123, 132, 109, 113, 129, 124, 122, 124, 115, 138,
130, 135, 154, 118, 119, 134, 128, 116, 119, 132, 123, 120)
x_bar <- mean(sample)
mu <- 120
sd <- sd(sample)
n <- 25
crit_val <- 2.576
alpha <- 0.01
df <- n-1
t_val <- qt(alpha/2, df, lower.tail = FALSE)
```

```
## [1] "critical value 2.576"
```

```
## [1] "t value: 2.79693950477446"
```

```
## [1] "t value > critical value"
```

```
## [1] "Null hypothesis is rejected."
```

Question 3

Part a:

```
df1 <- read.csv("summer2014.csv")
df2 <- read.csv("summer2021.csv")
```

```
summary(df1)
```

```
##     MONTH                DAY              HI              LO       
##  Length:92          Min.   : 1.00   Min.   :65.00   Min.   :48.00  
##  Class :character   1st Qu.: 8.00   1st Qu.:74.75   1st Qu.:54.00  
##  Mode  :character   Median :16.00   Median :79.00   Median :59.50  
##                     Mean   :15.84   Mean   :78.62   Mean   :59.09  
##                     3rd Qu.:23.25   3rd Qu.:83.00   3rd Qu.:64.00  
##                     Max.   :31.00   Max.   :90.00   Max.   :74.00  
##       AVG             DEP                HDD              CDD        
##  Min.   :57.0   Min.   :-10.0000   Min.   :0.0000   Min.   : 0.000  
##  1st Qu.:65.0   1st Qu.: -4.0000   1st Qu.:0.0000   1st Qu.: 0.000  
##  Median :69.5   Median :  1.0000   Median :0.0000   Median : 4.500  
##  Mean   :69.1   Mean   :  0.2609   Mean   :0.5652   Mean   : 4.663  
##  3rd Qu.:73.0   3rd Qu.:  3.2500   3rd Qu.:0.0000   3rd Qu.: 8.000  
##  Max.   :82.0   Max.   : 13.0000   Max.   :8.0000   Max.   :17.000  
##      PREC          
##  Length:92          
##  Class :character   
##  Mode  :character   
##                     
##                     
##
```

```
summary(df2)
```

```
##     MONTH                DAY             HIGH            LOW
##  Length:92          Min.   : 1.00   Min.   :64.00   Min.   :45.00
##  Class :character   1st Qu.: 8.00   1st Qu.:76.00   1st Qu.:57.00
##  Mode  :character   Median :16.00   Median :81.00   Median :62.00
##                     Mean   :15.84   Mean   :80.82   Mean   :62.01
##                     3rd Qu.:23.25   3rd Qu.:86.00   3rd Qu.:68.00
##                     Max.   :31.00   Max.   :94.00   Max.   :75.00
##
##       AVG              DEP               HDD              CDD
##  Min.   :56.00   Min.   :-13.400   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:67.00   1st Qu.: -4.125   1st Qu.:0.0000   1st Qu.: 2.000
##  Median :71.50   Median :  1.750   Median :0.0000   Median : 6.500
##  Mean   :71.43   Mean   :  1.130   Mean   :0.3696   Mean   : 6.989
##  3rd Qu.:76.00   3rd Qu.:  6.325   3rd Qu.:0.0000   3rd Qu.:11.000
##  Max.   :84.00   Max.   : 13.500   Max.   :9.0000   Max.   :19.000
##  NA's   :1
##      PREC                SNW
##  Length:92          Min.   :0
##  Class :character   1st Qu.:0
##  Mode  :character   Median :0
##                     Mean   :0
##                     3rd Qu.:0
##                     Max.   :0
##
```

```
## [1] "Maximum high temperature in 2014: 90"
```

```
## [1] "Maximum high temperature in 2021: 94"
```

Part b:

```
df3 <-  subset(df2, select = -c(SNW) )
summary(df3)
```

```
##     MONTH                DAY             HIGH           LOW
##  Length:92          Min.   : 1.00   Min.   :64.00   Min.   :45.00
##  Class :character   1st Qu.: 8.00   1st Qu.:76.00   1st Qu.:57.00
##  Mode  :character   Median :16.00   Median :81.00   Median :62.00
##                     Mean   :15.84   Mean   :80.82   Mean   :62.01
##                     3rd Qu.:23.25   3rd Qu.:86.00   3rd Qu.:68.00
##                     Max.   :31.00   Max.   :94.00   Max.   :75.00
##
##       AVG              DEP               HDD              CDD
##  Min.   :56.00   Min.   :-13.400   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:67.00   1st Qu.: -4.125   1st Qu.:0.0000   1st Qu.: 2.000
##  Median :71.50   Median :  1.750   Median :0.0000   Median : 6.500
##  Mean   :71.43   Mean   :  1.130   Mean   :0.3696   Mean   : 6.989
##  3rd Qu.:76.00   3rd Qu.:  6.325   3rd Qu.:0.0000   3rd Qu.:11.000
##  Max.   :84.00   Max.   : 13.500   Max.   :9.0000   Max.   :19.000
##  NA's   :1
##      PREC
##  Length:92
##  Class :character
##  Mode  :character
##
##
##
##
```

Part c:

```
names(df1) = names(df3)
names(df3)
```

```
## [1] "MONTH" "DAY"   "HIGH"  "LOW"   "AVG"   "DEP"   "HDD"   "CDD"   "PREC"
```

Part d:

```
df1$YEAR <- 2014
df3$YEAR <- 2021
head(df1)
```

```
##   MONTH DAY HIGH LOW AVG DEP HDD CDD PREC YEAR
## 1   JUN   1   82  48  65   3   0   0    0 2014
## 2   JUN   2   87  63  75  13   0  10    0 2014
## 3   JUN   3   83  57  70   7   0   5 0.38 2014
## 4   JUN   4   73  53  63   0   2   0    0 2014
## 5   JUN   5   65  49  57  -6   8   0    0 2014
## 6   JUN   6   74  51  63  -1   2   0    0 2014
```

```
head(df3)
```

```
##   MONTH DAY HIGH LOW  AVG  DEP HDD CDD PREC YEAR
## 1   JUN   1   77  54 65.5  1.8   0   1    0 2021
## 2   JUN   2   78  51 64.5  0.5   0   0 0.01 2021
## 3   JUN   3   74  62 68.0  3.7   0   3 0.17 2021
## 4   JUN   4   81  58 69.5  4.9   0   5    0 2021
## 5   JUN   5   87  65 76.0 11.1   0  11    0 2021
## 6   JUN   6   89  62 75.5 10.4   0  11    0 2021
```

Part e:

```
df4 <- rbind(df1, df3)
head(df4)
```

```
##   MONTH DAY HIGH LOW AVG DEP HDD CDD PREC YEAR
## 1   JUN   1   82  48  65   3   0   0    0 2014
## 2   JUN   2   87  63  75  13   0  10    0 2014
## 3   JUN   3   83  57  70   7   0   5 0.38 2014
## 4   JUN   4   73  53  63   0   2   0    0 2014
## 5   JUN   5   65  49  57  -6   8   0    0 2014
## 6   JUN   6   74  51  63  -1   2   0    0 2014
```

```
tail(df4)
```

```
##     MONTH DAY HIGH LOW  AVG  DEP HDD CDD PREC YEAR
## 179   AUG  26   89  69 79.0  9.5   0  14    T 2021
## 180   AUG  27   78  69 73.5  4.2   0   9    0 2021
## 181   AUG  28   89  63 76.0  6.9   0  11    0 2021
## 182   AUG  29   87  72 79.5 10.6   0  15 0.01 2021
## 183   AUG  30   83  67 75.0  6.3   0  10 0.13 2021
## 184   AUG  31   79  60 69.5  1.0   0   5    0 2021
```

Part f:

```
summary(df4)
```

```
##     MONTH                DAY             HIGH            LOW
##  Length:184         Min.   : 1.00   Min.   :64.00   Min.   :45.00
##  Class :character   1st Qu.: 8.00   1st Qu.:75.00   1st Qu.:55.75
##  Mode  :character   Median :16.00   Median :79.00   Median :60.00
##                     Mean   :15.84   Mean   :79.72   Mean   :60.55
##                     3rd Qu.:23.25   3rd Qu.:85.00   3rd Qu.:66.00
##                     Max.   :31.00   Max.   :94.00   Max.   :75.00
##
##       AVG              DEP               HDD              CDD
##  Min.   :56.00   Min.   :-13.4000   Min.   :0.0000   Min.   : 0.000
##  1st Qu.:66.00   1st Qu.: -4.0000   1st Qu.:0.0000   1st Qu.: 1.000
##  Median :70.00   Median :  1.1500   Median :0.0000   Median : 5.000
##  Mean   :70.26   Mean   :  0.6957   Mean   :0.4674   Mean   : 5.826
##  3rd Qu.:75.00   3rd Qu.:  4.8250   3rd Qu.:0.0000   3rd Qu.:10.000
##  Max.   :84.00   Max.   : 13.5000   Max.   :9.0000   Max.   :19.000
##  NA's   :1
##      PREC                YEAR
##  Length:184         Min.   :2014
##  Class :character   1st Qu.:2014
##  Mode  :character   Median :2018
##                     Mean   :2018
##                     3rd Qu.:2021
##                     Max.   :2021
##
```

```
## [1] "There is 1 missing value in the AVG column"
```

```
df4[is.na(df4$AVG), ] <- median(df4$AVG, na.rm = TRUE)
summary(df4)
```

```
##     MONTH                DAY             HIGH            LOW
##  Length:184         Min.   : 1.00   Min.   :64.00   Min.   :45.00
##  Class :character   1st Qu.: 8.00   1st Qu.:75.00   1st Qu.:55.75
##  Mode  :character   Median :16.00   Median :79.00   Median :60.50
##                     Mean   :16.11   Mean   :79.65   Mean   :60.61
##                     3rd Qu.:24.00   3rd Qu.:85.00   3rd Qu.:66.00
##                     Max.   :70.00   Max.   :94.00   Max.   :75.00
##       AVG              DEP               HDD              CDD
##  Min.   :56.00   Min.   :-13.400   Min.   : 0.0000   Min.   : 0.000
##  1st Qu.:66.00   1st Qu.: -4.000   1st Qu.: 0.0000   1st Qu.: 1.000
##  Median :70.00   Median :  1.300   Median : 0.0000   Median : 5.000
##  Mean   :70.26   Mean   :  1.070   Mean   : 0.8478   Mean   : 6.179
##  3rd Qu.:75.00   3rd Qu.:  4.925   3rd Qu.: 0.0000   3rd Qu.:10.000
##  Max.   :84.00   Max.   : 70.000   Max.   :70.0000   Max.   :70.000
##      PREC                YEAR
##  Length:184         Min.   :  70
##  Class :character   1st Qu.:2014
##  Mode  :character   Median :2014
##                     Mean   :2007
##                     3rd Qu.:2021
##                     Max.   :2021
```

```
## [1] "As can be observed above, there are no missing values in the AVG column anymor
e."
```

Part g:

```
jun <- df4[(df4$MONTH=="JUN"), "HIGH"]
jul <- df4[(df4$MONTH=="JUL"), "HIGH"]
```

```
## [1] "Variation in high temperature in June:  49.4938632378726"
```

```
## [1] "Variation in high temperature in July:  32.2815970386039"
```

```
## [1] "The month of june has a greater variation in the high temperatures."
```

```
## [1] "Difference in standard deviation for the high\ntemperatures for the two months:
1.35349787667144"
```

Part h:

```
df_jul <- df2[(df2$MONTH=="JUL"),]
#df_jul[(df_jul$PREC=="T"), "PREC"] <- 0
df_jul[(df_jul$PREC=="T"), "PREC"]
```

```
## [1] "T" "T"
```

```
df_jul$PREC <- as.numeric(gsub("T", 0, df_jul$PREC))
df_jul
```

```
##    MONTH DAY HIGH LOW  AVG   DEP HDD CDD PREC SNW
## 31   JUL   1   78  59 68.5  -2.8   0   4 0.56   0
## 32   JUL   2   69  59 64.0  -7.4   1   0 0.83   0
## 33   JUL   3   76  61 68.5  -3.1   0   4 0.30   0
## 34   JUL   4   75  59 67.0  -4.7   0   2 0.02   0
## 35   JUL   5   90  56 73.0   1.2   0   8 0.00   0
## 36   JUL   6   86  71 78.5   6.5   0  14 0.05   0
## 37   JUL   7   72  62 67.0  -5.1   0   2 0.57   0
## 38   JUL   8   78  62 70.0  -2.2   0   5 0.36   0
## 39   JUL   9   73  63 68.0  -4.3   0   3 0.28   0
## 40   JUL  10   75  61 68.0  -4.4   0   3 0.04   0
## 41   JUL  11   71  59 65.0  -7.4   0   0 0.43   0
## 42   JUL  12   76  62 69.0  -3.5   0   4 0.24   0
## 43   JUL  13   89  67 78.0   5.4   0  13 0.20   0
## 44   JUL  14   82  68 75.0   2.4   0  10 0.42   0
## 45   JUL  15   87  63 75.0   2.4   0  10 0.00   0
## 46   JUL  16   76  67 71.5  -1.2   0   7 0.05   0
## 47   JUL  17   67  59 63.0  -9.7   2   0 3.03   0
## 48   JUL  18   79  62 70.5  -2.2   0   6 0.02   0
## 49   JUL  19   83  63 73.0   0.3   0   8 0.00   0
## 50   JUL  20   85  64 74.5   1.8   0  10 0.12   0
## 51   JUL  21   75  61 68.0  -4.7   0   3 0.00   0
## 52   JUL  22   76  56 66.0  -6.7   0   1 0.00   0
## 53   JUL  23   78  57 67.5  -5.2   0   3 0.00   0
## 54   JUL  24   80  57 68.5  -4.1   0   4 0.00   0
## 55   JUL  25   85  69 77.0   4.4   0  12 0.02   0
## 56   JUL  26   87  63 75.0   2.4   0  10 0.00   0
## 57   JUL  27   80  62 71.0  -1.5   0   6 0.04   0
## 58   JUL  28   75  59 67.0  -5.5   0   2 0.00   0
## 59   JUL  29   76  57 66.5  -5.9   0   2 0.08   0
## 60   JUL  30   69  55 62.0 -10.4   3   0 0.00   0
## 61   JUL  31   73  50 61.5 -10.8   3   0 0.00   0
```

```
summary(df_jul)
```

```
##     MONTH                DAY             HIGH           LOW       
##  Length:31          Min.   : 1.0   Min.   :67.0   Min.   :50.00  
##  Class :character   1st Qu.: 8.5   1st Qu.:75.0   1st Qu.:59.00  
##  Mode  :character   Median :16.0   Median :76.0   Median :61.00  
##                     Mean   :16.0   Mean   :78.1   Mean   :61.06  
##                     3rd Qu.:23.5   3rd Qu.:82.5   3rd Qu.:63.00  
##                     Max.   :31.0   Max.   :90.0   Max.   :71.00  
##       AVG              DEP               HDD              CDD       
##  Min.   :61.50   Min.   :-10.800   Min.   :0.0000   Min.   : 0.000  
##  1st Qu.:67.00   1st Qu.: -5.350   1st Qu.:0.0000   1st Qu.: 2.000  
##  Median :68.50   Median : -3.500   Median :0.0000   Median : 4.000  
##  Mean   :69.58   Mean   : -2.774   Mean   :0.2903   Mean   : 5.032  
##  3rd Qu.:73.00   3rd Qu.:  0.750   3rd Qu.:0.0000   3rd Qu.: 8.000  
##  Max.   :78.50   Max.   :  6.500   Max.   :3.0000   Max.   :14.000  
##       PREC              SNW       
##  Min.   :0.0000   Min.   :0  
##  1st Qu.:0.0000   1st Qu.:0  
##  Median :0.0400   Median :0  
##  Mean   :0.2471   Mean   :0  
##  3rd Qu.:0.2900   3rd Qu.:0  
##  Max.   :3.0300   Max.   :0
```

```
## [1] "average daily rainfall for the period July 1, 2021 – July 31, 2021: 0.2471"
```

```
## [1] "standard deviation for the period July 1, 2021 – July 31, 2021:  0.5598404150493
07"
```

```
## [1] "I replaced the T with 0 in the dataframe."
```