QUIZ 1 - QUESTION 4

Amanjit Gill

October 31, 2022

```
library(GGally)
```

```
## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(MVN)

dev.new(width = 2.5, height = 2.5, unit = "in")

b <- as_tibble(read.csv("bridges.csv"))
head(b)
```

```
## # A tibble: 6 x 5
##   river erected purpose  length material
##   <chr>   <int> <chr>     <int> <chr>
## 1 A        1819 HIGHWAY    1037 WOOD
## 2 A        1837 HIGHWAY    1000 WOOD
## 3 A        1840 HIGHWAY     990 WOOD
## 4 A        1844 AQUEDUCT   1000 IRON
## 5 M        1846 HIGHWAY    1500 IRON
## 6 A        1851 HIGHWAY    1000 WOOD
```

```
# PART A - pairwise plots and normality test

b.subset <- b %>% select(erected, length)
b.subset
```

```
## # A tibble: 81 x 2
##    erected length
##      <int>  <int>
##  1    1819   1037
```

```
##  2    1837   1000
##  3    1840    990
##  4    1844   1000
##  5    1846   1500
##  6    1851   1000
##  7    1856   1200
##  8    1859   1030
##  9    1863   1000
## 10    1864   1200
## # ... with 71 more rows
```

```
ggpairs(b.subset)

mvn(b.subset, mvnTest = "mardia", univariateTest = "SW")
```

```
## $multivariateNormality
##              Test        Statistic              p value Result
## 1 Mardia Skewness 44.4848070981332 5.08777517344877e-09     NO
## 2 Mardia Kurtosis 3.78231712161238 0.000155375239301536     NO
## 3             MVN             <NA>                 <NA>     NO
##
## $univariateNormality
##           Test  Variable Statistic   p value Normality
## 1 Shapiro-Wilk  erected     0.9832  0.3674      YES
## 2 Shapiro-Wilk  length      0.8188  <0.001      NO
##
## $Descriptives
##          n     Mean   Std.Dev Median  Min  Max 25th 75th       Skew   Kurtosis
## erected 81 1910.222  34.76852   1910 1819 1978 1890 1931 -0.3205326 -0.3713876
## length  81 1567.469 747.49152   1300  804 4558 1000 2000  1.6723769  3.1325645
```

```
# PART B - Hypothesis Test

alpha <- 0.05

# separate into separate sets by river (M or A)
# river 1 = M, river 2 = A
x1 <- b[b$river == 'M', ] %>% select(erected, length)
x2 <- b[b$river == 'A', ] %>% select(erected, length)

p <- ncol(x1)

(n1 <- nrow(x1))
```

```
## [1] 32
```

```
(n2 <- nrow(x2))
```

```
## [1] 37
```

```
xbar1 <- colMeans(x1)
S1 <- cov(x1)

xbar2 <- colMeans(x2)
S2 <- cov(x2)

(Sp <- ((n1 - 1)*S1 + (n2 - 1)*S2)/(n1+n2-2))
```

```
##           erected      length
## erected 1240.460   4091.154
## length  4091.154 411748.065
```

```r
# approach 1: compare T^2 with critical T^2

(Tsq <- t(xbar1-xbar2) %*% solve(Sp*(1/n1 + 1/n2)) %*% (xbar1-xbar2))
```

```
##          [,1]
## [1,] 1.196022
```

```r
(Tsq_crit <- qf(1-alpha, p, n1+n2-p-1) * p*(n1+n2-2)/(n1+n2-p-1))
```

```
## [1] 6.366864
```

```r
# approach 2: compare F with critical F

(F <- Tsq * (n1+n2-p-1)/(p*(n1+n2-2)))
```

```
##           [,1]
## [1,] 0.5890852
```

```r
(F_crit <- qf(1-alpha, p, n1+n2-p-1))
```

```
## [1] 3.135918
```

```r
# approach 3: compare p-value with critical p (0.05)

(prob_of_F <- pf(F, p, n1+n2-p-1, lower.tail = FALSE))
```

```
##           [,1]
## [1,] 0.5577251
```