HW4

Minh Luc, Devin Pham, Kyle Moore

Friday of Week 4, 04/22/2022

Contents

We all contributed equally for this homework.

Question 0

Member 1:

- Name: Minh Luc
- Student ID: A17209607

Member 2:

- Name: Kyle Moore
- Student ID: A14271413

Member 3:

- Name: Devin Pham
- Student ID: A17198936

```r
# install the packages if needed by using
# install.packages("...")
library(tidyr)
library(readr)
library(tidytuesdayR)
urlRemote <- 'https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/'
pathGithub <- 'data/2020/2020-07-28/'
fileName <- 'penguins.csv'
penguins <- paste0(urlRemote, pathGithub, fileName) %>% read.csv(header = TRUE)
dfr <- drop_na(as.data.frame(penguins))
head(dfr)
```

```
##   species    island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
## 1  Adelie Torgersen           39.1          18.7               181        3750
```

```
## 2  Adelie Torgersen           39.5          17.4                186        3800
## 3  Adelie Torgersen           40.3          18.0                195        3250
## 4  Adelie Torgersen           36.7          19.3                193        3450
## 5  Adelie Torgersen           39.3          20.6                190        3650
## 6  Adelie Torgersen           38.9          17.8                181        3625
##      sex year
## 1   male 2007
## 2 female 2007
## 3 female 2007
## 4 female 2007
## 5   male 2007
## 6 female 2007
```

Question 1

- **(a)**

```
qf(.95, df1=3, df2=6, lower.tail = FALSE)
```

```
## [1] 0.1118488
```

- **(b)**

```
p <- 4
n <- 15
T2 <- 10
pf(((n - p) / ((n-1) * p)) * T2, df1=p, df2=n-p, lower.tail = FALSE)
```

```
## [1] 0.169808
```

Question 2

- **(a)**

  Null hypothesis: $H_0 : \mu = \mu_0$, or $\mu = (44, 17, 100, 4107)$

  Alternative hypothesis: $H_1 : \mu \neq \mu_0$, or $\mu \neq (44, 17, 100, 4107)$

- **(b)**

```
library(DescTools)

X <- dfr[1:50, 3:6]
mu <- colMeans(X)
mu0 <- c(44, 17, 200, 4207)

print(mu)     # print sample mean vector
```

```
##    bill_length_mm     bill_depth_mm flipper_length_mm       body_mass_g
##            38.916            18.774           187.140          3706.500
```

```
n <- nrow(X)   # num data points
p <- ncol(X)   # num variables
S <- cov(X)    # sample covariance matrix
```

```
(Tsquare = n * t(mu-mu0) %*% solve(S) %*% (mu-mu0)) # not scaled T^2 stat
```

```
##           [,1]
## [1,] 995.4982
```

```
HotellingsT2Test(x=X,mu=mu0, test="f")  # scaled T^2 stat
```

```
##
##  Hotelling's one sample T2-test
##
## data:  X
## T.2 = 233.64, df1 = 4, df2 = 46, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to c(44,17,200,4207)
```

- **(c)**

  We have $n = 50, p = 4$, therefore:

$$\frac{(n-1)p}{n-p} F_{p,n-p} = \frac{(50-1)4}{50-4} F_{4,46} = 4.261 F_{4,46}$$
$$T^2 \sim 4.261 F_{4,46}$$

- **(d)**

  We found the p-value of 2.2e-16 using the Hotelling's one sample T^2 test above. This p-value≈ 0, and given a significance level of 0.05 and our high T^2 statistic, we can reject H_0.

- **(e)**

```
alpha <- .1

F.quantile <- qf(alpha, p, n-p, lower.tail = FALSE)
print(F.quantile)
```

```
## [1] 2.071244
```

```
critical_value <- (n-1)*p/(n-p)*F.quantile
print(critical_value)
```

```
## [1] 8.8253
```

  We reject when $T^2 > 8.8253$

- **(f)**

  The 90% confidence ellipsoid for μ is:

$$\begin{aligned} R(X) &= \left\{ \mu : T^2(\mu) \leq \frac{(n-1)p}{(n-p)F_{p,n-p}(\alpha)} \right\} \\ &= \left\{ \mu : n(\bar{\boldsymbol{X}} - \mu)^T \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \mu) \leq \frac{(49)4}{(46)F_{4,46}(.1)} \right\} \\ &= \left\{ \mu : 50(\bar{\boldsymbol{X}} - \mu)^T \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \mu) \leq 8.8253 \right\} \end{aligned}$$

  Where

$$\bar{\boldsymbol{X}} = \begin{bmatrix} 38.916 & 18.774 & 187.140 & 3706.500 \end{bmatrix}$$

$$
\boldsymbol{S} = \begin{bmatrix} 6.2728000 & 0.9271592 & 2.9160816 & 451.2714286 \\ 0.9271592 & 1.3868612 & 3.7057551 & 310.1724490 \\ 2.916082 & 3.705755 & 41.796327 & 1234.785714 \\ 451.2714 & 310.1724 & 1234.7857 & 202010.4592 \end{bmatrix}
$$

- **(g)**

  The confidence intervals (CI) for each component are

$$
\begin{aligned}
\mu_j &\in \bar{\boldsymbol{x}}_j \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \times \frac{s_j}{\sqrt{n}}, j = 1, ..., p \\
\mu_1 &\in 38.916 \pm \sqrt{\frac{(49)4}{(46)} F_{4,46}(0.1)} \times \frac{\sqrt{6.278}}{\sqrt{50}} \\
&\in 38.916 \pm 8.8253 \times 0.354 \\
&\in 38.916 \pm 3.127 \\
&\in [35.789, 42.043] \\
\mu_2 &\in 18.774 \pm \sqrt{\frac{(49)4}{(46)} F_{4,46}(0.1)} \times \frac{\sqrt{1.387}}{\sqrt{50}} \\
&\in 18.774 \pm 8.8253 \times 0.167 \\
&\in 18.774 \pm 1.470 \\
&\in [17.304, 20.244] \\
\mu_3 &\in 187.14 \pm \sqrt{\frac{(49)4}{(46)} F_{4,46}(0.1)} \times \frac{\sqrt{41.796}}{\sqrt{50}} \\
&\in 187.14 \pm 8.8253 \times 0.914 \\
&\in 187.14 \pm 8.066 \\
&\in [179.074, 195.206] \\
\mu_4 &\in 3706.5 \pm \sqrt{\frac{(49)4}{(46)} F_{4,46}(0.1)} \times \frac{\sqrt{202010.459}}{\sqrt{50}} \\
&\in 3706.5 \pm 8.8253 \times 63.563 \\
&\in 3706.5 \pm 560.959 \\
&\in [3145.541, 4267.459]
\end{aligned}
$$

- **(h)**

```
bonferroni_t <- qt((.1/(2*4)), df=49, lower.tail = FALSE)
bonferroni_t
```

```
## [1] 2.312375
```

```
for (x in 1:4) {
  print(x)  # which mu iteration it is on
  print(mu[x])  # value of current mu
  print(S[x,x]) # variance of current mu
  print("left interval")
  print(mu[x] - bonferroni_t * sqrt(S[x,x]/n))  # find right interval
  print("right interval")
  print(mu[x] + bonferroni_t * sqrt(S[x,x]/n))  # find left interval
}
```

```
## [1] 1
## bill_length_mm
##         38.916
## [1] 6.2728
## [1] "left interval"
## bill_length_mm
##       38.09696
## [1] "right interval"
## bill_length_mm
##       39.73504
## [1] 2
## bill_depth_mm
##        18.774
## [1] 1.386861
## [1] "left interval"
## bill_depth_mm
##      18.38889
## [1] "right interval"
## bill_depth_mm
##      19.15911
## [1] 3
## flipper_length_mm
##            187.14
## [1] 41.79633
## [1] "left interval"
## flipper_length_mm
##          185.0258
## [1] "right interval"
## flipper_length_mm
##          189.2542
## [1] 4
## body_mass_g
##      3706.5
## [1] 202010.5
## [1] "left interval"
## body_mass_g
##    3559.519
## [1] "right interval"
## body_mass_g
##    3853.481
```

$$
\begin{aligned}
\mu_1 &\in [38.097, 39.735] \\
\mu_2 &\in [18.389, 19.159] \\
\mu_3 &\in [185.026, 189.254] \\
\mu_4 &\in [3559.519, 3853.481]
\end{aligned}
$$