HW2

Minh Luc, Devin Pham, Kyle Moore

Friday of Week 2, 04/08/2022

Contents

Question 0 1

Question 1 1

Question 2 2

Question 3 3

Question 4 3

Question 5 3

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

Question 1

- **(a)**

```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
```

- **(b)**

```
nrow(dfr) # number of rows
```

```
## [1] 333
```

```
ncol(dfr) # number of columns
```

```
## [1] 8
```

There are 333 rows and 8 columns in the dataframe(dfr).

Question 2

- **Find the mean vector, covariance matrix and correlation matrix of X:**

```
X <- dfr[,3:6] # assign all rows, but only columns 3-6 to X

colMeans(X) # mean vector containing the means for each column in X
```

```
##    bill_length_mm     bill_depth_mm flipper_length_mm       body_mass_g
##          43.99279          17.16486         200.96697        4207.05706
```

```
cov(X) # compute the covariance matrix of X
```

```
##                   bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
## bill_length_mm         29.906333     -2.462091          50.05819   2595.6233
## bill_depth_mm          -2.462091      3.877888         -15.94725   -748.4561
## flipper_length_mm      50.058195    -15.947248         196.44168   9852.1916
## body_mass_g          2595.623304   -748.456122        9852.19165 648372.4877
```

```
cor(X) # compute the correlation matrix of X
```

```
##                   bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
## bill_length_mm         1.0000000    -0.2286256         0.6530956   0.5894511
## bill_depth_mm         -0.2286256     1.0000000        -0.5777917  -0.4720157
## flipper_length_mm      0.6530956    -0.5777917         1.0000000   0.8729789
## body_mass_g            0.5894511    -0.4720157         0.8729789   1.0000000
```

- The variance-covariance matrix is a symmetric matrix that represents how the variables are correlated: positively correlated, negatively correlated, or uncorrelated. The diagonal represents the variance of each variable itself. It is symmetric due to the fact that $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- The correlation matrix is a standardized version of the variance-covariance matrix that represents the strength of the correlation between two variables where $-1 \leq correlation \leq 1$. Entries closer to 1 are more strongly positively correlated, those closer to -1 are strongly negatively correlated, and those near 0 are weakly or uncorrelated. The diagonals are all 1's because each variable is completely correlated with itself.

Question 3

- **(a)**:
- **(b)**:
- **(c)**:
- **(d)**:

Question 4

Question 5

- **(a)**:
- **(b)**:
- **(c)**:
- **(d)**:
