GENE613 - Homework 1

Daniel Osorio - dcosorioh@tamu.edu
Department of Veterinary Integrative Biosciences
Texas A&M University

0. Read the data

   ```
   > hw1Data <- read.delim(file = "../data/hw1data18.txt",
   +                       stringsAsFactors = FALSE
   +                       )
   ```

1. Give the minimum, the maximum, the mean, the variance, the standard deviation, and the coefficient of variation for the ADG and DDMI.

   ```
   > hw1pt1 <- function(data) {
   +   return(c(
   +     min = min(data),
   +     max = max(data),
   +     mean = mean(data),
   +     var = var(data),
   +     stdDev = sd(data),
   +     CV = (sd(data) / mean(data))
   +   ))
   + }
   > apply(hw1Data[, 3:4], 2, hw1pt1)
   ```

   ```
                  ADG        DDMI
   min     0.89000000  4.8300000
   max     1.93000000 10.7500000
   mean    1.46370370  8.0744444
   var     0.05047037  1.5012179
   stdDev  0.22465612  1.2252420
   CV      0.15348469  0.1517432
   ```

2. Calculate the covariance of ADG and DDMI, the correlation of ADG and DDMI, the regression of ADG on DDMI, and the regression of DDMI on ADG, and explain the differences between these values.

   ```
   > hw1pt2 <- function(data) {
   +   return(c(
   +     covariance = with(data, cov(ADG, DDMI)),
   +     correlation = with(data, cor(ADG, DDMI)),
   +     b1_ADGonDDMI = (coef(with(hw1Data, lm(ADG ~ DDMI)))[[2]]),
   +     b1_DDMIonADG = (coef(with(data, lm(DDMI ~ ADG)))[[2]])
   +   ))
   ```

```
+ }
> hw1pt2(hw1Data)

  covariance   correlation b1_ADGonDDMI b1_DDMIonADG
   0.2544098     0.9242592    0.1694689     5.0407759
```

Covariance: is a measure of how much two random variables vary together. It is measured as the mean value of the product of the deviations of two variables from their respective means. In this case, the covariance is positive, which means that the majority of the greater values of ADG mainly correspond with greater values of the DDMI variable.

Correlation: is a measure of the degree to which changes to the value of one variable predict the change in the value of another. In this case, the correlation es positive and close to 1 which means that if we know the changes in one variable, is highly probable to predict the changes in the another.

β_1: is the degree of change in the outcome variable for every 1-unit of change in the predictor variable.

3. Execute the following code (you may need to modify the trait names to the right of the $\leftarrow$ depending upon how you input your data) in R

```
> no3.adg.ddmi <-  with(hw1Data, lm(ADG~DDMI))
> no3.ddmi.adg <-  with(hw1Data, lm(DDMI~ADG))
> summary(no3.adg.ddmi)

Call:
lm(formula = ADG ~ DDMI)

Residuals:
      Min        1Q    Median        3Q       Max
-0.208707 -0.056347 -0.004309  0.058828  0.182275

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09534    0.11429   0.834    0.412
DDMI         0.16947    0.01400  12.105 5.96e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08746 on 25 degrees of freedom
Multiple R-squared:  0.8543,        Adjusted R-squared:  0.8484
F-statistic: 146.5 on 1 and 25 DF,  p-value: 5.957e-12

> summary(no3.ddmi.adg)

Call:
lm(formula = DDMI ~ ADG)

Residuals:
     Min       1Q   Median       3Q      Max
-1.10455 -0.36099 -0.01944  0.33891  1.09075
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6962     0.6164    1.13    0.269
ADG           5.0408     0.4164   12.11 5.96e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.477 on 25 degrees of freedom
Multiple R-squared:  0.8543,        Adjusted R-squared:  0.8484
F-statistic: 146.5 on 1 and 25 DF,  p-value: 5.957e-12
```

Explain the results by executing this code

Coefficients: The beta coefficient is the degree of change in the outcome variable for every 1-unit of change in the predictor variable.

R squared: The coefficient of determination is the percent of variance in the outcome variable that is explained by the set of predictor variables.

F stat: Is a statistic that tests the amount of variance in the outcome variable explained by the generated model. If the F-value is statistically significant (typically $p < .05$) we say that the model explains a significant amount of variance in the outcome variable.

4. Calculate means and SD for sires. If you wanted to increase ADG, which sire would you choose to be a parent in the next generation? If you wanted to decrease DDMI, which sire should you choose?

```
> hw1pt4 <- function(sireID){
+    selectedSire <- hw1Data[with(hw1Data, sire %in% sireID),]
+    c(MEAN=apply(selectedSire[,3:4],2,mean),
+    SD=apply(selectedSire[,3:4],2,sd))
+ }
> sapply(unique(hw1Data$sire),hw1pt4)

               551G       432H      437J       297J
MEAN.ADG  1.3988889 1.4900000 1.458571 1.5312500
MEAN.DDMI 7.5655556 7.9366667 8.300000 8.5012500
SD.ADG    0.2060609 0.1216553 0.258162 0.2581493
SD.DDMI   1.1858342 0.7460786 1.185580 1.4159342
```

Q1: If you wanted to increase ADG, which sire would you choose to be a parent in the next generation? I would select the sire 432H which shows the second highest ADG value and the lowest variability between sires.

Q2: If you wanted to decrease DDMI, which sire should you choose? I would select the sire 432H which shows the second lowest DDMI value and the lowest variability between sires.