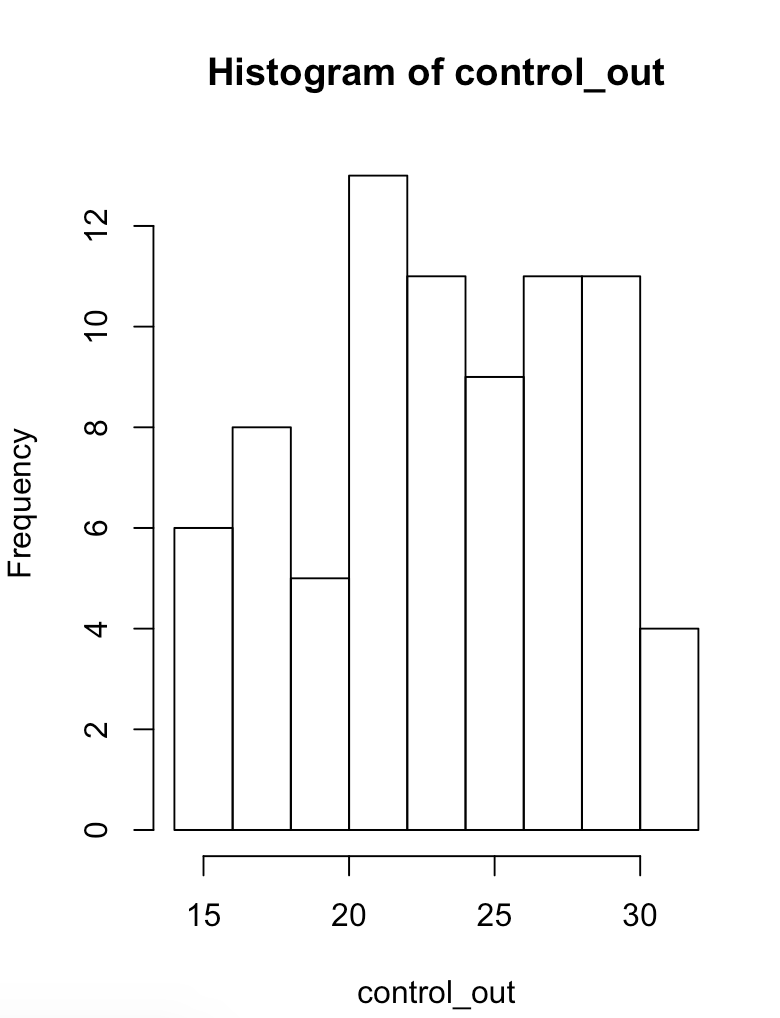
**INTRODUCTION TO STATISTICAL CONCEPTS**

**ASSIGNMENT 1B**

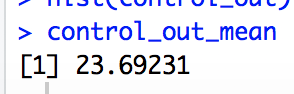
**ANGRY MOODS CASE QUESTIONS:**

**Ways to improve an Angry Mood: A Look at Gender and Sports Participation**

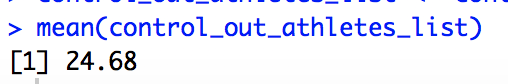
1. Plot a histogram of the distribution of the Control-Out scores.



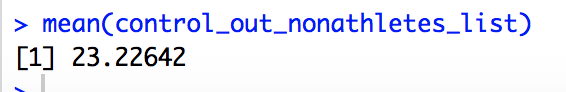
1. What is the overall mean Control-Out score? What is the mean Control-Out score for the athletes? What is the mean Control-Out score for the non-athletes?
   * Overall mean of control out score:



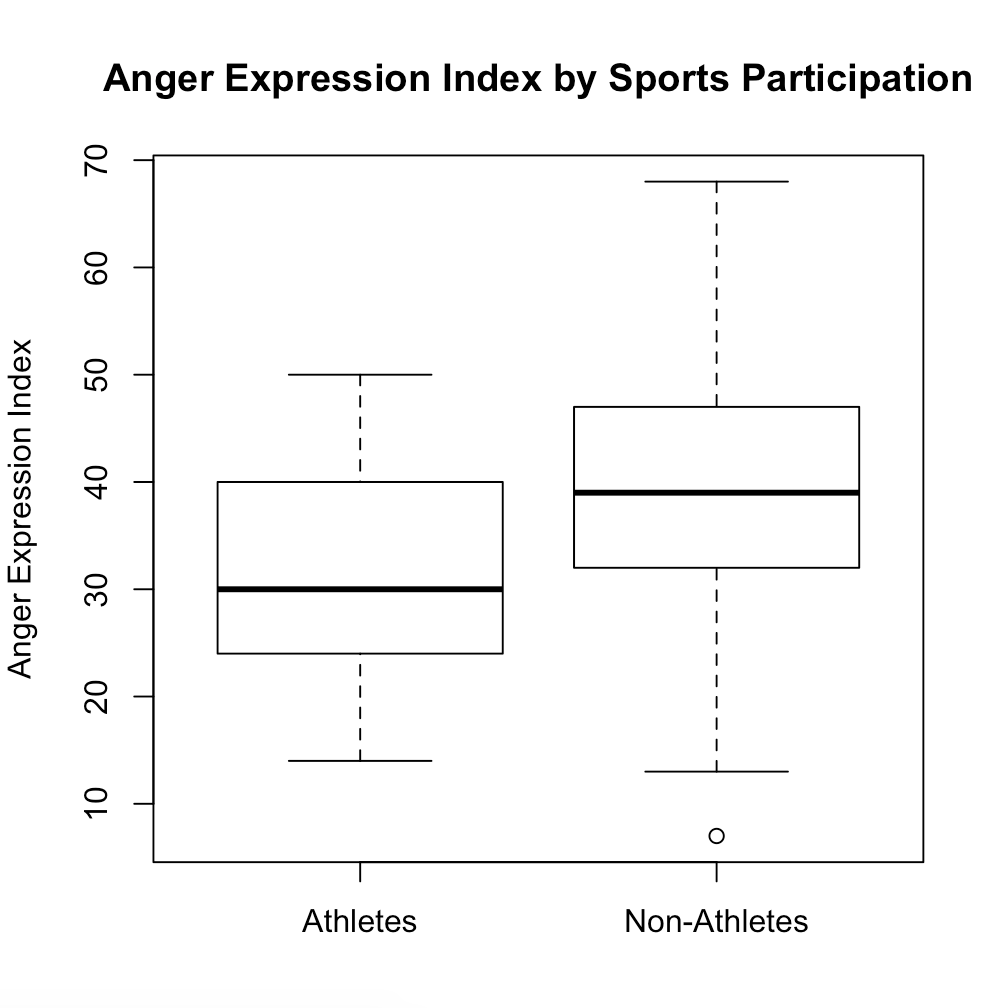
* + Mean for control out score for athletes:



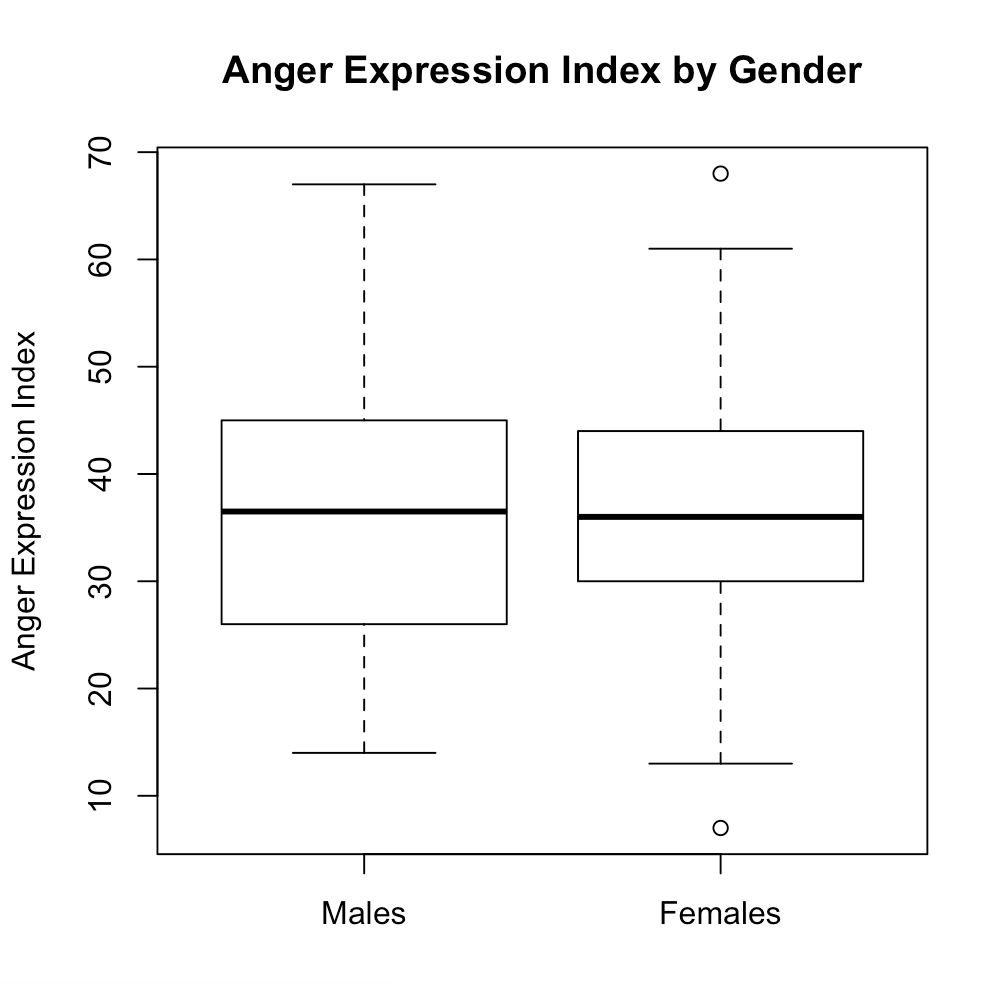
* + Mean for control out score for non-athletes:



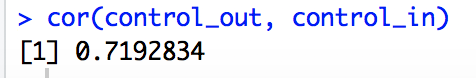
1. Plot parallel box plots of the Anger Expression Index by sports participation. Does it look like there are any outliers? Which group reported expressing more anger?
   * There seems to be one outlier in non-athletes, who has a low anger expression index despite not being an athlete.
   * The group that is expressing more anger is the non-athlete group, as evidenced by the anger expression index. The amount of anger experienced and expressed by people in this group exceeds how they control both outward and inward feelings or expressions (eg cooling off) of anger, as described by the formula for Anger Expression.



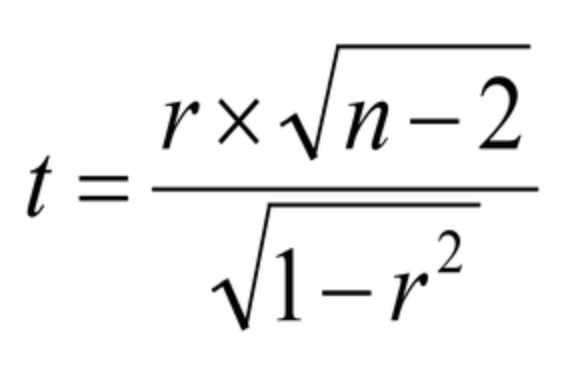
1. Plot parallel box plots of the Anger Expression Index by gender.



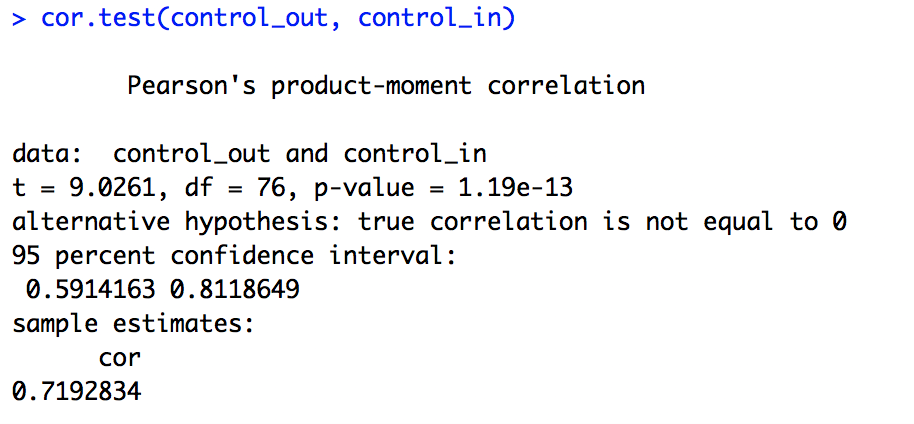
1. What is the correlation between the Control-In and Control-Out scores? Is this correlation statistically significant at the 0.01 level?
   * The correlation between control-in and control-out scores is

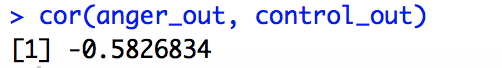


* + The correlation will be statistically significant at the 0.01 level, since the computed p-value from this correlation 0.719 and with N = 78 pairs of control in and control out scores is 0.00001. To get this number, we first needed to convert r (our correlation) to a t-score, through the formula:



* + From here, we then compute the p value at the 0.01 confidence interval. We can also do this in r to get the following, which does indeed tell us that this correlation will be significant at a 0.01 significant level:



1. Would you expect the correlation between the Anger-Out and Control-Out scores to be positive or negative? Compute this correlation.
   * I’d expect the correlation to be negative, people who express anger aggressively do not tend to control those outward expressions of angry feelings very well—and people who have a low anger out score (thus who do not express anger aggressively) will tend to have a higher control-out score (which means that they can control the outward expression of angry feelings).
   * 

Code:

|  |
| --- |
| #set wd getwd()  #read data moods <- read.csv("angry\_moods.csv", header=TRUE) moods  #pre-processing into df for further analysis angry\_df <- data.frame(moods) angry\_df  # Question 10 control\_out <- angry\_df$Control.Out hist(control\_out)  # Question 11 control\_out\_mean <- mean(control\_out) control\_out\_mean  control\_out\_athletes <- subset(angry\_df, Sports == 1, select = c("Control.Out", "Sports")) control\_out\_athletes control\_out\_athletes\_list <- control\_out\_athletes$Control.Out mean(control\_out\_athletes\_list)  control\_out\_nonathletes <- subset(angry\_df, Sports == 2, select = c("Control.Out", "Sports")) control\_out\_nonathletes  control\_out\_nonathletes\_list <- control\_out\_nonathletes$Control.Out mean(control\_out\_nonathletes\_list)  # Question 17 AE\_athletes <- subset(angry\_df, Sports == 1, select = c("Anger\_Expression")) AE\_athletes  AE\_nonathletes <- subset(angry\_df, Sports == 2, select = c("Anger\_Expression")) AE\_nonathletes  Anger\_exp <- angry\_df$Anger\_Expression Sport <- angry\_df$Sports  boxplot(Anger\_exp~Sport,         main="Anger Expression Index by Sports Participation",         names = c('Athletes', 'Non-Athletes'),         ylab = "Anger Expression Index") # is there a more efficient way to do this?  # Question 18 males <- subset(angry\_df, Gender == 1, select = c("Anger\_Expression")) males  females <- subset(angry\_df, Gender == 2, select = c("Anger\_Expression")) females  Anger\_exp <- angry\_df$Anger\_Expression gender <- angry\_df$Gender  boxplot(Anger\_exp~gender,         main="Anger Expression Index by Gender",         names = c('Males', 'Females'),         ylab = "Anger Expression Index")  # Question 20 control\_out <- angry\_df$Control.Out control\_in <- angry\_df$Control.In  cor(control\_out, control\_in) cor.test(control\_out, control\_in)  # Question 21 anger\_out <- angry\_df$Anger.Out cor(anger\_out, control\_out) |