Midterm 1

1. Given the results of the ANOVA, because the p value is less than 0.001, the intergroup differences are statistically significant. So, there is at least one area whose cotton candy levels are more dangerous.

```
1   data <- read.csv('testprob.csv', header=TRUE)
2
3
4   bubblegum <- data[data$Location=="Bubble Gum Room",]["CC.Level"]
5   chocolate <- data[data$Location=="Chocolate River",]["CC.Level"]
6   taffy <- data[data$Location=="Taffy Tank",]["CC.Level"]
7   caramel <- data[data$Location=="Caramel Chew Loo",]["CC.Level"]
8
9   nbubblegum <- bubblegum[!is.na(bubblegum)]
10  nchocolate <- chocolate[!is.na(chocolate)]
11  ntaffy <- taffy[!is.na(taffy)]
12  ncaramel <- caramel[!is.na(caramel)]
13
14  shapiro.test(nbubblegum)
15  shapiro.test(log(log(nchocolate+1)+1))
16  shapiro.test(ntaffy)
17  shapiro.test(ncaramel)
18
19  ndata <- rbind(nbubblegum, log(nchocolate), ntaffy, ncaramel)
20
21  #using the 'data' variable as opposed to the 'ndata' variable
22  #since running with 'ndata' yields the following results
23  ##Error in model.frame.default(formula = CC.Level ~ Location, data = ndata,   :
24  ## 'data' must be a data.frame, not a matrix or an array
25  test <- aov(CC.Level~Location, data=ndata)
26  summary(test)
```

2. (a) The true average difference in translation between dominant and nondominant arms for pitchers is 1.028 mm.

   (b) The true average difference in translation between dominant and nondominant arms for position players is 0.145 mm.

   (c) I would agree as the sde calculated for pitchers was 0.959, whereas the sde calcualted for position players was 0.367.

```
1   data <- read.csv('ex9-88.csv', header=TRUE)
2
3   posdom <- c(data["Pos_Dom_"][,])
4   posndt <- c(data["Pos_ND_T"][,])
5   pitdom <- c(data["PIT_Dom_"][,])
6   pitndt <- c(data["Pit_ND_T"][,])
7
8   npitdom <- pitdom[!is.na(pitdom)]
9   npitndt <- pitndt[!is.na(pitndt)]
10
11  diffpits = npitdom - npitndt
12  mdiffpits = mean(diffpits)
13  sddiffpits = sd(diffpits)/sqrt(length(diffpits))
14  sddiffpits
15  print('---')
16  tpits = mdiffpits/(sddiffpits*sqrt(length(diffpits)))
17  tpits
18  #diff pm z*sde
19  mdiffpits + qnorm(0.95)*sddiffpits
20  mdiffpits - qnorm(0.95)*sddiffpits
```

```
1   data <- read.csv('ex9-88.csv', header=TRUE)
2
3   posdom <- c(data["Pos_Dom_"][,])
4   posndt <- c(data["Pos_ND_T"][,])
5   pitdom <- c(data["PIT_Dom_"][,])
6   pitndt <- c(data["Pit_ND_T"][,])
7
8   nposdom <- pitdom[!is.na(posdom)]
9   nposndt <- pitndt[!is.na(posndt)]
10
11  diffpos = nposdom - nposndt
12  mdiffpos = mean(diffpos)
13  sddiffpos = sd(diffpos)/sqrt(length(diffpos))
14  sddiffpos
15  print('---')
16  tpos = mdiffpos/(sddiffpos*sqrt(length(diffpos)))
17  tpos
18  #diff pm z*sde
19  mdiffpos + qnorm(0.95)*sddiffpos
20  mdiffpos - qnorm(0.95)*sddiffpos
```

3. For this case, we would want to use a paired t-test, as the diets were randomly assigned, each population was normally distributed, and their variances were similar.

```
1  C <- c(15.00, 7.00, 2.44, 5.60, 3.63, 6.24, 4.12, 8.21, 3.90, 0.95)
2  B <- c(5.12, 9.38, 18.77, 15.03, 6.67, 7.91, 7.38, 15.09, 11.57, 8.98)
3
4  qqnorm(C)
5  qqline(C)
6  qqnorm(B)
7  qqline(B)
8
9  difff = C - B
10 sddiff = sd(difff)
11 mdiff = mean(difff)
12 n = length(difff)
13
14 t = mdiff/(sddiff/sqrt(n))
15 t
16 mdiff + t*(sddiff/sqrt(n))
17 mdiff - t*(sddiff/sqrt(n))
18
```
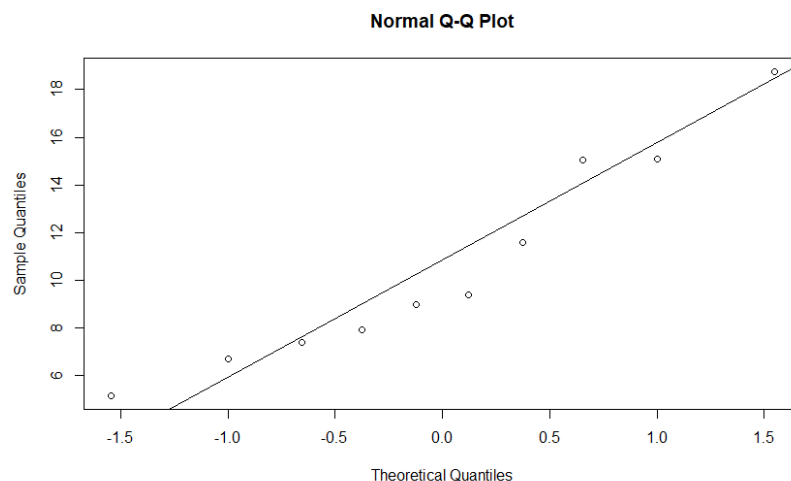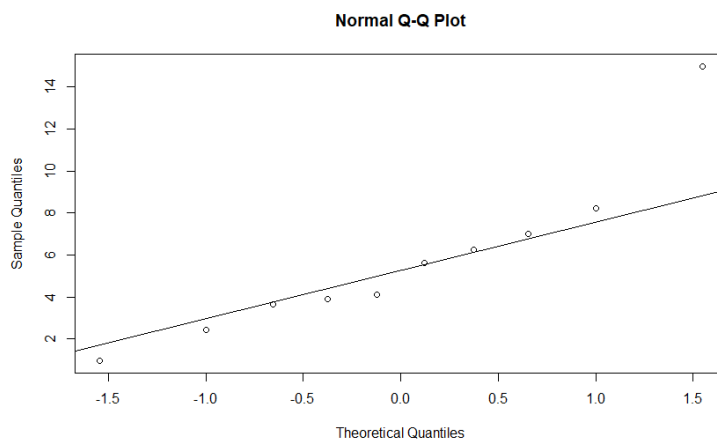

Figure 2: Proof of normality for population B.

Figure 3: Proof of normality for population C.

2

4. $\dfrac{\sigma_1^2}{\sigma_2^2} \cdot F_{1-\alpha/2, v_2, v_1} < \dfrac{\sigma_1^2}{\sigma_2^2} < \dfrac{\sigma_1^2}{\sigma_2^2} \cdot F_{\alpha/2, v_2, v_1}$