ASSIGNMENT 5

Brandon Lampe
STAT 527
Advanced Data Analysis I

October 20, 2014

1 REM with ethanol treatment:

Load and format data with:

```
# read table as it appears above, simply copy/paste
dat1 <- read.table(text="
          0 g/kg 88.6 73.2 91.4 68.0 75.2
          1 g/kg 63.0 53.9 69.2 50.1 71.5
          2 g/kg 44.9 59.5 40.2 56.3 38.7
          4 g/kg 31.0 39.6 45.3 25.2 22.7

", header=FALSE)
# extract only columns with data values
dat2 <- dat1[, 3:7]
# transpose the matrix so each treatment row is now in its own column
dat3 <- t(dat2)
# change matrix to a data.frame
rem <- as.data.frame(dat3)
# name the columns based on dosage
colnames(rem) <- c("0", "1", "2", "4")
# Convert to long format
rem.long <- melt(rem, variable.name = "dose", value.name = "minutes")
## No id variables; using all as measure variables
# add column to include units of dose
rem.long$Dose <- rep(NA, nrow(rem.long))
rem.long[(rem.long$dose == 0), "Dose"] <- "0 g/kg"
rem.long[(rem.long$dose == 1), "Dose"] <- "1 g/kg"
rem.long[(rem.long$dose == 2), "Dose"] <- "2 g/kg"
rem.long[(rem.long$dose == 4), "Dose"] <- "4 g/kg"
rem.long$Dose <- factor(rem.long$Dose)
```

1(a) **(10 pts) Make an appropriate graphical summary of the data to compare the groups. Compute, the sample means and standard deviations for the 4 treatment groups.**

```
#create box plot
dat1.p <- ggplot(rem.long, aes(x = Dose, y = minutes))
# plot global mean
dat1.p <- dat1.p + geom_hline(yintercept = mean(rem.long$minutes), color = "black",
                      linetype = "dashed", size = 0.3,  alpha = 0.5)
dat1.p <- dat1.p + geom_boxplot(size = 0.75, alpha = 0.5) # boxplot
dat1.p <- dat1.p + geom_point(position = position_jitter(w = 0.05, h = 0),
                      alpha = 0.5, size = 2)
dat1.p <- dat1.p + stat_summary(fun.y = mean, geom = "point", shape = 18,
                        size = 6, aes(color = Dose), alpha = 0.8)
dat1.p <- dat1.p + stat_summary(fun.data = "mean_cl_normal", geom = "errorbar",
                      width = .1, aes(color = Dose), alpha = 0.8)
```
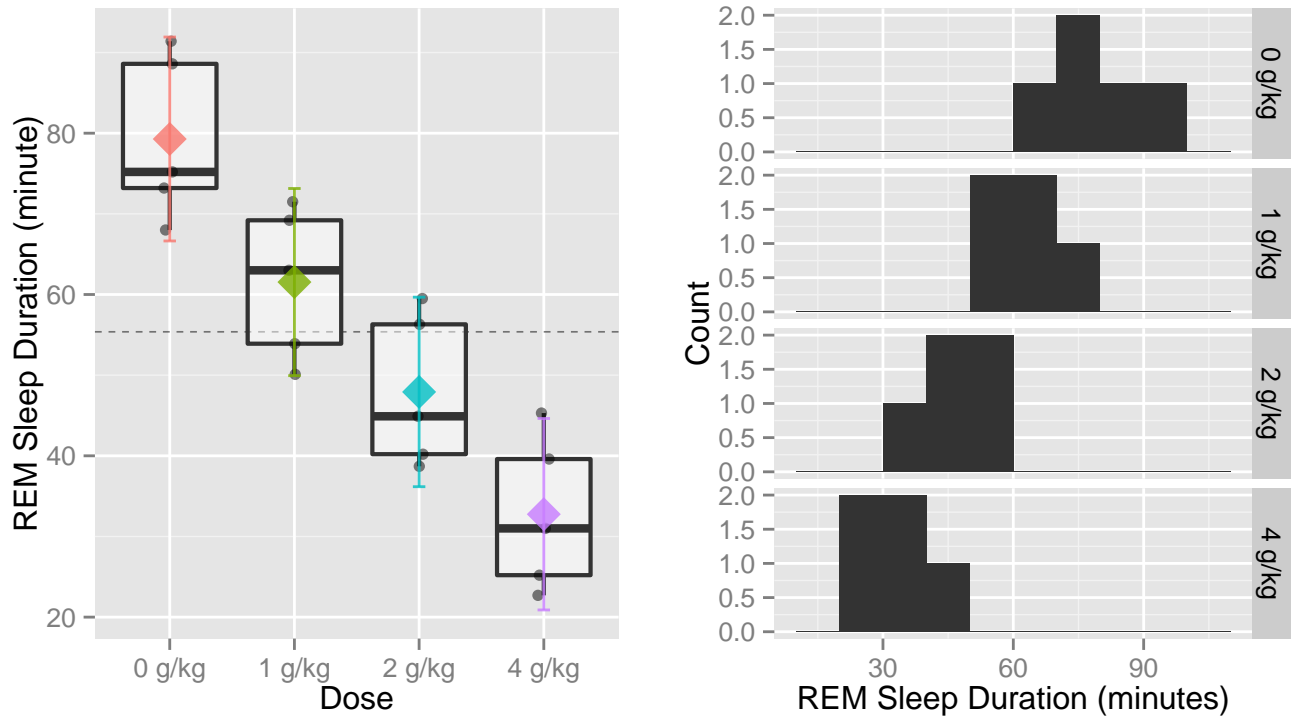
Figure 1: REM Sleep Duration Summary for Each Treatment of Ethanol (g of Ethanol per kg of Body Mass).

```r
dat1.p <- dat1.p + labs(y ="REM Sleep Duration (minute)",x = "Dose")

dat1.p <- dat1.p + guides(color = FALSE)
# create histograms
dat1.hist <- ggplot(rem.long, aes(x = minutes)) + geom_histogram(binwidth = 10)
dat1.hist <- dat1.hist + facet_grid(Dose ~ .)
dat1.hist <- dat1.hist+ labs(y = "Count", x = "REM Sleep Duration (minutes)")
# plot boxplot and historgram
grid.arrange(dat1.p, dat1.hist, ncol = 2)
```

```r
# global mean
Yglob <- mean(rem.long$minutes) # because each sample is the same size
Yglob
## [1] 55.375
stats <- data.frame(ddply(rem.long, .(dose), summarize, mean = round(mean(minutes), 1),
        sd = round(sd(minutes), 1), spread = round(max(minutes) - min(minutes)),
        iqr = IQR(minutes), sdOut = round((Yglob - mean(minutes))/sd(minutes),1)))
## Error in eval(expr, envir, enclos): argument "by" is missing, with no default
stats #error message is associated with KnitR bug... evaluation works fine in R
## Error in eval(expr, envir, enclos): object 'stats' not found
```

Treatment Group	Mean	Standard Deviation
0	79.3	10.2
1	61.5	9.3
2	47.9	9.5
4	32.8	9.6

Each of the five samples only contain 5 data points, which makes assessing normality difficult to disprove. The global mean is 55.4, and the spread and IQR for each of samples are nearly identical as they vary by less than 5%.

1(b) **(10 pts) Looking at the graphical and descriptive summaries from part (a), do there appear to be differences in the typical REM levels for the 4 groups? Describe the differences you see that appear to be most pronounced.**

Yes, there appear to be differences in the typical REM levels fro the 4 groups. Differences exist in the value of each samples mean and the relative distance of the sample mean to the global mean. Adjacent groups appear have similar means; however, groups that are not adjacent do not appear to have similar means. Only groups 1 and 2 have IQRs that include the global mean, while groups 0 and 4 do not even have outliers that cross the global mean.

Below is a summary of confidence levels and sample means relative to the global mean:

- Treatment Group 0: The 95% confidence level of the sample mean does not include the global mean. The global mean is less than the sample mean by 2.3 standard deviations of the sample mean.

- Treatment Group 1: The 95% confidence level of the sample mean does include the global mean and is within 0.7 standard deviations of the sample mean.

- Treatment Group 2: The 95% confidence level of the sample mean does include the global mean and is within 0.8 standard deviations of the sample mean.

- Treatment Group 4: The 95% confidence level of the sample mean does not include the global mean. The global mean is greater than the sample mean by 2.4 standard deviations of the sample mean.

1(c) **(10 pts) Do a one-way Analysis of Variance to compare the mean REM levels for the 4 groups. Are the groups significantly different at the 5% level?**

```
# perform ANalaysi Of VAriance
fit.REM <- aov(minutes ~ dose, data = rem.long)
summary(fit.REM)
##              Df Sum Sq Mean Sq F value   Pr(>F)
## dose          3   5882    1961   21.09 8.32e-06 ***
## Residuals    16   1487      93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit.REM
## Call:
##    aov(formula = minutes ~ dose, data = rem.long)
##
## Terms:
##                    dose Residuals
## Sum of Squares  5882.358  1487.400
## Deg. of Freedom        3        16
##
## Residual standard error: 9.641706
## Estimated effects may be unbalanced
```

- The population parameter for each sample is the mean REM sleep duration, μ_n = average sleep duration for sample n.

- The F-statistic is used to test if it is plausible that the mean REM sleep durations of the four samples are all equal; that is $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. Where the alternative is that any two or more of the mean REM sleep durations are not equal; that is H_A : not H_0.

- The p-value of the F-test is 8.3e-6; therefore, I reject the Null hypothesis in favor of the Alternative hypothesis. The mean REM sleep durations of the four populations are not all equal at the 5% level or even the 1% level.

1(d) **(10 pts) Compare all possible pairs of groups using Fishers LSD method, and summarize the results of the multiple comparison. Repeat for Tukeys HSD method, and using Bonferroni comparisons. Do the three methods find different groupings? If so, what accounts for that?**

```r
n.sample <- 4 # number of population samples for which a mean was taken

# family error rate:  alpha < FER < c*alpha
alpha <- 0.05
c <- n.sample*(n.sample-1)/2 #pairs of means to compare
c*alpha
## [1] 0.3
# Fisher's LSD method
pairwise.t.test(rem.long$minutes, rem.long$dose, pool.sd = TRUE, p.adjust.method = "none")
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  rem.long$minutes and rem.long$dose
##
##    0        1       2
## 1 0.01024 -       -
## 2 9.8e-05 0.04014 -
## 4 1.0e-06 0.00023 0.02435
##
## P value adjustment method: none
```

Fisher's LSD Method: At the 5% level, sufficient eveidence exists to conclude that all populations means are unique. Results from this analysis are shown below in groups, notice that no groups are paired.

```
4 g/kg      2 g/kg      1 g/kg      0 g/kg
------      ------      ------      ------
```

Family Error Rate: $\alpha < FER < c\alpha \rightarrow 0.05 < FER < 0.30$

```r
# Tukey's HSD method
TukeyHSD(fit.REM)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = minutes ~ dose, data = rem.long)
##
## $dose
##       diff       lwr        upr       p adj
## 1-0 -17.74 -35.18636  -0.2936428 0.0455781
## 2-0 -31.36 -48.80636 -13.9136428 0.0005142
## 4-0 -46.52 -63.96636 -29.0736428 0.0000056
## 2-1 -13.62 -31.06636   3.8263572 0.1563545
## 4-1 -28.78 -46.22636 -11.3336428 0.0011925
## 4-2 -15.16 -32.60636   2.2863572 0.1005398
```

Tukey's HSD Method: At the 5% level, insufficient evidence exists to conclude that the population mean REM sleep duration for 1 g/Kg dosage, 2 g/Kg dosage, and 4 g/Kg dosage are unique. Sufficient evidence exists to conclude the mean REM sleep duration for 0 g/Kg dosage is different from the other populations means at the 5% level. Additionally, sufficient evidence exists to conclude the population mean REM sleep durations of 4 g/Kg and 1 g/Kg are different at the 5% level. Results from this analysis are shown below in groups.

```
4 g/kg      2 g/kg      1 g/kg      0 g/kg
            ------------------
------------------
```

Family Error Rate: varies

```r
# family error rate:  alpha/c < FER < alpha
alpha/c # lower bound
## [1] 0.008333333
alpha # upper bound
## [1] 0.05
```

```r
# Bonferroni comparison method
pairwise.t.test(rem.long$minutes, rem.long$dose, pool.sd = TRUE, p.adjust.method = "bonf")
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  rem.long$minutes and rem.long$dose
##
##   0       1       2
## 1 0.06146 -       -
## 2 0.00059 0.24085 -
## 4 6.1e-06 0.00139 0.14608
##
## P value adjustment method: bonferroni
```

Bonferroni Comparison Method: At the 5% level, insufficient evidence exists to conclue that the population mean REM sleep durations of 1 g/Kg dosage and 2 g/Kg dosage, 2 g/Kg dosage and 4 g/Kg dosage, and 1 g/Kg dosage and 0 g/Kg dosage are different. Sufficent evidence exists to conclude the population mean REM sleep durations of 4 g/Kg differ from 1 and 0 g/Kg and that 2 g/Kg differ from 0 g/Kg. Results from this analysis are shown below in groups.

```
4 g/kg      2 g/kg       1 g/kg      0 g/kg
                         -----------------
           ------------------
-----------------
```

Family Error Rate: $\alpha/c < FER < \alpha \rightarrow 0.0083 < FER < 0.05$

Do the methods find different groupings? Yes, Fisher's method did not result in any groupings, Tukey's method resulted in two groupings, and Bonferroni's methods resulted in three groupings. Of the methods utilized, Fisher's Method is the most likely to result in a Type I error (reject a true null hypothesis) as it is the most liberal of the three. Bonferroni's method is the most conservative of the three methods and is most likely to reult in a Type II error (fail to reject a false null hypothesis). Regarding error, Tukey's method falls in between the other two methods as its erorr is equal to a t-test but it accounts for increased probability of making a Type I error when making multiple comparisons. Tukey's method feels right.

1(e) (10 pts) Are the results of the F -test in part (c), and the multiple comparisons in part (d) consistent with what you described in part (b)? Briefly discuss.

Results from the F-test in part (c) are consistent with what I described in part (b). That is, at the 5% level (approximately two standard deviations) the groups do not all have the same mean. Results in part (d) are not consistent with what I described in part (b). I was unable to distinguish paired groups (those with similar means at the 5% level) based on a graphical summary. Adjacent groups appear to have similar differences between ranges, IQRs, means, and medians. However, Tukey's method identified two sets of paired groups where I was unable to distinguish this in part (b).

1(f) (10pts) Looking at the numerical and graphical summaries does it appear that the distributions of REM levels are reasonably normal, and have constant variance across groups? Discuss.

Based on the histogram and box plots in part (b), the distributions of the REM levels appear normal have have nearly constant variance across each group. The spread and standard deviation of each group varies by less than 5% across all the sample data.

1(g) (10 pts) Comment on the Levenes and Bartletts formal tests of equal variances in light of what you see when you look at the data.

```r
#Bartlett's test of equal variances
bartlett.test(minutes ~ dose, data = rem.long)
##
##  Bartlett test of homogeneity of variances
##
## data:  minutes by dose
## Bartlett's K-squared = 0.0323, df = 3, p-value = 0.9985
```

```
#Levene's test of equal variances
leveneTest(minutes ~ dose, data = rem.long)
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.0058 0.9994
##       16
```

The Bartlett and Levene tests both have the null hypothesis that sample distributions all have equal variance and alternate hypothesis that at least one sample distribution variance is not equal to the variance of another sample. Results from both the Bartlett and Levene tests fail to reject the Null hypothesis of the tests. P-values of both the Bartlett and Levene tests are well above 0.05, 0.9985 and 0.994, respectively. Therefore, at the 5% level and 1% level insufficient evidence exists to conclude the four samples are from poulations with an unequal variance. These results agree with the inference made based on the calculated standard deviations, i.e., the variances are equal.

The Bartlett's test is sensitive to departures from normality where Levene's test is less sensitive to departures from normality. The histogram of data does not appear normal, but the symmetry shown in the box plots does indicate the data are normal. The normality assumption should be verified for these tests.

1(h) **(10 pts) Obtain the residuals (or centered values) and generate a normal probability plot based on them. This is an overall check on normality. Is the normality assumption reasonable here? Perform a formal test of normality on the residuals, report the p-value and comment. If the test says we did not sample from normally distributed populations, why does it appear to be saying that? Do you think there is a real difficulty here with applying the normal theory methods?**

```
# create histograms
resid <- data.frame(fit.REM$residuals)
colnames(resid) <- "r"

# create histograms
resid.hist <- ggplot(resid, aes(x = r))
resid.hist <- resid.hist + scale_x_continuous(limits = c(-15,15))
resid.hist <- resid.hist + geom_histogram(binwidth = 4, color = "black", fill = "white",
                                 aes(y = ..density..))
resid.hist <- resid.hist + geom_density(alpha = 0.1, fill = "#FF6666")
resid.hist <- resid.hist + labs(y = "Count", x = "Residuals of REM Sleep Duration (minutes)")

# boxplot of Unseeded Precip
resid.box <- ggplot(resid, aes(x = "ANOVA Residuals", y = r)) # boxplot
resid.box <- resid.box + geom_boxplot()
resid.box <- resid.box + coord_flip()
resid.box <- resid.box + stat_summary(fun.y = mean, geom = "point", shape = 3, size = 4)
resid.box <- resid.box + geom_point(position = position_jitter(w = 0.05, h = 0),
                          alpha = 0.5, size = 2)
resid.box <- resid.box + labs(y = "ANOVA Residuals", x = "")

grid.arrange(resid.hist, resid.box, nrow = 2)
```
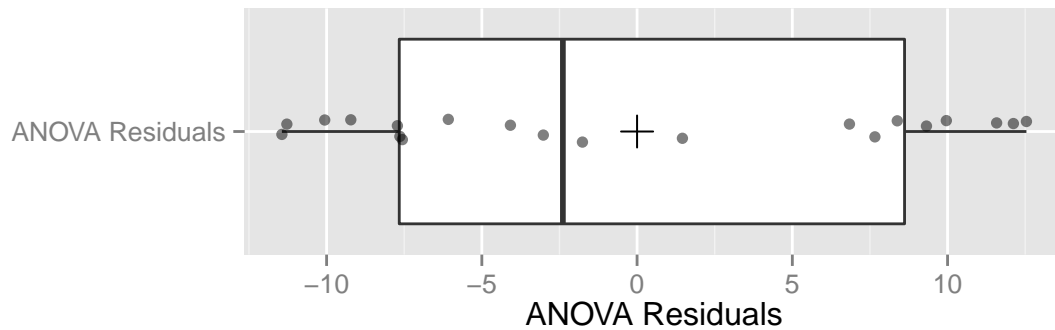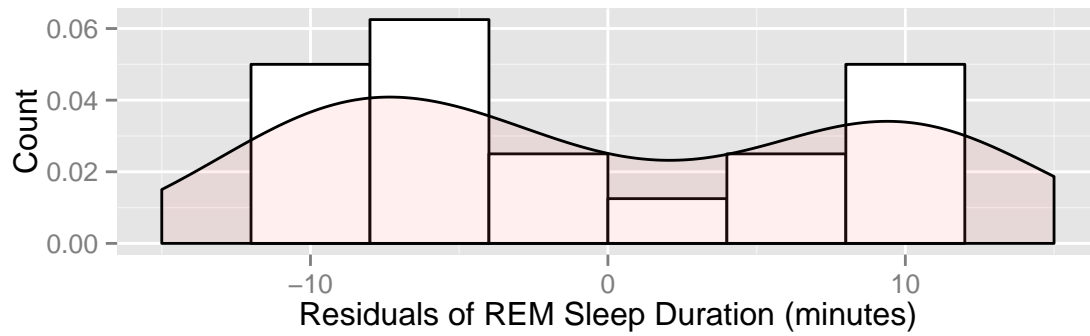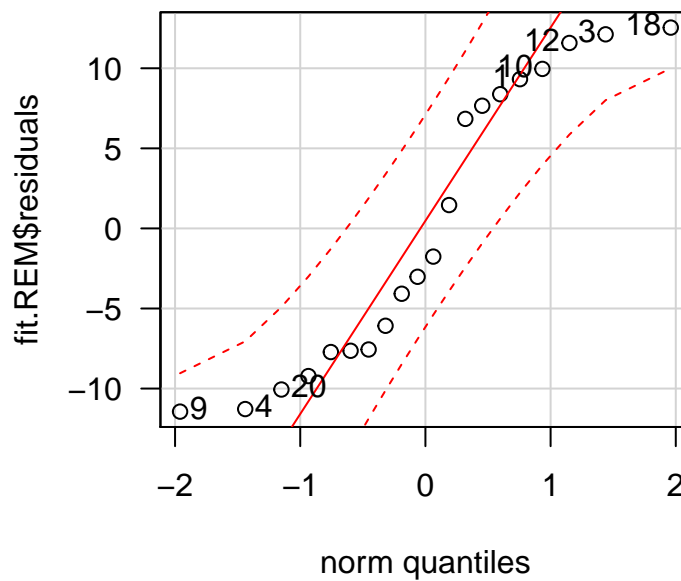
```
# QQ plot of residuals (normal probability plot of residuals)
qqPlot(fit.REM$residuals, las = 1, id.n = 8, id.cex = 1, lwd = 1, main = "QQ Plot of Residuals")
```

QQ Plot of Residuals


```
## 18  3 12  9  4 20 10  1
## 20 19 18  1  2  3 17 16
```

```
# shapiro-wilk test for normality
shapiro.test(fit.REM$residuals)
##
##  Shapiro-Wilk normality test
##
## data:  fit.REM$residuals
```

```
## W = 0.8781, p-value = 0.01633
# anderson - darling test for normality
library(nortest)
ad.test(fit.REM$residuals)
##
##  Anderson-Darling normality test
##
## data:  fit.REM$residuals
## A = 0.9091, p-value = 0.01651
```

Historgram and Boxplot of Residuals (data centered about the global mean): These plots show the residuals are potentially bimodal, although very few observations were made, which may result in a misleading distribution of the residuals. The residual distribution is symmetric about the mean. The mean is near the median and distribution has equal length tails, which are indications of a normal distribution of residuals.

Normal Probability Plot The QQ plot shows that all values fall within the normal distibution at the 5% level; therefore, there is no evidence from this plot that the residuals deviate from normality at this leve.

Formal Normality Tests: The Shapiro-Wilks and Anderson-Darling tests both have the null hypothesis that the samples are normally distributed and an alternate hypothesis of the samples not being normally distributed. Both of these tests reject the Null (residuals are not normally distributed) at the 5% level but fail to reject the Null (residuals are normally distributed) at 1% level. The tests appears to be rejecting the Null at the 5% level because of the small sample sizes, and I think the normal theory methods are suitable for this analysis.