

**Part I.** (80 points) Do all calculations in R. All R code for the assignment should be included with the part of the problem it addresses (for code and output use a fixed-width font, such as Courier). Code is used to calculate result. Text is used to report and interpret results. Do not report or interpret results in the code. Also:

1. Clearly define population parameters in each problem. That is, give a verbal description of what the population mean is in the context of the problem.
2. Clearly specify hypotheses when appropriate (not every problem involves a test of hypothesis).
3. Write a coherent conclusion based on each CI or test.

(80<sup>pts</sup>)

**1. REM with ethanol treatment:** An experiment was designed to investigate whether rapid eye movement (REM) sleep time depends on the amount of ethanol given in an injection. Four injection concentrations were considered: 0, 1, 2, and 4 grams per kilogram of body weight. Twenty rats were then chosen and randomly divided into four equal sized groups. Each group was given a different treatment (concentration of ethanol). The REM sleep time during the subsequent 24 hour period was recorded for each rat. We are interested in comparing the typical amount of REM sleep across treatments. The data are as follows:

Treatment	REM sleep in minutes				
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

Summarize all your results carefully in a report following the guided steps below (read all steps before beginning so you don't duplicate effort). Make certain you define all populations and parameters under consideration, and that you make clear with each step which parameter(s) you are analyzing. I am particularly concerned that you keep straight the distinction between sample and population means, and sample and population variances.

Here is a series of steps to get the table above into R in wide, then long, format. Study each step, looking at the intermediate `dat*` variables to see how I did it. Could you have done this on your own using the Chapter 5 notes?

```
# 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
library(reshape2)
rem.long <- melt(rem, variable.name = "dose", value.name = "minutes")
## Using as id variables
#str(rem.long)
```

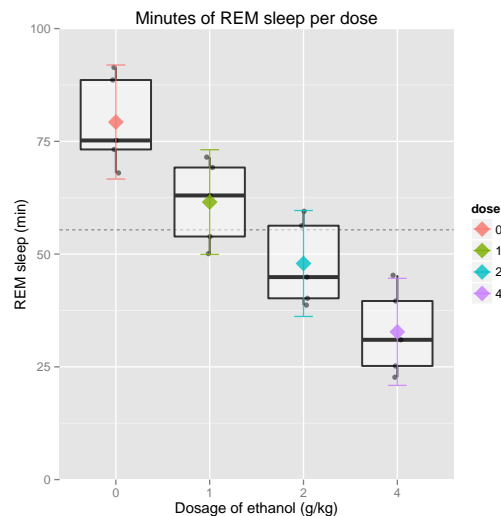
- (a) (10 pts) Make an appropriate graphical summary of the data to compare the groups. Compute, the sample means and standard deviations for the 4 treatment groups.

*Solution:* There appears to have decreasing sleep by dose, with spreads for each group roughly equal.

```
# Plot the data using ggplot
library(ggplot2)
p <- ggplot(rem.long, aes(x = dose, y = minutes))
# plot a reference line for the global mean (assuming no groups)
p <- p + geom_hline(aes(yintercept = mean(minutes)),
                    colour = "black", linetype = "dashed", size=.3, alpha=.5)
# boxplot, size=.75 to stand out behind CI
p <- p + geom_boxplot(size = 0.75, alpha = 0.5)
# points for observed data
p <- p + geom_point(position = position_jitter(w = 0.05, h = 0), alpha=0.5)
# diamond at mean for each group
p <- p + stat_summary(fun.y = mean, geom = "point", shape = 18, size = 6,
                    aes(colour=dose), alpha=.8)
# confidence limits based on normal distribution
p <- p + stat_summary(fun.data = "mean_cl_normal", geom = "errorbar",
                    width = .2, aes(colour=dose), alpha=0.8)
p <- p + coord_cartesian(ylim = c(0, 100))
p <- p + opts(title = "Minutes of REM sleep per dose") +
  ylab("REM sleep (min)") + xlab("Dosage of ethanol (g/kg)")

## 'opts' is deprecated. Use 'theme' instead. (Deprecated; last used in version 0.9.1)
## Setting the plot title with opts(title="...") is deprecated.
## Use labs(title="...") or ggtitle("...") instead. (Deprecated; last used in version 0.9.1)

print(p)
```



```
# Calculate the mean, sd, n, and se for the four rems
library(plyr)

##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:Hmisc':
##
##   is.discrete, summarize

rem.summary <- ddply(rem.long, "dose", function(X) { data.frame(
  m = mean(X$minutes), s = sd(X$minutes), n = length(X$minutes) ) } )
# standard errors
rem.summary$se <- rem.summary$s/sqrt(rem.summary$n)
# individual confidence limits
rem.summary$ci.l <- rem.summary$m - qt(1-.05/2, df=rem.summary$n-1) * rem.summary$se
rem.summary$ci.u <- rem.summary$m + qt(1-.05/2, df=rem.summary$n-1) * rem.summary$se
```

```
rem.summary
##   dose      m      s n      se ci.l ci.u
## 1    0 79.28 10.181 5 4.553 66.64 91.92
## 2    1 61.54  9.344 5 4.179 49.94 73.14
## 3    2 47.92  9.461 5 4.231 36.17 59.67
## 4    4 32.76  9.559 5 4.275 20.89 44.63
```

- (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.

*Solution:* Considering either the mean or the median as a measure of central tendency, the typical REM decreases as the administered dose of ethanol increases. The difference in means appears to be substantial relative to the spread in the distributions, which show little overlap.

- (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?

*Solution:* Let  $\mu_0$  = mean REM sleep in population of rats give a dose of 0 g/kg, and define  $\mu_1$ ,  $\mu_2$ , and  $\mu_4$  analogously. We are interested in testing  $H_0 : \mu_0 = \mu_1 = \mu_2 = \mu_4$  against  $H_a : \text{not } H_0$ . An Analysis of Variance table for performing the test is given below.

The p-value from the ANOVA table is  $p < 0.0001$ , which is less than the desired size of the test (0.05). Therefore, we reject  $H_0$  in favor of  $H_A$ , and conclude at the 5% level that the population mean REM depends on dose.

```
fit <- aov(minutes ~ dose, data = rem.long)
summary(fit)
##           Df Sum Sq Mean Sq F value Pr(>F)
## dose          3   5882    1961   21.1 8.3e-06 ***
## Residuals    16   1487     93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fit
## Call:
## aov(formula = minutes ~ dose, data = rem.long)
##
## Terms:
##           dose Residuals
## Sum of Squares 5882    1487
## Deg. of Freedom   3      16
##
## Residual standard error: 9.642
## Estimated effects may be unbalanced
```

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

*Solution:* Summaries for the three multiple comparison methods are given below. For the Bonferroni method, there are 6 possible pairwise comparisons, so a FER of 0.05 is guaranteed by setting the individual error level to  $0.05/6 = 0.0083$ .

```
# Fisher's LSD (FSD) uses "none"
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
#
# Tukey 95% Individual p-values
TukeyHSD(fit)
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = minutes ~ dose, data = rem.long)
##
## $dose
##      diff      lwr      upr p adj
## 1-0 -17.74 -35.19 -0.2936 0.0456
## 2-0 -31.36 -48.81 -13.9136 0.0005
## 4-0 -46.52 -63.97 -29.0736 0.0000
## 2-1 -13.62 -31.07  3.8264 0.1564
## 4-1 -28.78 -46.23 -11.3336 0.0012
## 4-2 -15.16 -32.61  2.2864 0.1005
#
# Bonferroni 95% Individual p-values
pairwise.t.test(rem.long$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

```

The three methods lead to the following groups:

```

              (ordered by decreasing mean)
dose:          0    1    2    4

Fisher        --- --- --- ---

Tukey         --- -----
              -----

Bonferroni    ----- -----
              -----

```

The 3 methods find different groupings. As expected, the Bonferroni groupings are coarsest (i.e., have more overlap) while the Fisher groupings are finer than the other two methods. The Fisher and Bonferroni comparisons are based on two-sample t-tests. Since the Bonferroni method uses a smaller individual error rate than Fisher's method, we expect the Bonferroni method to produce groupings with more overlap than Fisher's method. Tukey's method is expected to be intermediate to the other two methods, and this is what is seen here.

Noting that the mean REM decreases with increasing dose, Fisher's method leads to the conclusion that the decrease in REM is statistically significant for each increase in dose (because each treatment is in a separate group).

The Bonferroni comparisons indicate that the decrease in mean REM is statistically significant whenever the dose is increased by at least 2 levels (i.e., going from 0 to 2 or 4; and 1 to 4), but

that the mean REM at consecutive doses (i.e., 0 to 1, or 1 to 2, etc) are not significantly different. Tukey's method also leads to the Bonferroni conclusions, except that the differences between 0 and 1 g/kg are statistically significant.

- (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.

*Solution:* The results of the  $F$ -test and multiple comparisons are consistent with what we described earlier, and consistent with the general trend for the mean REM to decrease with increasing dose.

- (f) (10 pts) 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.

*Solution:* The sample sizes are very small so there is little information contained in each distribution to assess normality. Nonetheless, the boxplots show that the individual distributions are reasonably symmetric with no outliers, so there is no strong evidence of non-normality. In addition, the spreads in the distributions, as measured by either the IQR (measured indirectly from the widths of the boxes in the boxplots) or the standard deviation, are similar. The assumptions for the ANOVA appear to be met.

- (g) (10 pts) Comment on the Levene's and Bartlett's formal tests of equal variances in light of what you see when you look at the data.

*Solution:* The p-values for Bartlett's test and Levene's test (see below) exceed 0.90, which would lead us to not reject the hypothesis that the population variances are identical. The test results are not surprising, given the close agreement of the sample standard deviations.

```
## Test equal variance
# assumes populations are normal
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
# does not assume normality, requires car package
library(car)
leveneTest(minutes ~ dose, data = rem.long)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3    0.01     1
##      16
```

- (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?

*Solution:* A normal probability plot of the residuals is given below. The p-value for the Anderson-Darling test of normality is 0.017, which leads to a conclusion that the assumed common centered distribution for REM is not normal. The histogram suggests that the distribution of the residuals is bimodal. Although the normality assumption appears to be violated here (S-shape in normality plot, formal tests small p-values), typically only extreme skewness and outliers have a deleterious effect on the performance of statistical methods that assume normality. Since neither feature is present, I do not have serious concerns with applying the normal theory methods.

```
# plot of data
par(mfrow=c(3,1))
# Histogram overlaid with kernel density curve
```

```

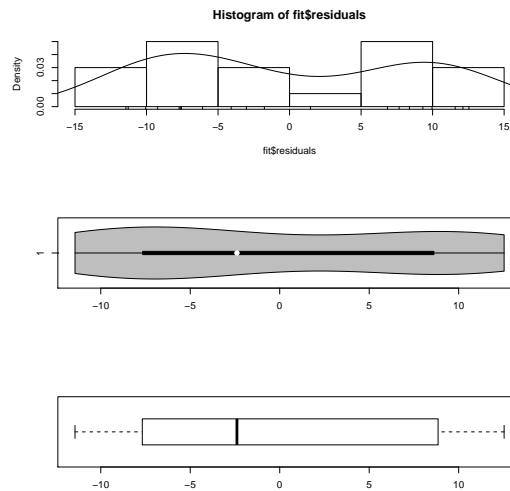
hist(fit$residuals, freq = FALSE, breaks = 5)
points(density(fit$residuals), type = "l")
rug(fit$residuals)

# violin plot
library(vioplplot)

## Loading required package: sm
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
vioplplot(fit$residuals, horizontal=TRUE, col="gray")

# boxplot
boxplot(fit$residuals, horizontal=TRUE)

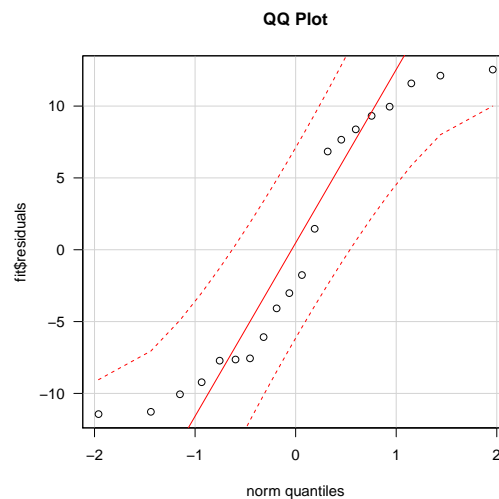
```



```

# QQ plot
par(mfrow=c(1,1))
library(car)
qqPlot(fit$residuals, las = 1, id.n = 0, id.cex = 1, lwd = 1, main="QQ Plot")

```



```

shapiro.test(fit$residuals)

##
## Shapiro-Wilk normality test
##

```

```
## data: fit$residuals
## W = 0.8781, p-value = 0.01633
library(nortest)
ad.test(fit$residuals)
##
## Anderson-Darling normality test
##
## data: fit$residuals
## A = 0.9091, p-value = 0.01651
cvm.test(fit$residuals)
##
## Cramer-von Mises normality test
##
## data: fit$residuals
## W = 0.1491, p-value = 0.02187
```