

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^{pts})

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.
- (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.
- (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?
- (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?
- (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.
- (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.
- (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.
- (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?