PubH 7401: Elements of Biostatistical Inference I
Homework 8, due Tuesday Dec. 2

1. In R, load the `nortest` package. Type `ad.test(rnorm(10))` Enter. This produces a p-value for testing the normality of *perfectly normal data* from 10 *iid* $N(0, 1)$ draws. Now hit ↑ Enter several times (e.g. say 20) in a row. Repeat for sample size $n = 100$. Do you ever see small p-values in either case (e.g. less than 0.1)? Would you ever reject normality from truly normal data?

2. Chapter 11 problem 21 a. Check the normality assumption graphically and formally.

3. Chapter 11 problem 22. Look at the *differences* in calcium measurements $Y_i = m_{i1} - m_{i2}$ where for $i = 1, \ldots, 118$ where m_{i1} is the standard method of measurement and m_{i2} is the flame photometry method. Then assume $Y_1, \ldots, Y_{118} \overset{iid}{\sim} N(\mu, \sigma^2)$ and test no difference, i.e. $H_0 : \mu = 0$ using the t-test procedure. Is the normality assumption okay here?

   This is known as paired data. Paired data occurs when two different treatments are given to the same experimental subject, or a subject is measured at two points in time. The resulting measurements are then differenced and a one-sample test is performed to test whether there's an overall mean difference in treatments within subjects. The subjects are often referred to as "being their own controls" in such a setup.

4. Chapter 11 problem 50 a, b.

5. Chapter 13 problem 4 (p. 531). Perform a hypothesis test and find a 95% CI for the difference in proportions. Use R's `prop.test()` function.

6. Chapter 13 problem 19 (p. 536). Just use the default test in R's `prop.test()` function.

7. Chapter 12 problem 28 (p. 510). This is a oneway ANOVA setup. Examine side-by-side boxplots of the data and comment on the standard ANOVA assumptions. Perform both Bonferroni and Tukey pairwise comparisons. Examine and comment on a histogram of the residuals.

8. Chapter 12 problem 33 (p. 511). This is a completely randomized design, which we'll discuss later.

   (a) Perform a standard ANOVA analysis followed by Tukey's procedure. Examine side-by-side boxplots of the data. Is normality a reasonable assumption? You might consider performing formal normality tests within one or more populations. Is constant variance reasonable? Does the overall F-test of $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ reject? Do a Tukey analysis on all $c = \binom{6}{2} = 15$ possible pairwise comparisons with an FER= 0.05. What significant differences do you find? Discuss. Look at a histogram of the residuals. Again, does normality seem reasonable? What are the results of a formal normality test on the residuals?

   (b) There are actually only $c = 3$ comparisons of interest conveniently given to you on page A44. Perform these three pairwise comparisons with an FER of 5% by following up your standard, normal-errors ANOVA fit with three pairwise comparisons, e.g.:

```
> d <- read.table("c:/tim/PubH7400/diet-and-longevity.txt",header=T)
> lifetime <- d[,1]
> treatment <- factor(d[,2])
> boxplot(lifetime~treatment)
> fit <- aov(lifetime~treatment)
> summary(fit)
> TukeyHSD(fit)
> pairwise.t.test(lifetime,treatment,p.adjust="none")
```

Multiply the three p-values *of interest* in the table produced by `pairwise.t.test()` by $c = 3$ and compare them to FER$= 0.05$. What do you conclude?

(c) Reanalyze the data by following a nonparametric Kruskal-Wallace test with pairwise comparisons of medians. This would go something like:

```
> np <- subset(lifetime,treatment=="NP")
> n.n85 <- subset(lifetime,treatment=="N/N85")
> lopro <- subset(lifetime,treatment=="lopro")
> n.r50 <- subset(lifetime,treatment=="N/R50")
> r.r50 <- subset(lifetime,treatment=="R/R50")
> n.r40 <- subset(lifetime,treatment=="N/R40")
> wilcox.test(n.r50,r.r50)
> wilcox.test(n.r50,lopro)
> wilcox.test(n.r50,n.r40)
```

Do you reject $H_0 : \eta_1 = \eta_2 = \eta_3 = \eta_4 = \eta_5 = \eta_6$? You should multiply the p-values from the three `wilcox.test()` evaluations by $c = 3$ and compare to the FER$= 0.05$. Do you reach the same conclusions as in part (b)?