Stat 371-003, Solutions to Homework #10

1. **11.38 (pg 516)**

   We first calculate the difference between the two averages, say $\bar{x}_E - \bar{x}_A = 5.38 - 4.37 = 1.01$.

   We next need to get an estimate of the standard error (SE) of $\bar{x}_E - \bar{x}_A$. We use $\sqrt{\text{MS(within)}} = \sqrt{0.2246} = 0.4739$ (from the ANOVA) as our pooled estimated of the within-group SD. And our estimated SE is

   $$\hat{\text{SE}}(\bar{x}_E - \bar{x}_A) = 0.4739\sqrt{\frac{1}{9} + \frac{1}{9}} = 0.2234$$

   We then find the appropriate quantile of the appropriate t distribution. We look at the t distribution with 40 degrees of freedoom (the within-group df from the ANOVA), as 40 = 45 − 5. (45 is the total sample size and 5 is the number of groups). And we consider the 1 - 0.025/10 = 0.9975 percentile, since we are doing a Bonferroni correction for the 10 possible pairwise comparisons.

   The 99.75 percentile of a t distribution with 40 degrees of freedom (from the table at the back of the book, or using `qt(0.9975, 40)` in R) is 2.971.

   Thus, our Bonferroni-corrected confidence interval is

   $$1.01 \pm 2.971 \times 0.2234 = 1.01 \pm 0.6638 = \textbf{(0.35, 1.67)}$$

2. **Kruskal-Wallis test**
   We read in the data as follows:

   ```
   dat <- read.csv("http://www.biostat.wisc.edu/%7Ekbroman/teaching/stat371/genenstein.csv")
   ```

   To perform the Kruskal-Wallis test, we type

   ```
   kruskal.test(response ~ genenstein, data=dat)
   ```

   Note that the response is on the left of the tilde and the treatment factor is on the right.

   The results are the following:

   ```
   Kruskal-Wallis rank sum test

   data:  response by genenstein
   Kruskal-Wallis chi-squared = 1.5166, df = 2, p-value = 0.4685
   ```

   Note that df = 2, since there are three treatment groups.

   The p-value for the test for a difference among the three treatment groups is 47%, and so we reject the null hypothesis and conclude that the data are consistent with the null hypothesis of no treatment effect. In other words, we have insufficient evidence to conclude that there is a treatment effect.

3. **11.17 (pg 497)**

   We read in the data as follows.

   ```
   dat <- read.csv("http://www.biostat.wisc.edu/%7Ekbroman/teaching/stat371/data_11-17.csv")
   ```
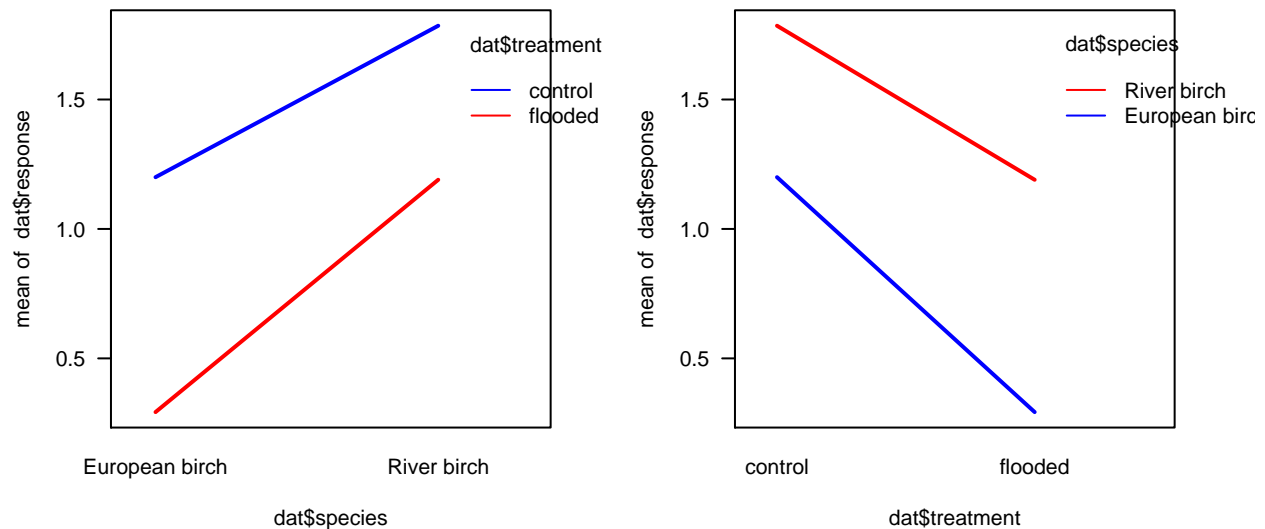
   To create the interaction plot, we type

   ```
   interaction.plot(dat$species, dat$treatment, dat$response)
   ```

   We could also have typed

   ```
   interaction.plot(dat$treatment, dat$species, dat$response)
   ```

   These give the following two versions of the interaction plot:

4. **11.18 (pgs 497–498)**

   (a) We can use R to construct the ANOVA table, or we can just do it by hand.
       In R, we would type,

       ```
       summary( aov(response ~ species * treatment, data=dat) )
       ```

       We get the following table.

Source	df	SS	MS	F	P
species	1	2.198	2.198	55.60	8×10^{-6}
treatment	1	2.258	2.258	57.11	7×10^{-6}
species $\times$ treatment	1	0.098	0.098	2.47	0.142
error	12	0.474	0.040		

       You could leave out the **F** and **P** columns, here.

(b) The F statistic for the species $\times$ treatment interaction is 0.098/0.040 = 2.47. With df = 1 and 12, Table 10 indicates that the P-value is between 0.1 and 0.2, or the calculations in R show that P $\approx$ 0.142.

We thus fail to reject the null hypothesis and conclude that there is no evidence for an interaction. In other words, the treatment effect in the two species could reasonably be the same.

(c) The F statistic for the effect of species is 2.198/0.040 = 55.60. With df = 1 and 12, Table 10 indicates that the P-value is < 0.0001, or the calculations in R show that P $\approx 8 \times 10^{-6}$.

Thus, we firmly reject the null hypothesis and conclude that species differ in ATP concentration.

(d) We estimate the within-group SD by $\sqrt{\text{MS(error)}} = \sqrt{0.040} \approx 0.199$.

5. **11.52 (pg 521)**

We first calculate the difference between the two averages, $\bar{x}_{\text{probucol}} - \bar{x}_{\text{placebo}} = 1.79 - 1.43 = 0.36$

We next need to get an estimate of the standard error (SE) of $\bar{x}_{\text{probucol}} - \bar{x}_{\text{placebo}}$. We will use $\sqrt{\text{MS(within)}}$, but we need to first calculate that.

SS(within) = SS(total) − SS(between) = 73.9945 − 5.4336 = 68.5609.

df(within) = (62+58+54+56) − 4 = 226.

Thus MS(within) = 68.5609 / 226 = 0.3034, and so out estimate of the within-group SD is $\sqrt{0.3034} = 0.5508$.

Our estimated SE is then

$$\hat{\text{SE}}(\bar{x}_{\text{probucol}} - \bar{x}_{\text{placebo}}) = 0.5508\sqrt{\frac{1}{62} + \frac{1}{58}} = 0.1006$$

We then find the appropriate quantile of the appropriate t distribution. We look at the t distribution with 226 degrees of freedoom (the within-group df from the ANOVA). And we consider the 97.5 percentile. (If we were doing a Bonferroni correction for the 6 possible pairwise comparisons, we would look at the 1 - 0.05/6 = 0.9917 percentile, but we weren't asked to do a Bonferroni correction; presumably this particular comparison was deemed important *in advance* of gathering the data.)

The 97.5 percentile of a t distribution with 226 degrees of freedom (from the table at the back of the book, or using `qt(0.975, 226)` in R) is 1.971.

Thus, our 95% confidence interval is

$$0.36 \pm 1.971 \times 0.1006 = 0.36 \pm 0.198 = \textbf{(0.16, 0.56)}$$