

Stat 401, Section F Homework 7

Due Date: Wednesday, October 17

1. A plant physiologist investigated the effect of flooding on root metabolism in two tree species: flood-tolerant river birch and the intolerant European birch. Four seedlings of each species were flooded for one day and four were used as controls. The concentration of adenosine triphosphate (ATP) in the roots of each plant was measured. The data (nmol ATP per mg tissue) are shown in the table:

River	Birch	European	Birch
Flooded	Control	Flooded	Control
1.45	1.70	.21	1.34
1.19	2.04	.58	.99
1.05	1.49	.11	1.17
1.07	1.91	.27	1.30

- (a) One way to check the appropriateness of the one-way ANOVA assumptions of normality and constant variance for all groups is to examine the residuals. Compute the (group specific) residuals for this data.
- (b) If the one-way ANOVA model assumptions of normality and constant standard deviation are approximately satisfied, (i) the collection of all residuals should have a distribution that is approximately normal, and (ii) the standard deviation of the residuals in each group should be about the same. Does this appear to be the case for this problem? Use **R** to draw a box plot for all the residuals and the residual plot in order to decide. (*a residual plot is simply a scatter-plot with the group means on the x-axis and the residuals on the y-axis*)
- (c) In the case of one-way ANOVA, the standard deviations of the observations in each group can be compared to assess whether the constant standard deviation assumption is approximately satisfied. Give the standard deviations for each of the groups in this problem and explain whether you are willing to accept the constant standard deviation assumption.
- (d) Could the ATP concentration for the four groups be the same? Provide a test statistic, p -value, and a conclusion.
- (e) Determine a 95% confidence interval for the mean ATP concentration of each of the four groups.
- (f) Define linear combinations (that is, specify the multipliers) to measure each of the following:
 - i. The effect of flooding in river birch
 - ii. The effect of flooding in European birch
 - iii. The difference between river birch and European birch with respect to the effect of flooding
- (g) Use a t -test to investigate whether flooding has the same effect in river birch and in European birch.
- (h) If the sample sizes were $n = 10$ rather than $n = 4$ for each group, but the means, standard deviations and s_{pooled} remained the same, how would the result above (part(g)) change?
- (i) Provide a 95% confidence interval for the contrast that compares the average of the Flooded and Control River birch mean ATP levels to the average of the Flooded and Control European birch mean ATP levels.
- (j) Test the claim that flooding has no effect on ATP level in river birch using the extra residual sums of squares principle (we also called it a comparison between the *Full and Reduced models* in class). Show all your work. You are encouraged but not required to examine the **R** code I wrote for the *doughnut example* we discussed in the lab to verify your calculations for the residuals sums of squares (`doughnut_RSS.r`).

- (k) Consider the null hypothesis that flooding has no effect on ATP level in river birch. This hypothesis could be tested in at least these two ways: employing a contrast, or with a two-sample t -test. Answer the following questions (do not actually carry out the tests):
1. In what way or ways do the two procedures differ?
 2. In what way or ways do the conditions for validity of the two procedures differ?
 3. One of the two procedures requires more conditions for its validity, but if the conditions are met, then this procedure has certain advantages over the other one. What are these advantages?
2. One measure of lung function is forced vital capacity (FVC), which is the maximal amount of air a person can expire in one breath. In a public health survey, researchers measured FVC in a large sample of people. The results for male ex-smokers, stratified by age, are shown in the following table:

Age(years)	n	FVC (liters)	
		Mean	SD
25–34	83	5.29	.76
35–44	102	5.05	.77
45–54	126	4.51	.74
55–64	97	4.24	.80
65–74	73	3.58	.82
25–74	481	4.56	

- (a) Suppose it is desired to calculate a summary value for FVC in male ex-smokers. One possibility would be simply to calculate the grand mean of the 481 observed values which is 4.56 liters. Can you think of any serious drawbacks if we adopt this approach?
 - (b) Construct a 95% confidence interval for the grand mean.
 - (c) Another summary measure is the “age-adjusted” mean, which is an estimate of the mean FVC value in a reference population with a specified age distribution. This is given by the following linear combination:

$$L = .27\bar{Y}_1 + .28\bar{Y}_2 + .21\bar{Y}_3 + .13\bar{Y}_4 + .11\bar{Y}_5 .$$
 Construct a 95% confidence interval for the age-adjusted mean.
 - (d) Is this a contrast? Why or why not?
 - (e) Based on the two confidence intervals that you obtained above, what could you conclude about the two measures of FVC (i.e. grand mean and age-adjusted mean)?
 - (f) Could the mean of the age group 35–44 and 55–64 be the same? Provide a test statistic, a p -value, and a conclusion.
 - (g) Are you entitled to use the techniques you employed above? Why or why not?
3. Use the extra residual sums of squares principle to solve Problem 23, Chapter 5 (page 146). You can do all the necessary calculations by hand (I don’t suggest you do so, there is too much work), or you can use the **R** code I wrote for the doughnut example (`doughnut_RSS.r`) and adapt to this problem. *Hint: You will have to perform more than one model comparison, read the problem with great care.*