

Handed out: Wednesday October 21, 2009

Due: Wednesday October 28, 2009

Reading:	Oct. 21	Finish randomized complete block design (Chap 21.1-21.5)
	Oct. 23 - Oct. 28	Correlation/simple regression (Chap 1.1-1.3, 1.6-1.8, 2.1-2.9, 2.11, 3.1-3.7, skim 4.1-4.7)
	Oct. 27	NOTE: Lecture during discussion
	Oct. 30	NOTE: No lecture (out of town)
	Nov. 2	Begin multiple regression (Chap 6-7, review Chap 5)
	Nov. 3	Extended discussion (with exam review)
	Nov. 4	MIDTERM (thru simple regression/correlation)
	Nov. 6	NOTE: No lecture (out of town)
	Nov. 9 - Nov. 13	Multiple regression (Chap 6-7)
	Nov. 10	NOTE: Lecture during discussion
	Nov. 11	HOLIDAY

NOTE: I am traveling Friday October 30 and November 6. See above for detailed schedule over this time period.

NOTE: Midterm exam is in class on Wed Nov 4. I will post last year’s midterm with solutions. We will be able to spend extended time on Nov 3 addressing any questions that you may have (during discussion and then I hope to arrange for additional hour immediately after so that we can continue).

1. **Pygmalion effect** – The Pygmalion effect refers to a psychological phenomenon whereby teacher or instructor expectations have an impact on student performance. More simply put if your teacher/instructor thinks that you will do well then you actually do perform well. It has been documented in schools. A study was carried out in the army to see if the Pygmalion effect occurs in army training as well. Nine companies of soldiers (a company is a big unit that works together) were chosen for the study. In each company there are three platoons. The platoons are randomly assigned to treatment – one platoon gets the “Pygmalion” treatment (it’s leader is told that this platoon tested especially high even though it is just an average group) and the two other platoons serve as controls. The data are provided below. The average score on a post-training exam is provided below for each platoon. This looks like paired data but not exactly; we walk thru an analysis in this question.

Company	Pygmalion	Controls
1	80.0	63.2, 69.2
2	83.9	63.1, 81.5
3	68.2	76.2, 84.1
4	76.5	59.5, 73.5
5	87.8	73.9, 78.5
6	89.8	78.9, 84.7
7	76.1	60.6, 69.6
8	71.5	67.8, 73.2
9	69.5	72.3, 73.9

- (a) The entire platoon’s score is treated as a single observation instead of using individual soldier scores. Explain why this is the appropriate thing to do.
 - (b) It is natural to assume there is a company effect because the different platoons in the company work together. This means the platoon scores in the Pygmalion and control groups are not independent. Calculate the difference between the Pygmalion platoon score and the average of the controls for each company. Test the hypothesis of no Pygmalion effect versus a one-sided alternative. State your result.
 - (c) Describe the assumptions required for the test procedure that you used in (b). Be as clear as possible.
 - (d) Note that we have two controls in each company. We can estimate the platoon-to-platoon variation within a company by calculating $\frac{1}{9} \sum_{i=1}^9 (C_{i1} - \bar{C}_i)^2 + (C_{i2} - \bar{C}_i)^2$ where C_{i1} is the first control score in company i , C_{i2} is the second control score in company i and \bar{C}_i is the average of the two control scores in company i . Find this estimated variance.
 - (e) How does this estimated variance compare to the variance of the differences you used in your hypothesis test. The two are measuring different quantities but are related. Explain.
2. **SAS for randomized block ANOVA: weed control** – An experiment was carried out to compare three treatments for weed control on a soybean farm. The treatments are no herbicide (control group), herbicide 2 weeks after planting, and herbicide 4 weeks after planting. The response is the total weed biomass (in kilograms/hectare). Note that low

values of the response (less weeds) indicate a more effective treatment. The experiment was repeated with sixteen different soil varieties. The data are available in the file weedctrl (weedctrl.txt, weedctrl.csv, weedctrl.xls) on the website. Note that there are four variables variety, treat (0, 2 or 4 for control, 2 weeks, 4 weeks), loc (1 or 2), and biomass (the response). Complete the analysis using only data from location 1. (SAS Hints: Use 'IF LOC = 1;' in the data step to select only the cases from location 1. Note that the variety variable is a character variable. You indicate this to SAS by including a \$ after the variable name on the INPUT command. See also below regarding SAS for this problem.)

- Plot the response vs treatment group with an indication of block/plot (as was demonstrated in class). What does this plot suggest about the assumption of no interaction between treatment and block?
- Run the randomized block ANOVA in SAS (recall a sample randomized block program is on the web site). Plot the residuals vs treatment. Plot the residuals vs block. Prepare a normal probability plot for the residuals. Discuss the validity of the randomized block ANOVA assumptions.
- Does it appear that all treatments are equally effective? Give the p -value of the relevant test and discuss the results.
- Which treatments appear to be the most effective? Does this make sense to you.
- Does blocking appear to have been effective here? Explain.

SAS: Start with the sample randomized block program on the web site. The randomized block program on the website shows you how to make the plot to examine the block effects. When you make the residual plots for (b) you need to undo the special graphics commands that told SAS to connect the dots. The easiest way is to add "symbol i=none;" to the residual plot procedure – this tells it not to connect the points (no interpolation).

3. Random/fixed effects in studies with blocking/pairing – The model for paired or block data can be written as $Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$ with Y_{ij} denoting the observation for treatment j in block/pair i with $i = 1, \dots, I$ and $j = 1, \dots, J$. The β_i are the block/pair effects, the τ_j are the treatment effects, and the ϵ_{ij} are $N(0, \sigma^2)$ random variables. The overall mean μ and the treatment effects τ_j are considered fixed but unknown constants (i.e., not random variables).

- The model is expressed in "factor effects" form where we typically assume that $\sum_j \tau_j = 0$. Explain why this assumption is necessary. Hint: you can just consider the case with $J = 2$ treatments.
- Consider inference for the mean response of a treatment.
 - Suppose the block effects are considered fixed effects with $\sum_i \beta_i = 0$. Show that $\bar{Y}_{.j} = \frac{1}{I} \sum_i Y_{ij}$ is $N(\mu + \tau_j, \sigma^2/I)$ and explain how this can be used for inference about the mean response to treatment j .
 - Suppose the block effects are considered random effects with $\beta_i \sim N(0, \sigma_b^2)$ and the β_i independent of the ϵ_{ij} 's. Show that $\bar{Y}_{.j}$ has the same mean as above but different variance. Provide an explanation for why there should be different inferences for the population mean response to treatment j in the two cases.
- Consider inference for the difference in the mean response of two treatments, $(\mu + \tau_j) - (\mu + \tau_{j'}) = \tau_j - \tau_{j'}$. Show that the distribution of $\bar{Y}_{.j} - \bar{Y}_{.j'}$ is the same regardless of how we think about block effects (fixed or random). Explain.
- Does it matter whether blocks are considered fixed or random in the case where we are interested in inference for a contrast $\sum_j c_j \tau_j$? Explain.

4. Expected mean squares – The expected values of the entries in the mean squares column in an ANOVA table are critical for developing the ANOVA test for a treatment effect. Computing expected values of mean squares is not a ton of fun – but it is worth doing once!

- Read through Sections 21.2 and 21.3 to familiarize yourself with the book's notation for a randomized complete block model (which differs slightly from mine). Derive the expected value of MSTR (mean squares for treatment) that is given in Table 21.2, i.e., show that $E(MSTR) = \sigma^2 + n_b \frac{\sum \tau_j^2}{r-1}$.
- Use the result to explain the logic behind the F-test for treatments.
- What is a treatment block interaction? Explain why the test is conservative if there are treatment block interactions.