Spring 2003 (Newton) 15.075 Applied Statistics

Final Exam

Wednesday, 21 May 2003
1:30 to 4:30

Open textbook and notes. Please use exam books and show all work. All problems will be weighted equally. Unless otherwise specified use alpha=0.05. Good luck!

The Headmaster of the School of Wizardry is interested in determining which of two magic wand types (Type A or Type B) is better. In this study, 5 students were randomly selected and assigned Type A wands, and 5 were randomly selected and assigned Type B wands. Points scored in a game of skill were recorded. The results are as follows:

Table of Results:

Type A	3	3	4	4	5
Type B	4	4	5	6	7

Part A:

(1) What kind of study is this: observational or experimental, comparative or descriptive? Make a histogram of each sample (Type A and Type B) and comment on the results. Give a measure of location and a measure of dispersion for each sample. Suggest ways to improve the study.

Part B:

For Parts B and C, assume that the samples are independent.

(2) Use a parametric test (Chapter 8) to determine if there a significant difference between the two wand types. Can you pool the sample variances? What are your assumptions, hypotheses and conclusions?

(3) Use a nonparametric test to determine if there is a significant difference between the two wand types. What are your assumptions, hypotheses and conclusions?

Part C:

(4) Use a linear model to determine if there is a significant difference between the two wand types. What is the model? What are the assumptions and hypotheses?

(5) Your analysis will be easiest if you choose an explanatory variable that is orthogonal to the intercept (column of 1's). Show the X matrix, the X'X matrix, the X'X inverse matrix, the X'y matrix, and the coefficient estimates.

(6) Show the covariance matrix of the coefficients and the correlation matrix of the coefficients. Calculate 95% confidence intervals for each of the coefficients. Interpret the confidence intervals.

(7) Show the fitted values and the residuals. What are the Residual Standard Error and Multiple R-Squared (coefficient of determination)?

(8) Make a coefficient table (such as provided by the S-Plus summary.lm function) with a column for coefficient values, a column for standard errors, a column for t-values and a column for P-values. For the P-values, just say whether they are less than 0.05 or not. Interpret the P-values. What do you conclude from this table?

(9) Make an ANOVA table (such as provided by the S-Plus summary.aov function) with a column for degrees of freedom, a column for sums of squares, a column for mean squares, a column for F values and a column for P-values (again, just less than 0.05 or not). What are your conclusions from this table?

Part D:

Actually, in the Table of Results, each column represents scores from the same game. That is, scores reported in column 1 are from game 1 of the season, column 2 from game 2 of the season, etc. Questions 9 and 10 ask you to repeat the analyses in Part B using this information.

(10) Use a parametric test (Chapter 8) to determine if there is a significant difference between the two wand types. What are your assumptions, hypotheses and conclusions?

(11) Use a nonparametric test to determine if there is a significant difference between the two wand types. What are your assumptions, hypotheses and conclusions?

Part E:

(12) How would the analysis for **Part C** change? What is the model? What are the assumptions and hypotheses?

(13) Complete the following ANOVA table. What are your conclusions from this table?

Source	Sum Sq	DF	Mean Sq	F value	P value
as.factor(wandtype)	4.9				
as.factor(game)	9.0				
Error	0.6				

Part F: Extra Credit

(14) There is particular interest in a score greater than 5. Denote a score greater than 5 as a success and a score less than or equal to 5 as a failure. Use Fisher's Exact Test to decide if the number of successes is related to wand type.