Statistical Methods II
Assignment 12

Due: Thursday, April 14, 2011

Here we begin a new section of the course. In this section, there is much more emphasis placed on experimental design. Make sure you know the basics of each design, especially when it is most appropriate and when it is not appropriate. Make sure you can distinguich between and among the experimental designs. The analysis of the designs requires little additional knowledge than what we did in the first section of the course. Of course, there are subtleties that must be noticed and new variations on the formula that must be respected.

⋆ ⋆ ⋆

Before you start this assignment, you may want to go back through the scripts for the week. They may help you in completing this assignment, especially in how to get the data from this assignment into R. I am not providing an external data file for you this time.

Do not worry about non-parametric tests for this assignment. Just use the analysis of variance procedure without testing its assumptions (unless specifically requested). However, please think of the assumptions and how you would test them.

As always, if you have questions, ask.

Problem 1 [[3]]

Researchers in child development are interested in developing ways to increase the spatial-temporal reasoning of preschool children. Spatialtemporal reasoning relates to the child's ability to visualize spatial patterns and mentally manipulate them over a time-ordered sequence of spatial transformations. This ability, often referred to as thinking in pictures, is important for generating and conceptualizing solutions to multi-step problems and is crucial in early child development.

The researchers want to design a study to evaluate which of several methods proposed to accelerate the growth in spatialtemporal reasoning yields the greatest increase in a child's development in this area. There are three methods proposed: three months of playing piano lessons, three months of playing specially developed computer video games, and specially designed games in small groups supervised by a trained instructor. The researchers measure the effectiveness of the three programs by assessing the children and assigning them a reasoning score both before and after their participation in the program. The difference in these two scores is the response variable. A control group is also included to measure the change in reasoning for children not given any special instruction. A pilot study with only 20 students was to be conducted prior to the complete study to determine potential problems.

(1) What is the dependent variable? What is the independent variable? How many levels are in that factor?
(2) Number the students 1 to 20. Assign 5 of the 20 students to each of the 4 types of instruction so that the assignment is completely random.
(3) Evaluate the use of completely randomized design in this case.

Problem 2 [[4]]

A petroleum company was interested in comparing the miles per gallon achieved by four different gasoline blends (A, B, C, and D). Because there can be considerable variability due to differences in driving characteristics and car models, these two extraneous sources of variability were included as "blocking" variables in the study. The researcher selected four different brands of cars and four different drivers. The drivers and brands of cars were assigned to blends in the manner displayed in the table below. The mileage (in mpg) obtained over each test run was recorded as follows.

(1) What is the dependent variable? What is(are) the independent variable(s)? How many levels are in the(each) factor? What is(are) the blocking variable(s)?
(2) Assuming an additive model, estimate the effect of car model on miles per gallon.
(3) Assuming an additive model, estimate the effect of driver number on miles per gallon.
(4) Assuming an additive model, estimate the effect of gasoline blend on miles per gallon.
(5) Which gasoline blend is the best? Is it significantly better than the other blends?

	Car Model			
Driver	1	2	3	4
1	(A) 15.5	(B) 33.8	(C) 13.7	(D) 29.2
2	(B) 16.3	(C) 26.4	(D) 19.1	(A) 22.5
3	(C) 10.5	(D) 31.5	(A) 17.5	(B) 30.1
4	(D) 14.0	(A) 34.5	(B) 19.7	(C) 21.6

Table 1. *Data to accompany Problem 12.2. The letter in parentheses represents the gasoline blend.*

Problem 3 [[3]]

The yields of wheat (in pounds) are shown here for five farms. Five plots are selected based on their soil fertility at each farm with the most fertile plots designated as 1. The treatments (fertilizers) applied to each plot are shown in parentheses.

(1) Identify the design used. Why is this design superior to the completely randomized design?

(2) Do an analysis of variance and draw conclusions concerning the five fertilizers.

(3) Which fertilizer was best? Was it significantly better than the others?

	Fertility				
Farm	1	2	3	4	5
1	(D) 10.3	(E) 8.6	(A) 6.7	(C) 7.6	(B) 5.8
2	(E) 8.8	(B) 6.7	(C) 6.7	(A) 4.8	(D) 6.0
3	(A) 6.3	(C) 8.3	(B) 6.8	(D) 8.0	(E) 8.8
4	(C) 8.9	(D) 7.4	(E) 8.2	(B) 6.2	(A) 4.4
5	(B) 7.3	(A) 4.4	(D) 7.7	(E) 6.8	(C) 6.7

Table 2. *Data to accompany Problem 12.3. The letter in parentheses represents the fertilizer type.*