Fall, 2007

APPLIED STATISTICS COMPREHENSIVE EXAM

METHODS I & II

Answers to all questions require complete explanations to receive full credit. Statistical tables will be provided. Calculators are permitted for this exam.

1. (20 points) The values listed below are counts of the number of times each digit 0 to 9 appears among the first 1000 digits of pi. Using an appropriate level 0.05 test, decide if these data are consistent with the theory that each digit is equally likely. Make sure to verify the conditions needed for this test to be valid.

Digit	Number of Occurrences
0	93
1	116
2	103
3	102
4	93
5	97
6	94
7	95
8	101
9	106

2. (25 points) An experiment is run to compare the efficacy of Drug A to that of Drug B. Each patient in the experiment is first administered one of the drugs, and a measure of drug efficacy is recorded. After a washout period, each patient is administered the other of the two drugs, and the same measure of drug efficacy is again recorded. The order in which the drugs are administered is randomized for each patient. The data are given below. Assuming that higher values indicate higher efficacy, use an appropriate level 0.05 test to decide whether there is evidence that Drug B is more effective than Drug A. State any necessary assumptions and verify them, when possible.

Patient	A	B
1	120	150
2	140	180
3	130	150
4	120	110
5	160	190
6	180	220

3. (25 points) A farmer plants corn, using one of five different types of fertilizer and one of three different types of seed. She plants two plots of corn for each fertilizer-seed combination, for a total of $5 \times 3 \times 2 = 30$ plots of corn. The yield of each plot was then measured. The sums of squares are as follows:

$$SS_{Fertilizer} = 449.5,\ SS_{Seed} = 512.9,\ SS_{Interaction} = 143,\ SS_{Within} = 136$$

   a. Create the ANOVA table, including columns for source, SS, df, MS, and F-statistic.
   b. Using the ANOVA table, make appropriate statistical inferences.
   c. What would be your next course of action for this analysis?

4. (30 points) Each year, Villanova's Department of Public Safety collects data on crimes committed at Villanova, in accordance with state and federal regulations. Included in these data is the number of "liquor referrals" (when they were called for or observed potential violations in University alcohol policies). The following data represent the number of liquor referrals (y) at Villanova over the past five years ($x = year - 2000$).

x	y
2	629
3	793
4	830
5	792
6	906

   A linear regression model produced an equation in the following form:

$$\hat{y}_i = b_0 + b_1 x_i$$

   and the following statistics were calculated:

$$b_1 = 55.3,\ MSE = 3470,\ \bar{x} = 4,\ \bar{y} = 790$$

   a. Calculate the value of b_0.
   b. Interpret the values of b_0 and b_1 in the context of the problem.
   c. Determine if there is a significant linear relationship between these variables at the 5% level of significance.
   d. The Department of Public Safety predicts that there will be 1030 liquor referrals in 2007. Based on the data from the previous years, do you think that its prediction is appropriate?