**STAT 302H MIDTERM #2 SPRING 2007**

1. A pharmaceutical company wants to study two, new non-steroidal formulas for a new pain reliever. For each formula, they gave the drug to 100 patients with headaches. Sixty-two of the patients receiving Formula 1 indicated improvement and forty-eight of those receiving Formula 2 indicated improvement.
    a. Formally state and test the appropriate hypotheses that the two formulas are equivalent.
    b. Construct a 90% confidence interval for the difference between the two true proportions.
    c. Now, assume that the company actually tested a 3rd formula on 100 patients, and seventy-two declare improvement. Knowing what you know so far, how might you compare the three formulas? Explain you answer.

2. Environmental chemists are interested in looking at four different adjustments to a basic bioremediation formula by adding an additional supplement. Each of the four different treatments was tested on three different samples needing remediation. The four different groups and the results of the tests are listed below. The higher the score, the better the group performed.

| **10% Additive** | **20% Additive** | **30% Additive** | **40% Additive** |
|---|---|---|---|
| 45 | 58 | 42 | 25 |
| 39 | 64 | 46 | 28 |
| 41 | 66 | 45 | 29 |
| $\bar{X}$ =41.667 | 62.667 | 44.333 | 27.333 |
| $S^2$ = 9.333 | 17.333 | 4.333 | 4.333 |

    a. Formally state and test the hypotheses to determine if the four groups performed differently.
    b. How would you analyze the data if you decided to use regression?

3. Consider the data set below where we wish to study the relationship between X=Age (in Days) and Y=Somite Level (a measure of development) in six mice.

| | | | | | | |
|---|---|---|---|---|---|---|
| **X:** | 7 | 9 | 8 | 24 | 25 | 22 |
| **Y:** | 9 | 6 | 7 | 15 | 14 | 18 |

    a. Compute both the sample correlation, r, and the Spearman correlation, $r_s$.
    b. Test the null hypothesis that the two variables are uncorrelated.
    c. Does the correlation coefficient do a good job of measuring the relationship? Why?

4. STATTOOLS was used to compare five different brands of allergy medicine. The Brands, labeled A-E, were each compared on six different subjects. The results are listed below. The higher the score the better the performance.

| ***OneWay ANOVA Table*** | **Sum of Squares** | **Degrees of Freedom** | **Mean Squares** | **F-Ratio** | **p-Value** |
|---|---|---|---|---|---|
| **Between Variation** | 3432.133 | 4 | 858.033 | 126.678 (*) | < 0.0001 |
| **Within Variation** | 169.333 | 25 | 6.773 | | |
| **Total Variation** | 3601.467 | 29 | | | |

| ***Confidence Interval Tests*** | **Difference of Means** | **No Correction** | | **Bonferroni** | |
|---|---|---|---|---|---|
| | | **Lower** | **Upper** | **Lower** | **Upper** |
| **A-B** | -7.333 | -10.4279761 | -4.238690566 | -11.95860616 | -2.70806051 |
| **A-C** | -3.333 | -6.4279761 | -0.238690566 | -7.958606156 | 1.29193949 |
| **A-D** | -14.833 | -17.9279761 | -11.73869057 | -19.45860616 | -10.20806051 |
| **A-E** | 17.333 | 14.23869057 | 20.4279761 | 12.70806051 | 21.95860616 |
| **B-C** | 4.000 | 0.905357233 | 7.094642767 | -0.625272823 | 8.625272823 |
| **B-D** | -7.500 | -10.59464277 | -4.405357233 | -12.12527282 | -2.874727177 |
| **B-E** | 24.667 | 21.5720239 | 27.76130943 | 20.04139384 | 29.29193949 |
| **C-D** | -11.500 | -14.59464277 | -8.405357233 | -16.12527282 | -6.874727177 |
| **C-E** | 20.667 | 17.5720239 | 23.76130943 | 16.04139384 | 25.29193949 |
| **D-E** | 32.167 | 29.0720239 | 35.26130943 | 27.54139384 | 36.79193949 |

a. What are the null and alternative hypotheses tested in (*)?
b. What is the difference between the "No Correction" intervals and the "Bonferroni" intervals?
c. Which Brands are different?
d. Which Brand(s) worked best? Explain.

5. Biologists have attempted to use the size of the individual islands in the Galapagos Islands to predict the number of species that live on the island. Below is a small set of data.

| | | | | | |
|---|---|---|---|---|---|
| **Island Size:** | 26 | 32 | 20 | 36 | 90 |
| **# Species:** | 23 | 33 | 24 | 35 | 65 |

a. Find the regression line using Island Size to predict Number of Species.
b. Estimate the number of species for an island size 60 square miles.
c. Do you have any concerns about reporting this result? Explain.