<u>06-713: Homework 4</u>
Due Oct. 10

1. (a) One famous event in National Football League History was the "perfect season" played by the Miami Dolphins in 1972. In that year, the Dolphin won all 17 games they played. In this problem, you will examine how unusual this is. First, assume that every team has a 50% chance of winning each game purely at random. If there are 32 teams in the NFL, how likely is it that one or more teams will win all 17 games in a season in one year? Based on this, how often should you expect a "perfect season" to happen?
   (b) We could refine our analysis by assuming that some teams really are better than other teams. If a team was good enough to win 65% of its games (again, at random), what is the probability that that team will play a perfect 17 game season in a given year?
   (c) Look up the Pittsburgh Pirates win/loss record for the 2007 season as of October 1. Considering each game as an independent measurement of the Pirates' ability (or lack therof) to win, provide an interval estimate for the Pirates' winning percentage using a 95% confidence interval.

2. (a) If events B and C are independent, $P(B \cap C) = P(B)P(C)$. Develop a χ^2 test to test if two outcomes, B and C are independent. (Hint: Divide experimental outcomes into $B \cap C, B \cap \overline{C}, \overline{B} \cap C, \overline{B} \cap \overline{C}$.)
   (b) In the manufacturing division you manage, 70% of the new hires are chemical engineers and 39% of the new hires leave in less than one year. Your CEO believes that turnover can be reduced by reducing the fraction of chemical engineers that are hired. You randomly sample your records and find that in 120 new hires, 42 chemical engineers left within one year compared to 6 non-chemical engineers. What do you recommend to your CEO?

3. A new production process is introduced at your plant. The number of defective units per hour observed in a 22 hour period are [3 0 5 4 2 5 4 1 5 3 6 4 0 7 3 2 4 4 6 5 7 9]. One member of your group believes the number of defective units follows a Poisson distribution, $P(x = k) = e^{-a} a^k / k!$, while another group member suggests a Binomial distribution, $P(x = k-1) = \binom{n}{k} p^k (1-p)^{n-k}$ with n=12. What values of a and p give acceptable hypotheses for the data with a significance level of 0.05?

4. [*Take home exam problem from 2006*] On the class website, an Excel spreadsheet containing numerical data from a set of experimental measurements is available. This data represents 47 distinct experiments, each of which measured three quantities, A, B, and C. The data has been ordered from the largest to the smallest values of A.

(a) Is it reasonable to correlate A, B, and C using a linear relation of the form $A = c_1 B + c_2 C + c_3$? Find the best values for c_i in this expression and explain why you think this expression is or is not justified by the data.

(b) If you assume that the values of A were sampled from a normal distribution, give interval estimates for the mean and variance of this distribution. Repeat this exercise for the values of B and the values of C.

(c) Is the assumption that you made in part (b) that the values of C were sampled from a normal distribution statistically plausible? If not, propose at least one other probability distribution that could be the source of the values of C and test whether the data is consistent with your proposed distribution.

Optional Problems

1. Consider two independent RVs with probability densities f(x) and g(x). The probability density of the sum of these RVs is given by a convolution integral, $(f+g)(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)dy$. Use this fact to examine the accuracy of the Central Limit Theorem for the sum of n independent RVs uniformly distributed on (0,1) for n=2 and 3.

2. Measuring a voltage with a new voltmeter in your lab introduces an error v with mean 0. Measuring a calibrated source with V=3.0 volts four times gives V=2.90, 3.15, 3.05, 2.96. Assume that v is normal and find the 0.95 confidence interval of the error's standard deviation. How useful is this result? What would you change if you were going to use this information in your research?