**Biology 683 Final Exam**

**1)** Analysis of our genome assembly showed that 127 genes were present in tandem arrays with at least 3 copies of each gene. These genes were significantly enriched relative to the rest of the genome for the GO term "indole biosynthetic process" (p-value = 0.015). Significance of the observed number of GO terms was determined by 1000 Monte Carlo simulations. Briefly describe in plain English what this means.

**2)**

What does the term indicated in red represent in Baye's Theorem

**3)** Use a GLM to analyze datafile 1 (growth of beetles on corn and wheat flour) from the course website. Copy and paste one equation that you used to analyze this data. Please paste the equation that you are using to estimate parameters from and are inferring something about the biology of this system. In analyzing this data it may be important to know that these beetles have a preferred temperature of 37c.

What predictor variables do you believe are important in determining the growth of beetles? Check all that apply.

Write a couple of sentences describing your findings. For instance, tell me what variables are important in determining growth and whether their effect is positive, negative, large, or small.

**4)** Upload a publication quality plot demonstrating an interesting relationship or relationships that you discovered in the data. Your figure should be a PDF file, vector art, and publication quality. The maximum file size is 10MB.

**5)** You have measured gene expression for 2000 genes in cell cultures under 10 different types of stress. You would like to know whether certain types of stress lead to similar changes in gene expression. Describe a method you might use to answer this question.

**6)** A longitudinal study tested whether a variety of lifestyle habits impacted longevity in 10000 60-year-old individuals over a 30 year period. Analyze datafile 2 from the course website and determine whether these risk categories have an important impact on longevity in the study.