STAT 471 **Quiz 2** **Newton: 3–22–06.**

Record all answers in the blue books. Show your work.

1. An $n \times p$ design matrix X and an $n \times 1$ response vector Y are used in a multiple linear regression model; the $p \times 1$ coefficient vector $\hat{\theta}$ satisfying $X^T X \hat{\theta} = X^T Y$ minimizes squared deviations between Y and its linear predictor $X\theta$.

   (a) What is the QR decomposition of X?

   (b) How can the QR decomposition be used to compute $\hat{\theta}$?

2. A study of genetic mutation in cells involves measuring m wells; in well i there is a large, known number n_i of cells. A random number Z_i of these have aquired a certain mutation. The model is $Z_i \sim \text{Poisson}(n_i\theta)$, where $\theta > 0$ is an unknown parameter. [Recall that a Poisson(λ) random variable Z has probability mass function $f(z) = \exp(-\lambda)\lambda^z/z!$ for $z = 0, 1, 2, \cdots$.]

   (a) Suppose that realized values z_i of Z_i are observed for all i.

      i. Write the likelihood $L(\theta)$ and the loglikelihood $l(\theta)$.

      ii. Derive the score $S(\theta) = l'(\theta)$ and compute the MLE $\hat{\theta}$.

      iii. Derive the Fisher information $I(\theta)$.

   (b) Suppose that the Z_i's are not observable; instead we can only observe the binary indicators $X_i = 1[Z_i > 0]$.

      i. What is the distribution of X_i?

      ii. Argue that for realized data $x_1, x_2, \ldots, x_n$, the likelihood is

      $$L(\theta) = \prod_{i=1}^{n} \{1 - \exp(-n_i\theta)\}^{x_i} \{\exp(-n_i\theta)\}^{1-x_i}.$$

      Note that unless all $x_i = 0$ or all $x_i = 1$, there is no closed form solution for the MLE $\hat{\theta}$.

      iii. Outline how Newton's method can be used to compute $\hat{\theta}$.

      iv. Compute $\hat{z}_i = E(Z_i | X_i = x_i)$ and outline how the EM algorithm can be used to compute $\hat{\theta}$.