Advanced Methods in Biostatistics 1 (140.751)
Fall 2005

Homework Assignment #3
(Due Wednesday, October 12, 2005)

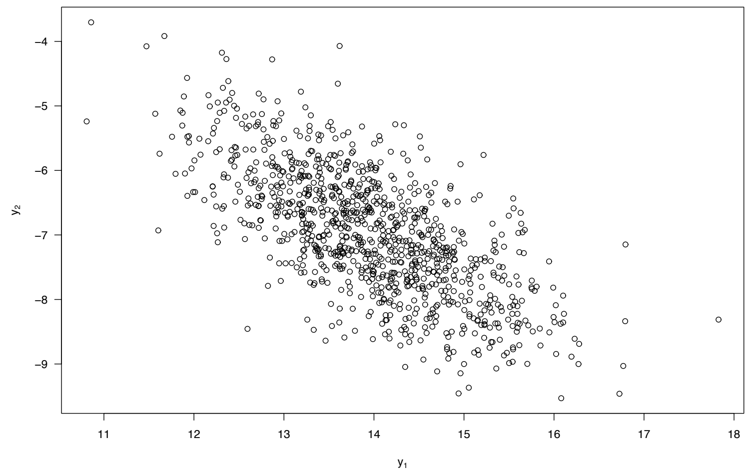
Please hand in a hard copy of your R code, and send an electronic version of it to Kenny (kshum@jhsph.edu).

1. Assume that $\mathbf{A}$ is a symmetric $n \times n$ matrix with real eigenvalues, and that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are linear independent eigenvectors of $\mathbf{A}$. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Show that:
   (a) $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ for some diagonal matrix $\mathbf{\Lambda}$.
   (b) $\mathbf{I} + \mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$ for some diagonal matrix $\mathbf{D}$.
   (c) If $\mathbf{D}$ is non-singular, then $\mathbf{I} + \mathbf{A}$ is non-singular, and $(\mathbf{I} + \mathbf{A})^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^{-1}$.

2. Let $\mathbf{J}$ be the $n \times n$ matrix with all entries equal to 1.
   (a) Show that $\mathbf{u}_1 = (1, \ldots, 1)'$, $\mathbf{u}_2 = (1, -1, 0, \ldots, 0)'$, $\mathbf{u}_3 = (1, 0, -1, 0, \ldots, 0)'$, $\mathbf{u}_4 = (1, 0, 0, -1, 0, \ldots, 0)'$, $\ldots$, $\mathbf{u}_n = (1, 0, \ldots, 0, -1)'$ are eigenvectors of $\mathbf{J}$, and that they are linearly independent.
   (b) Find $(\mathbf{I} + a\mathbf{J})^{-1}$ for the numbers $a \in \mathbb{R}$ for which the inverse exists.
   (c) Show that $(\mathbf{I} + a\mathbf{J})^{-1}$ is of the form $\mathbf{I} + b\mathbf{J}$, and find b.

3. Consider the random variables $Y_1, \ldots, Y_n$ defined as $Y_i = U + Z_i$, where $U \sim N(\xi, \tau^2)$, $Z_i \sim iidN(\mu, \sigma^2)$, and U and Z_i are independent. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$.
   (a) What is the distribution Y_i?
   (b) Find $cov(Y_i, Y_j)$ and $corr(Y_i, Y_j)$ for $i \neq j$.
   (c) What is the distribution of $\mathbf{Y}$?
   (d) Consider the estimator $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Is $\bar{Y}$ an unbiased estimator for $E[Y_1]$?
   (e) Let $\hat{\mathbf{Y}}$ be the $n-$vector whose entries are all $\bar{Y}$. Show that $\hat{\mathbf{Y}} = \mathbf{M}\mathbf{Y}$ for some projection matrix $\mathbf{M}$.
   (f) Consider the estimator $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. Show that $(n-1)S^2 = \mathbf{Y}'\mathbf{P}\mathbf{Y}$ for some projection matrix $\mathbf{P}$.
   (g) What is the distribution of S^2?
   (h) Is S^2 an unbiased estimator for $var(Y_1)$?
   (i) Let $\mathbf{V} = var(\mathbf{Y})$. Find the inverse $\mathbf{V}^{-1}$ and the determinant $det(\mathbf{V})$.

4. Write an R function `myrmvn()` that generates samples from a multivariate normal distribution, starting with the standard normal distribution (i. e. using `rnorm()`). Your function takes as arguments `mu` (the mean vector of length n), `sigma` (the $n \times n$ covariance matrix), and `hm` (the number of independent samples from the multivariate distribution).



Generate a scatter plot of 1000 independent samples of your favorite bivariate normal distribution (with mean not equal to (0,0)', and non-zero off-diagonal elements in the covariance matrix).

5. (a) Write an R function `myrchisq()` that generates independent random samples from the non-central χ^2 distribution, using only the R function `rnorm()`. Your function `myrchisq()` takes as arguments `n` (the number of independent samples), `df` (the degrees of freedom), and `lambda` (the non-centrality parameter).

   (b) Using the above `myrchisq(n,df,lambda)`, write a function `mypchisq(q,n,df,lambda)` that returns $F(q) = P(\chi^2_{df}(\lambda) \leq q)$, approximated by simulation.

   (c) How large do you have to choose `n` to guarantee that your estimate has a standard deviation less than 0.01?