Biost/Stat 533
Emerson, Spr 14
Homework #2
April 7, 2014

Written problems to be handed in Wednesday, April 16.

Under no circumstances may you refer to a homework key from this or other classes. While you may work with other students to derive a solution, when you write up your final solution, you may not refer to any other source. You must be able to develop your answer as if it were being done in a closed book, closed notes examination. You must provide a signed pledge to that effect:

On my honor I have neither given nor received unauthorized aid on the completion of this homework.

1. The χ^2, t, and F distributions are important "sampling distributions" commonly used in statistical inference. These distributions are derived as the exact distribution of certain statistics computed on normally distributed data. We are often ultimately interested the distribution-free interpretation of these statistics.

   a. Rigorously show that as n becomes large, a normal distribution provides a good approximation to the χ^2_n distribution. Make clear the parameters of the normal distribution, as well as the sense in which the approximation is valid.

   b. Rigorously show that as n becomes large, a normal distribution provides a good approximation to the t_n distribution. Make clear the parameters of the normal distribution, as well as the sense in which the approximation is valid.

   c. Rigorously show that as n becomes large, a χ^2 distribution provides a good approximation to the $F_{m,n}$ distribution. Make clear the parameters of the χ^2 distribution, as well as the sense in which the approximation is valid.

2. Let $\mathbf{A}$ be a $p \times p$ invertible matrix that is partitioned into $r \times r$ matrix $\mathbf{R}$, $s \times s$ matrix $\mathbf{S}$, and $r \times s$ matrix $\mathbf{T}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{T}^T & \mathbf{S} \end{bmatrix}.$$

   Prove that the inverse of $\mathbf{A}$ can be found as either

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{R}^{-1} + \mathbf{R}^{-1}\mathbf{T}(\mathbf{S} - \mathbf{T}^T\mathbf{R}^{-1}\mathbf{T})^{-1}\mathbf{T}^T\mathbf{R}^{-1} & -\mathbf{R}^{-1}\mathbf{T}(\mathbf{S} - \mathbf{T}^T\mathbf{R}^{-1}\mathbf{T})^{-1} \\ -(\mathbf{S} - \mathbf{T}^T\mathbf{R}^{-1}\mathbf{T})^{-1}\mathbf{T}^T\mathbf{R}^{-1} & (\mathbf{S} - \mathbf{T}^T\mathbf{R}^{-1}\mathbf{T})^{-1} \end{bmatrix}$$

   or

$$\mathbf{A}^{-1} = \begin{bmatrix} (\mathbf{R} - \mathbf{T}\mathbf{S}^{-1}\mathbf{T}^T)^{-1} & -(\mathbf{R} - \mathbf{T}\mathbf{S}^{-1}\mathbf{T}^T)^{-1}\mathbf{T}\mathbf{S}^{-1} \\ -\mathbf{S}^{-1}\mathbf{T}^T(\mathbf{R} - \mathbf{T}\mathbf{S}^{-1}\mathbf{T}^T)^{-1} & \mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{T}^T(\mathbf{R} - \mathbf{T}\mathbf{S}^{-1}\mathbf{T}^T)^{-1}\mathbf{T}\mathbf{S}^{-1} \end{bmatrix}.$$

For problems 3 - 7, we consider a linear regression model in which we have $p-1$ covariates $X_1, \ldots, X_{p-1}$ with outcome variable Y_i modeled as

$$(Y_i \,|\, \vec{X}_i = \vec{x}_i) = \beta_0 + x_{i,1}\beta_1 + \cdots + x_{i,p-1} + \epsilon_i \qquad \text{(or in matrix notation} \quad (\vec{Y} \,|\, \mathbf{X}) = \mathbf{X}\vec{\beta} + \vec{\epsilon}).$$

We consider least squares estimators (LSE) that minimize the sum of squared errors

$$SSE = \sum_{i=1}^{n} (Y_i - \vec{x}_i^T \vec{\beta})^2 \qquad \Rightarrow \qquad \hat{\vec{\beta}} = \text{argmin}_{\vec{\beta} \in \mathcal{R}^p} (\vec{Y} - \mathbf{X}\vec{\beta})^T (\vec{Y} - \mathbf{X}\vec{\beta}).$$

Differentiating the RHS wrt $\vec{\beta}$ and setting equal to $\vec{0}$:

$$\frac{\partial}{\partial \vec{\beta}} (\vec{Y} - \mathbf{X}\vec{\beta})^T (\vec{Y} - \mathbf{X}\vec{\beta})|_{\vec{\beta} = \hat{\vec{\beta}}} = \vec{0}$$

yields

$$\mathbf{X}^T \mathbf{X} \hat{\vec{\beta}} = \mathbf{X}^T \vec{Y}.$$

When $\mathbf{X}$ is of full rank, then $\mathbf{X}^T\mathbf{X}$ is invertible, and we have LSE

$$\hat{\vec{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \vec{Y}.$$

3. Consider linear regression model $(Y_i \,| X_i = x_i) = \beta_0 + x_i\beta_1 + \epsilon_i$. Derive expressions for LSE $\hat{\beta}_0$ and $\hat{\beta}_1$ in terms of simple sample means, sample variances, and sample covariances, instead of matrix notation.
4. Consider linear regression model $(Y_i \,| X_i = x_i, W_i = w_i) = \gamma_0 + x_i\gamma_1 + w_i\gamma_2 + \epsilon_i^*$, where it is assumed that the design matrix for the problem would be of full rank. Derive expressions for LSE $\hat{\gamma}_0$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ in terms of simple sample means, sample variances, and sample covariances, instead of matrix notation.
5. Consider the settings of problems 2 and 3, along with a linear regression model $(W_i \,|\, X_i = x_i) = \alpha_0 + x_i\alpha_+\epsilon_i^{**}$ having LSE $\hat{\vec{\alpha}}$. Now suppose we define new variables as the residuals $Z_i = Y_i - (\hat{\beta}_0 + x_i\hat{\beta}_1)$ and $U_i = W_i - (\hat{\alpha}_0 + x_i\hat{\alpha}_1)$. Show that the LSE $\hat{\eta}_1$ computed for linear regression model $(Z_i \,|\, U_i = u_i) = \eta_0 + u_i\eta_1 + \epsilon_i^{\#}$ is equal to the LSE $\hat{\gamma}_2$ from problem 4.
6. Under the assumption that $E[\epsilon_i] = E[\epsilon_i^*] = 0$, derive expressions for β_0 and β_1 in terms of γ_0, γ_1, and γ_2. Simplify these expressions in the case where X is a binary variable.
7. Generalize the results of problem 6 to compare the slope from the unadjusted model for binary predictor of interest X to the corresponding slope for X from a model adjusting for an arbitrary number of additional covariates.