Biost/Stat 533
Emerson, Spr 14
Homework #6
May 21, 2014

Written problems to be handed in Wednesday, May 28.

**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. Suppose $W_i$ is a categorical variable taking on one of the values $a_1, a_2, \ldots, a_p$. Consider a linear regression model

$$\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$$

in which $\vec{\beta}^T = (\beta_0, \ldots, \beta_{p-1})$ and $\vec{W}$ is modeled with dummy variables. That is, we consider a model where the first column of the design matrix is filled with 1's (so we are fitting an intercept), and the $j$th column of the design matrix is an indicator that $W_i = a_j$ for $j = 2, \ldots, p$ (so $X_{i1} = 1$, and for $j = 2, \ldots, p$, $X_{ij} = 1$ if $W_i = a_j$ and $X_{ij} = 0$ otherwise). Assume that $var(\vec{\epsilon}) = \sigma^2 \mathbf{I}_n$.

   a. Find expressions for $\widehat{\vec{\beta}}$ in terms of the group sample means $\overline{Y}_j$ where

$$\overline{Y}_j = \sum_{i=1}^{n} Y_i 1_{[W_i = a_j]} / \sum_{i=1}^{n} 1_{[W_i = a_j]}$$

   for $j = 1, \ldots, p$.

   b. Derive an asymptotic test of $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$. Show that this is equivalent to a one-way analysis of variance to compare $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$, where it is assumed that independent observations $Y_i \sim (\mu_j, \sigma^2)$ when $W_i = a_j$.

2. Let independent random vectors $(X_i, Y_i)$ for $i = 1, \ldots, n$ be distributed according to a bivariate normal distribution with $X_i \sim (\mu, \sigma^2)$, $Y_i \sim (\nu, \tau^2)$, and $corr(X_i, Y_i) = \rho$. Let $\vec{X} = (X_1, \ldots, X_n)^T$ and $\vec{Y} = (Y_1, \ldots, Y_n)^T$.

   a. Derive the conditional distribution of $Y_i | X_i = x$ and $X_i | Y_i = y$.

   b. Suppose we fit linear regression model $\vec{Y} = \beta_0 + \beta_1 \vec{X} + \vec{\epsilon}$. Is asymptotic inference for OLSE of the regression parameters valid for this model? Justify your answer. For what function of parameters $\mu$, $\nu$, $\sigma^2$, $\tau^2$, and $\rho$ is OLSE $\widehat{\vec{\beta}}$ an unbiased estimator?

   c. Suppose we fit linear regression model $\vec{X} = \gamma_0 + \gamma_1 \vec{Y} + \vec{\delta}$. Is asymptotic inference for OLSE of the regression parameters valid for this model? Justify your answer. For what function of parameters $\mu$, $\nu$, $\sigma^2$, $\tau^2$, and $\rho$ is OLSE $\widehat{\vec{\gamma}}$ an unbiased estimator?

   d. Under what conditions will $y = \hat{\beta}_0 + \hat{\beta}_1 x$ and $x = \hat{\gamma}_0 + \hat{\gamma}_1 y$ be the same line?

3. Consider again the setting of problem 2.

   a. Show that OLSE estimator $\widehat{\vec{\beta}}$ is the maximum likelihood estimator.

   b. Show that OLSE estimator $\widehat{\vec{\beta}}$ is UMVUE.

c. Show that OLSE estimator $\widehat{\vec{\beta}}$ is efficient.

4. Consider an "error in the variables" model in which there is a true relationship between response $Y$ and predictor $W$ given by $Y = \beta_0 + \beta_1 W + \epsilon$ with $\epsilon_i \sim (0, \sigma^2)$ totally independent. Suppose that $W$ is unobserved, and we instead have $Z$, an imprecise measurement of $W$ which follows the relation $Z = \alpha_0 + \alpha_1 W + \delta$, with $\delta_i \sim (0, \tau^2)$ totally independent of each other and the $\epsilon$'s. We then fit a regression model $E[Y] = \gamma_0 + \gamma_1 Z$, and use this model to make inference about an association between $Y$ and $W$.

   a. Under what conditions is OLSE $\hat{\gamma}_1$ unbiased for $\beta_1$?

   b. What can we say in this general setting about the standard error of $\hat{\gamma}_1$ compared to the standard error of $\hat{\beta}_1$ (if we had $W$)?

   c. Now suppose we further assume that $W_i$, $\delta_i$, and $\epsilon_i$ are jointly normally distributed and totally independent. How does the standard error of $\hat{\gamma}_1$ compare to the standard error of $\hat{\beta}_1$ (if we had $W$)? What does this suggest about our ability to test for associations in such a model? How much do we lose by having errors in the predictors?