201b Homework 2, due Wed Feb 13 2008

1. Suppose that $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, for $i = 1, 2, ..., n$, with $E(\epsilon_i) = 0$, and $cov(\epsilon) = \sigma^2 I$, and suppose also that $\sum_{i=1}^{n} x_i = 0$. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimates of β_0 and β_1, respectively. Show that $cov(\hat{\beta}_0, \hat{\beta}_1) = 0$.

2. Suppose that $z_i = \alpha_0 + \alpha_1 w_i + \delta_i$, for $i = 1, 2, ..., n$, where α_0 and α_1 are constants, z and w are vectors of length n, and δ is a vector of random variables with $E(\delta_i) = 0$, and for some constants b_i, the covariance matrix of δ satisfies $cov(\delta) = V$, where V is an $(n \times n)$ diagonal matrix such that $V[i, i] = b_i$.

(a) Re-write this model in terms of an ordinary univariate regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, and show that $E(\epsilon_i) = 0$, and $cov(\epsilon) = \sigma^2 I$.

(b) What are the least squares estimates of α_0 and α_1?

(c) Let $\hat{\alpha}_0$ and $\hat{\alpha}_1$ denote the estimates in part (b). What are $var(\hat{\alpha}_0)$ and $var(\hat{\alpha}_1)$?

3. Suppose that $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, for $i = 1, 2, ..., n+1$, with $E(\epsilon_i) = 0$, and $cov(\epsilon) = \sigma^2 I$. First, n observations of x and y are recorded, and the ordinary linear regression model is fit. Then later, a new observation (x_{n+1}, y_{n+1}), is recorded. Let $\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$.

(a) What is the variance of the predicted y-value corresponding to x_{n+1}? (In other words, what is $var(\hat{y}_{n+1})$?)

(b) What is the variance of the (n+1)st residual, $e_{n+1} = \hat{y}_{n+1} - y_{n+1}$?

4. For the standard multi-variate regression model $y = X\beta + \epsilon$, where y is a vector of length n, X is a fixed matrix of size $n \times p$ whose first column is all 1's (i.e. there is an intercept), and with $E(\epsilon_i) = 0$, and $cov(\epsilon) = \sigma^2 I$, let $e = y - \hat{y}$ denote the vector of residuals.

(a) Show that e is orthogonal to each column of X.

(b) Show that $\sum_{i=1}^{n} e_i = 0$.

5. Suppose that $y = x_1\beta_1 + x_2\beta_2 + \epsilon$, where y, x_1, and x_2 are fixed vectors of length n, and $E(\epsilon) = 0$ and $cov(\epsilon) = \sigma^2 V$, where σ^2 and V are known. Derive an appropriate test statistic for the hypothesis $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$. What is the distribution of the test statistic under the null hypothesis?

6. In what types of situations will an M-estimator be unlikely to give satisfactory results? Construct a fake dataset with a single explanatory variable (x) and $n = 10$ observations, to illustrate. In addition, construct another example dataset with two explanatory variables (x_1 and x_2), and $n = 10$ observations, where again the M-estimator would be unlikely to give satisfactory results.