University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Practice problems - week 8

Answer the following questions:

a. Consider the multiple regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Suppose we are estimating $\boldsymbol{\beta}$ under the equality constraints $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbf{C}$ is $m \times (k+1)$ matrix of known constants and $\boldsymbol{\gamma}$ is $m \times 1$ vector of known constants. Show that $\mathbf{e_c}'\mathbf{e_c} = \mathbf{e}'\mathbf{e} + (\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}_{\boldsymbol{c}} - \hat{\boldsymbol{\beta}})$, where $\mathbf{e_c}$ are the constrained residuals and $\hat{\boldsymbol{\beta}}_{\boldsymbol{c}}$ is the constrained least squares estimator of $\boldsymbol{\beta}$.

b. Refer to question (a). Find $\text{cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{c}})$.

c. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. Show that the regression sum of squares can be expressed as: $SSR = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\bar{y}^2$. Find $E(SSR)$ using this equation.

d. Consider the multiple regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, subject to a set of m linear restrictions of the form $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$. The matrix $\mathbf{C}$ is $m \times (k+1)$ and $\boldsymbol{\gamma}$ is $m \times 1$ vector. Suppose $\mathbf{C}$ is partitioned as $(\mathbf{C_1}, \mathbf{C_2})$, where $\mathbf{C_2}$ is nonsingular. (The columns of $\mathbf{C_2}$ are the last m columns of $\mathbf{C}$). Transform the model to the canonical form as $\mathbf{Yr} = \mathbf{Xr}\boldsymbol{\beta_1} + \boldsymbol{\epsilon}$ and explain how to estimate $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\beta}}_{\mathbf{1c}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{2c}}$ be the two constrained sub-vectors using the canonical form. Give an expression of $\text{cov}(\hat{\boldsymbol{\beta}}_{\mathbf{1c}}, \hat{\boldsymbol{\beta}}_{\mathbf{2c}})$.

e. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The Gauss-Markov conditions hold and also $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Evaluate $E\left(\mathbf{Y}'\mathbf{A}\mathbf{Y} - \sigma^2\right)^2$ where $\mathbf{A} = \frac{1}{n-k+1}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]$.

f. Let $\mathbf{X_1}$ equal a constant (column of ones) plus three predictors. Let $\mathbf{X_2}$ contain three more predictors. Let $\mathbf{y}$ be the response variable. The Gauss-Markov conditions hold. Regress each of the three variables in $\mathbf{X_2}$ on $\mathbf{X_1}$ and obtain the residuals $\mathbf{X_2^*}$. Regress $\mathbf{y}$ on $\mathbf{X_1}$ and $\mathbf{X_2^*}$. How do your results compare to the results of the regression of $\mathbf{y}$ on $\mathbf{X_1}$ and $\mathbf{X_2}$? The comparison you are making is between the least squares coefficients of the two regression models. Derive the result theoretically.

g. Consider the multiple regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ subject to a set of linear constraints of the form $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbf{C}$ is $m \times (k+1)$ matrix and $\boldsymbol{\gamma}$ is $m \times 1$ vector. The Gauss-Markov conditions hold and also $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Transform the model into the canonical form and find the distribution of $\frac{\left(\hat{\boldsymbol{\beta}}_{\mathbf{1c}} - \boldsymbol{\beta}_{\mathbf{1c}}\right)'\mathbf{X}_{\mathbf{1r}}'\mathbf{X}_{\mathbf{1r}}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1c}} - \boldsymbol{\beta}_{\mathbf{1c}}\right)}{\sigma^2}$.

h. Consider the multiple regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, subject to a set of m linear restrictions of the form $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$ with $\boldsymbol{\gamma} \neq \mathbf{0}$. The matrix $\mathbf{C}$ is $m \times (k+1)$ and it can be partitioned in $(\mathbf{C_1}, \mathbf{C_2})$, where either $\mathbf{C_1}$ or $\mathbf{C_2}$ is nonsingular. Transform the model to the canonical form as $\mathbf{Yr} = \mathbf{Xr}\boldsymbol{\beta_1} + \boldsymbol{\epsilon}$ and explain how to estimate the slopes of the transformed columns of $\mathbf{X}$ using partial regression.

i. Consider the multiple regression model. Show that $\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \mathbf{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')'(\mathbf{H} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{Y} = SSR$, where $\mathbf{1} = (1, 1, \ldots, 1)'$.

j. Consider the usual hat matrix $\mathbf{H}$ which is symmetric and idempotent. We can use spectral decomposition and write $\mathbf{H} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$. The eigenvalues of $\mathbf{H}$ are 0 or 1, and there $k+1$ eigenvalues equal to 1, because $tr(\mathbf{H}) = k+1$. Partition $\mathbf{P} = (\mathbf{E}, \mathbf{F})$ where $\mathbf{E}$ is $n \times (k+1)$ and show that $\mathbf{E}\mathbf{E}' = \mathbf{H}$ and $\mathbf{E}'\mathbf{E} = \mathbf{I}_{k+1}$.