University of California, Los Angeles
Department of Statistics

Statistics 100C **Instructor: Nicolas Christou**

Homework 6

Exercise 1
Consider the regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \epsilon_i, \quad i = 1, \ldots, n.$$

Also, $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$, and $\text{var}(\epsilon_i) = \sigma^2$. Define the matrix $\mathbf{C}$ and the vector $\boldsymbol{\gamma}$ using the notation discussed in class for testing the following hypotheses:

a. $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$.

b. $\beta_1 = 4\beta_6$.

c. $\beta_5 = 0$.

d. $\beta_5 = \beta_6$.

e. $\beta_2 = \beta_4 + \beta_5$.

Exercise 2
Consider the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Let $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. Show that $\hat{\mathbf{y}}$ and $\mathbf{e}$ are statistically independent.

Exercise 3
Use the restaurant data from homework 5, exercise 4:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics100C/restaurant.txt", header=TRUE)
```

and Consider the multiple regression model

$$cost_i = \beta_0 + \beta_1 food_i + \beta_2 decor_i + \beta_3 ser_i + \epsilon_i$$

a. Test the hypothesis using $\alpha = 0.05$:

$$H_0 : (\beta_1, \beta_2, \beta_3)' = \mathbf{0}$$
$$H_a : (\beta_1, \beta_2, \beta_3)' \neq \mathbf{0}$$

Use the F test discussed in class with the appropriate $\mathbf{C}$ matrix and $\boldsymbol{\gamma}$ vector. Submit the `R` code for computing this test statistic with your conclusion.

b. Test the hypothesis $\alpha = 0.05$:

$$H_0 : (\beta_1, \beta_3)' = \mathbf{0}$$
$$H_a : (\beta_1, \beta_3)' \neq \mathbf{0}$$

Use the F test discussed in class with the appropriate $\mathbf{C}$ matrix and $\boldsymbol{\gamma}$ vector. Submit the `R` code for computing this test statistic with your conclusion.

Hint: Here is the test statistic for (a) and (b):

$$F_{m,n-k-1} = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})' \left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})}{m s_e^2},$$

where m is the rank of the matrix $\mathbf{C}$. For example, when we test the overall significance of the model ($\beta_1 = \beta_2 = \ldots = \beta_k = 0$) we have $m = k$. Reject H_0 if $F_{m,n-k-1} > F_{\alpha;m,n-k-1}$.

Exercise 4

Consider the bivariate random vector with variance covariance matrix

$$\mathbf{A} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \text{where } \rho \text{ is the correlation coefficient } (-1 \leq \rho \leq 1).$$

a. Show that the eigenvalues of the matrix $\mathbf{A}$ are $\lambda_1 = 1 + \rho$ and $\lambda_2 = 1 - \rho$.

b. Show that the normalized eigenvectors that correspond to these two eigenvalues are:

$$\mathbf{v_1} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ and } \mathbf{v_2} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

c. Verify that $\mathbf{P\Lambda P'} = \mathbf{A}$, where $\mathbf{P}$ is the matrix with columns the two normalized eigenvectors from part (b).

d. Generate $n = 50$ independent random vectors from the bivariate normal distribution with mean zero and variance covariance matrix $\mathbf{A}$.

   Hint: The vector $\mathbf{y} = (y_1, y_2)'$ can be generated as follows:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

   where X_1, X_2 are independent with $X_1 \sim N(0, \sqrt{1+\rho})$ and $X_2 \sim N(0, \sqrt{1-\rho})$. Using `R` you can easily generate random samples from X_1 an X_2 and then use the expression above to generate random vectors (y_1, y_2). Note: Choose a value of $-1 < \rho < 1$.