STA 302f13 Assignment Six[1]

These problems are preparation for the quiz in tutorial on Friday October 25th, and are not to be handed in. For reference, the general linear model with normal error terms is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the columns of $\mathbf{X}$ are linearly independent, and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

1. In the general linear model, what is the distribution of $\mathbf{Y}$?
2. You know that the least squares estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. What is the distribution of $\widehat{\boldsymbol{\beta}}$? Show the calculations.
3. Let $\widehat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. What is the distribution of $\widehat{\mathbf{Y}}$? Show the calculations.
4. Let the vector of residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}}$. What is the distribution of $\hat{\boldsymbol{\epsilon}}$? Show the calculations. Simplify both the expected value (which is zero) and the covariance matrix.
5. Recall from an earlier homework problem that if $\mathbf{T}$ is a random vector with expected value $\boldsymbol{\mu}$, then $cov(\mathbf{T}) = E(\mathbf{T}\mathbf{T}') - \boldsymbol{\mu}\boldsymbol{\mu}'$. Using this fact, give expressions for
   (a) $E(\mathbf{Y}\mathbf{Y}')$
   (b) $E(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}')$

   These may be helpful in the next question.
6. For the general linear regression model, show that the $n \times (k+1)$ matrix of covariances $C(\hat{\boldsymbol{\epsilon}}, \widehat{\boldsymbol{\beta}}) = \mathbf{0}$. Why does this show that $SSE = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$ and $\widehat{\boldsymbol{\beta}}$ are independent?
7. In Assignment 4, you proved that

   $$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

   Starting with this expression, show that $SSE/\sigma^2 \sim \chi^2(n-k-1)$. A result you proved in Assignment 2 will be useful.
8. For the general fixed effects linear regression model, tests and confidence intervals for linear combinations of regression coefficients are very useful. Derive the appropriate t distribution and some applications by following these steps. Let $\mathbf{a}$ be a $p \times 1$ vector of constants.
   (a) What is the distribution of $\mathbf{a}'\widehat{\boldsymbol{\beta}}$? Show a little work. Your answer includes both the expected value and the variance.
   (b) Now standardize the difference (subtract off the mean and divide by the standard deviation) to obtain a standard normal.

[1] Copyright information is at the end of the last page.

(c) Divide by the square root of a well-chosen chi-squared random variable, divided by its degrees of freedom, and simplify. Call the result T.

(d) How do you know numerator and denominator are independent?

(e) Suppose you wanted to test $H_0 : \mathbf{a}'\boldsymbol{\beta} = c$. Write down a formula for the test statistic.

(f) Suppose you wanted to test $H_0 : \beta_2 = 0$. Give the vector $\mathbf{a}$.

(g) Suppose you wanted to test $H_0 : \beta_1 = \beta_2$. Give the vector $\mathbf{a}$.

(h) Letting $t_{\alpha/2}$ denote the point cutting off the top $\alpha/2$ of the t distribution with $n - k - 1$ degrees of freedom, derive the $(1 - \alpha) \times 100\%$ confidence interval for $\mathbf{a}'\boldsymbol{\beta}$.

9. Letting $SST = \sum_{i=1}^n (Y_i - \overline{Y})^2$, $SSE = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2$ and $SSR = \sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2$, show $SST = SSR + SSE$.

10. Show that $\overline{Y}$ is a function of $\widehat{\boldsymbol{\beta}}$. Why does this establish that SSR and SSE are independent?

11. If $H_0 : \beta_1 = \cdots = \beta_k = 0$ is true,

    (a) What is the distribution of Y_i?

    (b) What is the distribution of $\frac{SST}{\sigma^2}$? Just write down the answer. You already did it in Assignment 2.

12. Still assuming $H_0 : \beta_1 = \cdots = \beta_k = 0$ is true, what is the distribution of SSR/σ^2? Again you may use material from Assignment 2.

13. Suppose $H_0 : \beta_1 = \cdots = \beta_k = 0$ were *false*. Would you expect SSR to be bigger, or would you expect it to be smaller? Which one, and why?

14. Recall the definition of the F distribution. If $W_1 \sim \chi^2(\nu_1)$ and $W_2 \sim \chi^2(\nu_2)$ are independent, $F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2)$. How do you know $F = \frac{SSR/k}{SSE/(n-k-1)}$ has an F distribution under $H_0 : \beta_1 = \cdots = \beta_k = 0$? List the numbers of the questions that establish the necessary facts.
