STA 302f15 Assignment Eight[1]

In the general linear model, assume that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ Also assume that the columns of the $\mathbf{X}$ matrix are linearly independent, so that the formulas for $\widehat{\boldsymbol{\beta}}$ and related quantities apply. You may use anything from the formula sheet unless you are explicitly asked to prove it, or are instructed otherwise. Use moment-generating functions *only* if the question directly asks you to do it.

1. Label each of the following statements True (meaning always true) or False (meaning not always true), and show your work or explain.
   (a) $\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
   (b) $\mathbf{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\epsilon}}$.
   (c) $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\epsilon}}$
   (d) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$
   (e) $\mathbf{X}'\boldsymbol{\epsilon} = \mathbf{0}$
   (f) $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$.
   (g) $\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \mathbf{0}$
   (h) $\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \mathbf{y}'\widehat{\boldsymbol{\epsilon}}$.
   (i) $W = \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{\sigma^2}$ has a chi-squared distribution.
   (j) $E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) = 0$
   (k) $E(\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}) = 0$

2. What is the distribution of $\mathbf{s}_1 = \mathbf{X}'\boldsymbol{\epsilon}$? Show the calculation of expected value and variance-covariance matrix.

3. What is the distribution of $\mathbf{s}_2 = \mathbf{X}'\widehat{\boldsymbol{\epsilon}}$?
   (a) Answer the question.
   (b) Show the calculation of expected value and variance-covariance matrix.
   (c) Is this a surprise? Answer Yes or No.
   (d) What is the probability that $\mathbf{s}_2 = \mathbf{0}$? The answer is a single number.

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

4. The following are some distribution facts you are expected to know. Just give the answers. Only re-derive them if you can't remember.

   (a) Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, where a and b are constants. What is the distribution of Y?

   (b) Let $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$. What is the distribution of Z?

   (c) Let $Z \sim N(0, 1)$. What is the distribution of $Y = Z^2$?

   (d) Let $X_1, \ldots, X_n$ independent $N(\mu, \sigma^2)$ random variables. What is the distribution of the sample mean $\overline{X}$?

   (e) Let $X_1, \ldots, X_n$ independent $N(\mu, \sigma^2)$ random variables. What is the distribution of $Z = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$?

   (f) Let $W_1, \ldots, W_n$ be independent $\chi^2(1)$ random variables. What is the distribution of $Y = \sum_{i=1}^n W_i$?

   (g) Let $X_1, \ldots, X_n$ independent $N(\mu, \sigma^2)$ random variables. What is the distribution of $Y = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2$?

   (h) Let $Y = X_1 + X_2$, where X_1 and X_2 are independent, $X_1 \sim \chi^2(\nu_1)$ and $Y \sim \chi^2(\nu_1 + \nu_2)$, where ν_1 and ν_2 are both positive. What is the distribution of X_2?

5. In an earlier Assignment, you proved that

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

   Starting with this expression, show that $SSE/\sigma^2 \sim \chi^2(n - k - 1)$. Use the formula sheet.

6. The t distribution is defined as follows. Let $Z \sim N(0, 1)$ and $W \sim \chi^2(\nu)$, with Z and W independent. Then $T = \frac{Z}{\sqrt{W/\nu}}$ is said to have a t distribution with ν degrees of freedom, and we write $T \sim t(\nu)$.

   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.

   (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) For a regression model with four independent variables, suppose you wanted to test $H_0 : \beta_2 = 0$. Give the vector $\mathbf{a}$.

(g) For a regression model with four independent variables, 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 $\mathbf{a}'\boldsymbol{\beta}$. "Derive" means show the High School algebra.

7. For a multiple regression model with an intercept, let $SST = \sum_{i=1}^n (Y_i - \overline{Y})^2$, $SSE = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2$ and $SSR = \sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2$, show $SST = SSR + SSE$.

8. Still for a multiple regression model with an intercept, show that $\overline{Y}$ is a function of $\widehat{\boldsymbol{\beta}}$. Why does this establish that SSR and SSE are independent?

9. Continue assuming that the regression model has an intercept. 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, and again in Assignment 5.

10. Still assuming $H_0 : \beta_1 = \cdots = \beta_k = 0$ is true, what is the distribution of SSR/σ^2? Use the formula sheet and show your work.

11. 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, $F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2)$. Show that $F = \frac{SSR/k}{SSE/(n-k-1)}$ has an F distribution under $H_0 : \beta_1 = \cdots = \beta_k = 0$? Refer to the results of questions above as you use them.

12. The null hypothesis $H_0 : \beta_1 = \cdots = \beta_k = 0$ is less and less believable as R^2 becomes larger. Show that the F statistic of Question 11 is an increasing function of R^2 for fixed n and k. This mean it makes sense to reject H_0 for large values of F.
