STA 302f16 Assignment Six[1]

These problems are preparation for the quiz in tutorial on Thursday October 27th, and are not to be handed in.

1. Let the continuous random vectors $\mathbf{y_1}$ and $\mathbf{y_2}$ be independent. Show that their joint moment-generating function is the product of their moment-generating functions. Since $\mathbf{y_1}$ and $\mathbf{y_2}$ are continuous, you will integrate. It is okay to represent a multiple integral with a single integral sign. Start with the partitioned random vector

$$\mathbf{y} = \left(\begin{array}{c} \mathbf{y_1} \\ \hline \mathbf{y_2} \end{array} \right) \text{ and corresponding } \mathbf{t} = \left(\begin{array}{c} \mathbf{t_1} \\ \hline \mathbf{t_2} \end{array} \right)$$

2. Recall that $\mathbf{e} \sim N(\mathbf{0}, \sigma^2(I - H))$. What is the distribution of $\mathbf{w} = \mathbf{X}'\,\mathbf{e}$?
   (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 $\mathbf{w} = \mathbf{0}$? The answer is a single number.

3. In the multiple linear regression model, let the columns of the X matrix be linearly independent, so that the columns of $X'X$ are linearly independent as well (no need for a proof because you did this in an earlier assignment). Either (a) show that $(X'X)^{-1/2}$ is symmetric, or (b) show by a simple numerical example that $(X'X)^{-1/2}$ may not be symmetric.

4. Assume the general linear regression model with normal errors. Label each of the following statements True (meaning always true) or False (meaning not always true). You should be able to justify your answers. It may help to know that four of the statements are true.
   (a) $\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
   (b) $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$.
   (c) $\widehat{\mathbf{y}} = \mathbf{X}\mathbf{b} + \mathbf{e}$
   (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) $\mathbf{e}'\,\mathbf{e} = \mathbf{0}$
   (h) $\mathbf{e}'\,\mathbf{e} = \mathbf{y}'\,\mathbf{e}$.
   (i) $w = \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{\sigma^2}$ has a chi-squared distribution.
   (j) $E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) = 0$
   (k) $E(\mathbf{e}'\,\mathbf{e}) = 0$

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

5. This is a repeat from last week. Just write down the answers.

   (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 $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $y = \sum_{i=1}^n x_i$?

   (d) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of the sample mean $\overline{x}$?

   (e) Let $x_1, \ldots, x_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $z = \frac{\sqrt{n}(\overline{x}-\mu)}{\sigma}$?

   (f) Let $x_1, \ldots, x_n$ be independent random variables, with $x_i \sim N(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ be constants. What is the distribution of $y = \sum_{i=1}^n a_i x_i$?

   (g) Let $x_1, \ldots, x_n$ be independent random variables with $x_i \sim \chi^2(\nu_i)$ for $i = 1, \ldots, n$. What is the distribution of $y = \sum_{i=1}^n x_i$?

   (h) Let $z \sim N(0, 1)$. What is the distribution of $y = z^2$?

   (i) Let $x_1, \ldots, x_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. What is the distribution of $y = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$?

   (j) 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?

6. Show that if $\mathbf{w} \sim N_p(\boldsymbol{\mu}, \Sigma)$, with Σ positive definite, then $y = (\mathbf{w} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{w} - \boldsymbol{\mu})$ has a chi-squared distribution with p degrees of freedom.

7. Let $y_1, \ldots, y_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. The sample variance is $s^2 = \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{n-1}$.

   (a) Show $Cov(\overline{y}, (y_j - \overline{y})) = 0$ for any $j = 1, \ldots, n$.

   (b) How do you know that $\overline{y}$ and s^2 are independent?

   (c) Show that

   $$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

   Hint: $\sum_{i=1}^n (y_i - \mu)^2 = \sum_{i=1}^n (y_i - \overline{y} + \overline{y} - \mu)^2 = \ldots$

8. Recall the definition of the t distribution. If $z \sim N(0, 1)$, $w \sim \chi^2(\nu)$ and z and w are independent, then $t = \frac{z}{\sqrt{w/\nu}}$ is said to have a t distribution with ν degrees of freedom, and we write $t \sim t(\nu)$. As in the last question, let $y_1, \ldots, y_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Show that $t = \frac{\sqrt{n}(\overline{y}-\mu)}{s} \sim t(n-1)$.

9. For the general linear regression model with normal error terms, show that the $n \times (k+1)$ matrix of covariances $C(\mathbf{e}, \mathbf{b}) = \mathbf{0}$. Why does this show that $SSE = \mathbf{e}'\mathbf{e}$ and $\mathbf{b}$ are independent?

10. Calculate $C(\mathbf{e}, \widehat{\mathbf{y}})$; show your work. Why should you have known this answer without doing the calculation, assuming normal error terms? Why does the assumption of normality matter?

11. In an earlier Assignment, you proved that

$$(\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta}) = \mathbf{e}'\,\mathbf{e} + (\mathbf{b} - \boldsymbol{\beta})'X'X(\mathbf{b} - \boldsymbol{\beta}).$$

Starting with this expression and assuming normality, show that $\mathbf{e}'\,\mathbf{e}/\sigma^2 \sim \chi^2(n-k-1)$. Use the formula sheet.

12. 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 $\boldsymbol{\ell}$ be a $k+1 \times 1$ vector of constants.

(a) What is the distribution of $\boldsymbol{\ell}'\mathbf{b}$? 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 $H_0 : \boldsymbol{\ell}'\boldsymbol{\beta} = c$. Write down a formula for the test statistic. A statistic is a function of the sample data that is *not* a function of any unknown parameters.

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

(g) For a regression model with four independent variables, suppose you wanted to test $H_0 : \beta_1 = \beta_2$. Give the vector $\boldsymbol{\ell}$.

(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 $\boldsymbol{\ell}'\boldsymbol{\beta}$. "Derive" means show the High School algebra.

13. 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 $SST = SSR + SSE$. Hint: Add and subtract $\widehat{y}_i$. Switch to matrix notation partway through the calculation.

14. For the general linear regression model with normal error terms, show that if the model has an intercept, $\mathbf{e}$ and $\overline{y}$ are independent. Here are some ingredients to start you out. For the model with intercept,

(a) What does $X'\mathbf{e} = \mathbf{0}$ tell you about $\sum_{i=1}^n e_i$?

(b) Therefore what do you know about $\sum_{i=1}^n y_i$ and $\sum_{i=1}^n \widehat{y}_i$?

(c) Show that the least squares plane must pass through the point $(\overline{x}_1, , \ldots \overline{x}_k, \overline{y})$. Start with a scalar expression for $\widehat{y}_i$.

(d) Now show that $\mathbf{e}$ and $\overline{y}$ are independent.

15. Continue assuming that the regression model has an intercept. Many statistical programs automatically provide an *overall* test that says none of the independent variables makes any difference. If you can't reject that, you're in trouble. 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. Check Problem 7.

16. 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.

17. 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.

18. 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 17 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.
