Stat705 PS 4. Due in class October 6th

NOTE 1: It is fine to work with other students on problem sets, and that is encouraged. Each person's write up must be done separately though, and identical answers (to relatively complicated problems) from different students will not be graded.

NOTE 2: If you use a source other than the textbook to do these problems, you must list the source. Additionally, it is never OK to copy from another source verbatim! If we notice that, the problem won't be graded.

Please read chapter 2 in Plane Answers...

1. Suppose $\mathbf{y} \sim N(\mathbf{X}_0\boldsymbol{\gamma}, \sigma^2\mathbf{I}_n)$. Let $C(\mathbf{X}_0) \subset C(\mathbf{X})$. Let $\mathbf{P}_{X_0}$ and $\mathbf{P}_X$ be the perpendicular projection matrices onto $C(\mathbf{X})$ and $C(\mathbf{X}_0)$ respectively. (It's OK to quote theorems or results without proof to show the following, but you may need to verify that the theorem's requirements are true.)

   (a) Find the distributions of $N = \mathbf{y}^{\mathrm{T}}(\mathbf{P}_X - \mathbf{P}_{X_0})\mathbf{y}$ and $D = \mathbf{y}^{\mathrm{T}}(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$.

   (b) Are N and D independent? Why or why not?

   (c) What is the distribution of $\frac{N/(rank(\mathbf{X})-rank(\mathbf{X}_0))}{D/(n-rank(\mathbf{X}))}$.

2. Suppose $\mathbf{y} \sim N(\mathbf{1}\beta_0 + \beta_1\mathbf{x}, \sigma^2\mathbf{I}_n)$.

   (a) What is an unbiased estimator of $Var(\widehat{\beta}_1)$?

   (b) Consider $t = \widehat{\beta}_1/\widehat{Var}(\widehat{\beta}_1)$ where the denominator is your estimator from the previous question. Are the numerator and denominator independent? Why or why not?

   (c) If $\beta_1 = 0$, what is the distribution of t?

3. Suppose you have responses $\mathbf{y}$ and covariates $\mathbf{x}_1$ and $\mathbf{x}_2$. Let $\mathbf{x}_j^* = \mathbf{x}_j - \mathbf{1}^{\mathrm{T}}\mathbf{x}_j/n$. Consider two models: $\mathbf{y} = \beta_0\mathbf{1}_n + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{y} = \gamma_0\mathbf{1}_n + \gamma_1\mathbf{x}_1^* + \gamma_2\mathbf{x}_2^* + \mathbf{e}*, \mathbf{e}^* \sim N(\mathbf{0}, \tau^2\mathbf{I}_n)$.

   (a) What is the relationship between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$?

   (b) Are the $\widehat{\mathbf{y}}$s the same for both models? Why or why not?

   (c) Are the residuals the same for both models? Why or why not?

   (d) What is the relationship between MSEs from both models (i.e $\widehat{\sigma}^2$ and $\widehat{\tau}^2$)?

4. Let $\mathbf{y} = \beta_0\mathbf{1}_n + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Let $\mathbf{P}_{\mathbf{x}_1,\mathbf{x}_2}, P_{\mathbf{x}_1}$, and $\mathbf{P}_{\mathbf{x}_2}$ be perpendicular projection matrices that project onto $C(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2)$, $C(\mathbf{1}, \mathbf{x}_1)$, and $C(\mathbf{1}, \mathbf{x}_2)$ respectively. Assume $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2)$ has full rank. In class we showed that $SSR(\mathbf{x}_2) = SSR(\mathbf{x}_2|\mathbf{x}_1)$ when $(\mathbf{P}_{x_2} - \frac{1}{n}\mathbf{J}) = (\mathbf{P}_{x_1,x_2} - \mathbf{P}_{x_1})$. Assume without loss of generality (see previous question) that $\mathbf{1}^{\mathrm{T}}\mathbf{x}_1 = \mathbf{1}^{\mathrm{T}}\mathbf{x}_2 = 0$. Write out the matrix elements in $\mathbf{P}_{x_2}, \frac{1}{n}\mathbf{J}, \mathbf{P}_{x_1,x_2}$, and $\mathbf{P}_{x_1}$ and show that $\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = 0$ implies $(\mathbf{P}_{x_2} - \frac{1}{n}\mathbf{J}) = (\mathbf{P}_{x_1,x_2} - \mathbf{P}_{x_1})$.