Biostatistics 140.653 Third Term, 2021 March 1, 2021

Quiz 2

The purpose of this quiz is to assess your knowledge of the course materials covered during the second two weeks of class and covered in Problem Set 2.

Instructions:

- This is an open book quiz; you may consult your course notes and handouts.
- You should not discuss this quiz with any other student during Monday March 1st.
- This quiz is designed to be completed in 20-30 minutes.
- You may provide your solution by editing the word version of this quiz, annotating the pdf version of this quiz or writing your solution on paper and submitting a picture of your solution.

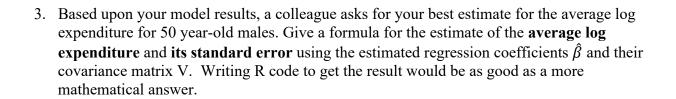
By signing my name, I enter agree to abide by the instructions above and the Johns Hopkins University School of Public Health Academic Code:

Name (Print):		
Signature:		

Suppose you use the Problem Set 2 NMES data for persons <u>aged 19 to 94</u> to address the question of whether males and females use roughly the same quantity of medical services at each age by regressing <u>log total expenditures</u> on a non-linear function of age, sex and the interactions. Specifically, you regress:

 $Y = LN(totalexp + 1) = log_e(totalexp + 1)$, on:

- Intercept
- age variables: age-40, $age sp1 = (age-40)^+$, $age sp2 = (age-65)^+$
- female= 1 if female, 0 if male
- the interaction of the three age variables with *female*
- 1. The coefficient for *female* estimates the difference in average log expenditures comparing:
 - a. 40 year-old males to females (i.e. males minus females)
 - b. 40 year-old females to males
 - c. 65 year-old males to females
 - d. 65 year-old females to males
- 2. To determine whether there is a difference in the average log expenditures between males and females at any age you would:
 - a. Fit a second model including only the *intercept* and perform the ANOVA F-test
 - b. Fit a second model excluding the 2 main effect terms of age_sp1 and age_sp2 as well as the 2 interaction terms of female with age_sp1 and age_sp2 and perform the ANOVA F-test
 - c. Fit a second model excluding the 3 interaction terms between the age variables and female and perform the ANOVA F-test
 - d. Fit a second model excluding the main effect for *female* and the 3 interaction terms between the age variables and perform the ANOVA F-test



4. This same colleague realizes that interpreting the average log expenditure is difficult and asks that you provide your best estimate for the <u>average expenditure for 50 year-old males</u> in dollars.

Recall that Y is log normal if $LN(Y) \sim N(\mu, \sigma^2)$.

If Y is log-normal, then
$$E(Y) = \exp(\mu + \sigma^2/2)$$
, median $(Y) = \exp(\mu)$.

Give a formula for the estimate of the <u>average expenditure for 50 year-old males</u> using the estimated regression coefficients and estimate of σ^2 , if necessary. You DO NOT have to derive the standard error for this mean.

5. In most cases, we want to also provide a 95% CI for means of interest. Describe, in words, an approach to generate a 95% confidence interval for the average expenditure for 50 year-old males.