In-Class Exercise - Inference with Multiple Linear Regression

Using 2018 GSS data, you estimate the following population model:

$$emailhr_i = \beta_0 + \beta_1 childs_i + \beta_2 age_i + \beta_3 hsgrad_i + \beta_4 somecol_i + \beta_5 college_i + \beta_6 postgrad_i + u_i$$

Where the variables are defined as follows:

- $emailhr$ - email hours per week
- $childs$ - number of children
- age - age in years
- $hsgrad$ - 1 if complete high school only, 0 otherwise
- $somecol$ - 1 if completed some college, 0 otherwise
- $college$ - 1 if completed college, 0 otherwise
- $postgrad$ - 1 if completed some postgraduate education, 0 otherwise.

Note that the education binary variables are mutually exclusive: a college graduate would have $hsgrad = somecol = postgrad = 0$ and $college = 1$.

```
. regress emailhr childs age hsgrad somecol college postgrad
```

Source	SS	df	MS	Number of obs	=	1,410
				F(6, 1403)	=	13.55
Model	10708.9133	6	1784.81888	Prob > F	=	0.0000
Residual	184794.92	1,403	131.714127	R-squared	=	0.0548
				Adj R-squared	=	0.0507
Total	195503.833	1,409	138.753608	Root MSE	=	11.477

emailhr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
childs	.0477113	.2083668	0.23	0.819	-.3610327	.4564554
age	-.0654925	.0192602	-3.40	0.001	-.1032743	-.0277106
hsgrad	1.970668	1.143132	1.72	0.085	-.2717642	4.2131
somecol	3.620887	1.159198	3.12	0.002	1.346939	5.894835
college	7.565025	1.220929	6.20	0.000	5.169982	9.960067
postgrad	7.322089	1.278366	5.73	0.000	4.814375	9.829804
_cons	6.127687	1.296195	4.73	0.000	3.584998	8.670375

You also estimate the following population model:

$$emailhr_i = \beta_0 + \beta_1 childs_i + \beta_2 age_i + u_i$$

```
. regress emailhr childs age
```

Source	SS	df	MS	Number of obs	=	1,410
				F(2, 1407)	=	4.74
Model	1309.67915	2	654.839575	Prob > F	=	0.0088
Residual	194194.154	1,407	138.02001	R-squared	=	0.0067
				Adj R-squared	=	0.0053
Total	195503.833	1,409	138.753608	Root MSE	=	11.748

emailhr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
childs	-.2399636	.2088007	-1.15	0.251	-.6495578	.1696306
age	-.0438551	.0194919	-2.25	0.025	-.0820915	-.0056187
_cons	9.732497	.9346207	10.41	0.000	7.899097	11.5659

1. In words, interpret the coefficient on $childs$ from the regression that includes binary education variables. Is it statistically significant at the 5% level? How do you know?

2. Suppose that you want to test the joint significance of the following regression: $emailhr_i = \beta_0 + \beta_1 childs_i + \beta_2 age_i + \beta_3 hsgrad_i + \beta_4 somecol_i + \beta_5 college_i + \beta_6 postgrad_i + u_i$. What is the F-test statistic?

3. Suppose you want to test whether the four binary education variables are jointly significant. What is the null hypothesis? What is the alternate hypothesis? Can you reject the null at the 5% level? What is the F-test statistic?

4. Suppose you want to test whether there is a statistically significant difference in hours spent on e-mail between people who have completed college and people who have completed some postgraduate education. Use the 5% significance level.

   You also conducted the following hypothesis tests:

```
. test hsgrad=somecol

 ( 1)  hsgrad - somecol = 0

       F(  1,  1403) =    4.03
            Prob > F =    0.0448

. test hsgrad=somecol=0

 ( 1)  hsgrad - somecol = 0
 ( 2)  hsgrad = 0

       F(  2,  1403) =    5.29
            Prob > F =    0.0051

. test college=postgrad

 ( 1)  college - postgrad = 0

       F(  1,  1403) =    0.05
            Prob > F =    0.8167

. test college=postgrad=0

 ( 1)  college - postgrad = 0
 ( 2)  college = 0

       F(  2,  1403) =   21.71
            Prob > F =    0.0000
```

   (a) Write your null hypothesis, alternative hypothesis, and test statistic.

   (b) What is the critical value for this test?

   (c) What is the F-test statistic?