Practice Final Exam
G572
Spring 2023

The actual final exam will be Wednesday, February 22nd during the regularly scheduled course time and you will have up to 80 minutes to complete it during that window. There will be 20 questions in the following format:
- 18 multiple choice questions
- 2 open answer questions (these will be along the lines of the "explain"/ "why" – type questions from problem sets).
  - None of the above will require Stata

The final exam will be cumulative. I am providing you with a selection of practice questions from the latter part of the course.

1. Suppose you regressed annual wage on the following variables, with the estimated coefficients in parentheses behind the variable: female (1,862), YearsofEducation (3,418). Assume the coefficients are significant at the 99% confidence level and (for the sake of this question) that the model is exogenous. These estimates would imply that an extra year of education
   a. Causes non-female income to increase by more than female income
   b. Causes female income to increase by more than non-female income
   c. **Causes female and non-female income to increase by the same amount**

Use the following image to answer questions 2-3:

. reg wage YearsEdu female

Source	SS	df	MS		Number of obs	=	199
					F(2, 196)	=	19033.07
Model	4.8943e+10	2	2.4472e+10		Prob > F	=	0.0000
Residual	252005628	196	1285743		R-squared	=	0.9949
					Adj R-squared	=	0.9948
Total	4.9195e+10	198	248461044		Root MSE	=	1133.9

wage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
YearsEdu	4617.528	24.57666	187.88	0.000	4569.06	4665.997
female	12190.49	163.649	74.49	0.000	11867.75	12513.23
_cons	-40726.08	399.0624	-102.05	0.000	-41513.08	-39939.07

2. Passively predict the average annual wage of someone who is female with 16 years of education.

-40,726.08 +12,190.49+4,617.53*16 = $45,344.89

3. Passively predict the difference in income of two people of the same gender, but one has 4 more years of education.

4*4,617.53 = $18,470.12

4. Which of the following brands has the lowest quantity sold, controlling for price:

. reg Q_shoes P_shoes i.Brand

Source	SS	df	MS		Number of obs	=	200
					F(4, 195)	=	282.32
Model	28502.0671	4	7125.51677		Prob > F	=	0.0000
Residual	4921.55237	195	25.2387301		R-squared	=	0.8528
					Adj R-squared	=	0.8497
Total	33423.6194	199	167.957887		Root MSE	=	5.0238

Q_shoes	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
P_shoes	-.272678	.0083508	-32.65	0.000	-.2891474	-.2562087
Brand						
2	.1135662	1.011078	0.11	0.911	-1.880486	2.107618
3	2.744283	.9757593	2.81	0.005	.8198869	4.66868
4	1.993289	1.032385	1.93	0.055	-.0427849	4.029363
_cons	2758.786	1.216468	2267.87	0.000	2756.387	2761.185

a. 1
b. 2
c. 3
d. 4

5. A good proxy variable must be correlated with:

a. Other exogenous X variables in the model
b. The omitted variable that it is proxying for
c. The variable of interest
d. All of the above
e. None of the above

6. If two confounding factors are positively correlated with and X in our model, we will have overestimated the true causal impact of X on Y.
   a. True
   b. False
   c. Not enough information to tell

7. Which of the following models may have a variable of interest:
   a. Model to be used for passive prediction
   b. Model to be used for active prediction

8. The model with the highest R-squared is always the best model to use for passive prediction.
   a. True
   b. False

9. Any variable that would have been in U would be a good control variable.
   a. True
   b. False

10. Which of the following models would Amazon be most likely to add control variables to?
    a. A model of warehouse storage space as a function of Prime subscription, given that they are considering changing their Prime subscription cost.
    b. A model of warehouse storage space as a function of weather and time of year.
    c. Both of the above
    d. None of the above