Final Exam
EC295 – Winter 2021
Due: Monday, April 19 at 6:00 PM

Structure

- This is a take-home test over 8 hours.
- You will answer a total of five questions, each containing one to eight sub-questions. Make sure to answer all sub-questions. **You must show your work to get the full grade**.
- This is open book and open internet.
- You are prohibited from collaborating in any way with other people on the exam. This means no discussing the questions in person, by phone, using chat services, question boards, email, etc.
- If you draw your answer from another source (e.g. the textbook, the internet), you must reference it. The normal rules of plagiarism apply to this exam.
- Instructor help on the test will be limited to clarification questions only. I will be available via zoom from 10 am to 10.20 am, from 1 pm to 1.20 pm. Please use the corresponding zoom link. You can also contact me via email; I will answer within an hour.
- I have attached the t-distribution and F-distribution tables (at the end of this document). You can use them to calculate critical values.

Submission Instructions

- Submit your exam to Gradescope when complete. Note that you can submit as many times as you wish before the deadline. **Late submissions will receive a zero grade.**
- You are required to **hand-write** your responses to each question on paper (i.e. do not use a word processor unless you have accommodations through ALC).
- I would suggest answering each sub-question on a separate piece of paper.
- Sign your name below the response to each question.
- Upload your hand-written responses to Gradescope in one of two ways:
  - Take a **legible** photo of your response to each question separately, and then upload each image. **Make sure to tag your questions**.
  - Scan your questions into a single PDF, then upload it and tag each question. **Make sure to tag your questions**.

Question 1 - 24 points

For each statement below, evaluate whether it is true, false, or uncertain, and fully explain your answer.

a) (5 points) In a multiple regression, when control variables are correlated with the unobserved error term, the OLS slope estimate on the variable of interest is always biased.

b) (5 points) A high (but not perfect) amount of collinearity between independent variables in a regression model will result in higher standard errors of the OLS slope estimates.

c) (5 points) If the variables x_1 and x_2 are uncorrelated with each other, the multiple regression of y on x_1 and x_2 will always produce different slope estimates from separate simple regressions of y on x_1 and y on x_2.

d) (4 points) Suppose you estimate the parameters of the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Based on an F-test, you reject the null hypothesis that $\beta_1 = 0$ and $\beta_2 = 0$. The appropriate conclusion is that both $\beta_1 \neq 0$ and $\beta_2 \neq 0$.

e) (5 points) The intuition behind the homoskedasticity-only F-test is that if we impose restrictions on the model, and as a result the fit improves, we accept the restrictions.

Question 2 - 8 points

Consider the regression model relating crop yield to two types of fertilizer:

$$\text{yield} = \beta_0 + \beta_1 \text{fertilizer}_1 + \beta_2 \text{fertilizer}_2 + u$$

Use "approach 2" to transform the regression so that you can use a t-statistic to test the hypothesis that fertilizer 1 has three times the effect fertilizer 2 has on yields. Define the null and alternative hypotheses, as well as ALL the steps you would follow to perform the test.

Question 3 - 12 points

Suppose you are given the following sample regression function:

$$\ln(\widehat{\text{hwage}}) = 1.5 + 0.1\text{exper} - 0.2 \text{ female} - 0.1 \text{ immigrant} - 0.05 \text{ (female*immigrant)}$$

where ln(hwage) is the natural logarithm of hourly wages, exper is years of experience, female is a variable equal to 1 if the person is female and 0 if the person is male, and immigrant is a variable equal to 1 if the person is immigrant and 0 if it is non-immigrant. The sample size is 85.

a) (6 points) Precisely interpret the estimate on female, the estimate on immigrant, and the interaction term between female and immigrant.

b) (6 points) Suppose you wanted to test whether the regression function that relates hourly wages with experience and gender is the same for immigrants and non-immigrants. Explain the null and alternative hypotheses you would use for the test, as well as the steps you would follow to perform the test. Make sure to find the corresponding critical value when possible.

Question 4 - 36 points

Suppose you are given the following sample regression function:

$$\text{price} = \beta_0 + \beta_{score}\text{Score} + \beta_{France}\text{France} + \beta_{Australia}\text{Australia} + u \quad (1)$$

where price is the wine price measured in dollars, score is the score given by a wine expert (scale 1 to 100 points), Europe and Australia are dummy variables indicating the wine's country of origin. The reference category is United States (USA). The sample size is 50.

a) (5 points) Suppose that all of the wines in the sample came from France and USA. What problem, if any, would this cause for estimating regression (1) using OLS?

You estimate regression (1) and obtain the following estimates. The standard errors are shown in parenthesis below the estimate.

$$\widehat{price} = \underset{(0.10)}{5} + \underset{(0.045)}{0.1}\text{ Score} + \underset{(0.70)}{1.2}\text{ France} - \underset{(0.30)}{0.5}\text{ Australia}$$

b) (5 points) Precisely interpret the estimate on the intercept, and the slope estimates on the dummy variables France and Australia.

c) (4 points) Using an F-test, test the null hypothesis that the effect of score on wine prices is zero against the alternative that it is not at the 5% significance level.

d) (4 points) Construct a 95% confidence interval for the effect of a 10-point increase in scores on wine prices.

e) (5 points) Jointly test the null hypothesis that $\beta_{Australia} = 0$ and $\beta_{France} = 0$, at a 5% significance level. Assume the t-statistics for testing each restriction individually have a correlation of 0.5.

f) (5 points) Suppose you instead wanted to use France as the reference category. Write down the new regression model and find the estimates for the intercept and all slope coefficients.

g) (4 points) Suppose you decide that oak ageing matters for determining the price of a wine, so you add it to the regression model to get

$$\text{price} = \beta_0 + \beta_{score}\text{Score} + \beta_{France}\text{France} + \beta_{Australia}\text{Australia} + \beta_{oak}\text{oak} + u \qquad (2)$$

where oak represents the number of years of oak ageing. Would the R^2 and the adjusted R^2 in regression (2) be higher or lower than the one without oak (i.e., regression (1))? Explain.

h) (4 points) Suppose you wanted to extend regression (1) to let the effect of scores on wine price to depend on the level of score. Write down the regression model that accounts for this effect.

Question 5 - 20 points

Consider the following Stata output. The dependent variable in all regressions is baby birthweight in kilograms. The independent variables are as follows: ln(income) is the natural logarithm of the family income, age is mother's age in years, age^2 is mother's age squared, and smoker=1 if the mother smoked while pregnant and 0 otherwise. The variable of interest is income. The regression model errors are homoskedastic.

Variables	(1)	(2)	(3)
ln(income)	0.317***	0.116***	0.115***
	(0.0273)	(0.0251)	(0.0251)
age		-0.0862***	-0.0862***
		(0.00953)	(0.00953)
age^2		-0.0003*	-0.0003*
		(0.000174)	(0.000174)
smoker			-0.113***
			(0.00956)
Constant	2.265***	5.392***	5.431***
	(0.0828)	(0.144)	(0.144)
Observations	76,072	76,072	76,072
R-squared	0.002	0.161	0.162

Standard errors in parentheses
*** p< 0.01, ** p< 0.05, * p< 0.1

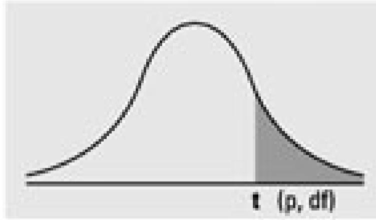
a) (5 points) Interpret the coefficient on ln(income) in regression (1), and explain why it is larger in absolute terms than in regressions (2) and (3).

b) (5 points) Precisely explain what is the effect of age on birthweight. Use the results from regression (2).

c) (5 points) Test, at a 5% significance level, the null hypothesis that the effect of age on birthweight is linear.

d) (5 points) In regression (3), for the OLS estimate of $\beta_{\ln(\text{income})}$ to be unbiased, you can make one of two assumptions: 1) $E(u|\ln(\text{income}), \text{age}, \text{age}^2, \text{smoker}) = 0$ (mean independence of the error), or 2) $E(u|\ln(\text{income}), \text{age}, \text{age}^2, \text{smoker}) = E(u|\text{age}, \text{age}^2, \text{smoker})$(conditional mean independence of the error). Explain the key differences between these two assumptions in this context.

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	—	—	80%	90%	95%	98%	99%	99.9%

Percentage points of *Fisher*'s distribution

$f_{0.05,v_1,v_2}$

v_2 \ v_1	Degrees of freedom of the numerator (v_1)																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.64	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.31	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Degrees of freedom of the denominator (v_2)