Econ 3040 – Final Exam, Dec. 20th, 2020

Professor: Ryan Godwin

- The exam consists of 100 marks. Answer all questions.
- You must submit your exam by 3:45 pm on December 20th in the UM Learn "Final Exam" Dropbox, under the assignment tab. If you are unable to submit on UM Learn, email me your exam directly.
- You do not need to use R Studio for the final exam, but you can. You should at least have a calculator available.
- The exam is open book / open notes, however, you must provide answers to the questions **in your own words**. You may not plagiarize, or copy-paste, responses to questions from the internet, or from anywhere.
- You must complete the exam individually. Please be aware that there are ways to detect cheating, even electronically.
- A table of standard normal probabilities is provided at the back of the exam.

Part A – Short Answer – 8 marks each, except Q6

1) Why are estimators random? Why are the least squares estimators, $b_0, b_1, \ldots, b_k$, considered to be "good" estimators for $\beta_0, \beta_1, \ldots, \beta_k$?

2) Consider the population model:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Suppose that the sample correlation between X and Y is exactly 0. What is the estimated value b_1? Explain your answer carefully.

Hints: the formula for calculating b_1 is

$$b_1 = \frac{\sum_{i=1}^{n}[(X_i - \bar{X})(Y_i - \bar{Y})]}{\sum_{i=1}^{n}[(X_i - \bar{X})^2]}$$

and the sample covariance between X and Y is

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

3) Suppose that your data is: *wage* – the hourly wage of a worker; *male* – a dummy variable; and *educ* – a categorical variable which equals "high" if the worker has obtained a high-school degree, "ugrad" if a undergraduate university degree, "MA" if a Master's degree, and "PhD" if a PhD. Explain how this categorical variable would be used in an analysis of the effects of education on wage.

4) Why do you reject a null hypothesis when the p-value is small?

5) Explain why it is important to use adjusted-R-square ($\bar{R}^2$) instead of R-square (R^2) in a multiple regression model. How does $\bar{R}^2$ fix the problem that occurs with R^2?

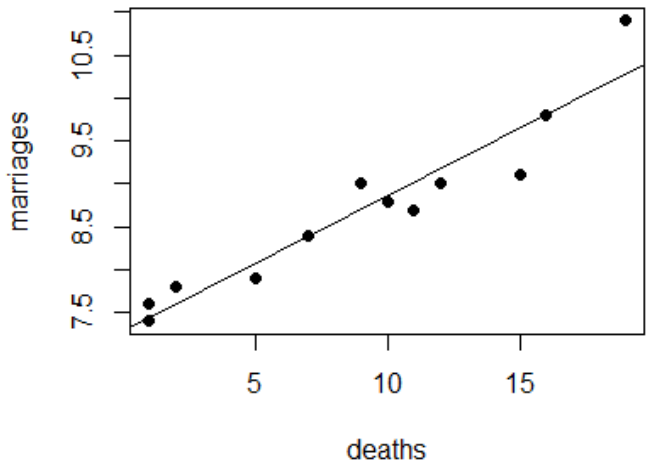
6) [10 marks] The following question uses data from the U.S. The two variables are *marriages* – the number of marriages per 1000 people in Kentucky, and *deaths* – the number of people who drowned after falling out of a fishing boat in the U.S. The data was collected annually from the year 1999 to 2010.

	Year											
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
marriages	10.9	9.8	9	9	9.1	8.8	8.7	8.4	7.8	7.9	7.6	7.4
deaths	19	16	9	12	15	10	11	7	2	5	1	1

The model estimated by OLS, is:

$$\widehat{marriages} = 7.28 + 0.16 \times deaths \quad , \quad R^2 = 0.907$$
$$(0.17) \quad (0.02)$$

Below is a plot of the data, where the line indicates the LS fitted model:



Test the null hypothesis that *deaths* has **no effect** on marriages. Using the result of this test, the above figure, and the R^2, comment on whether drowning deaths on fishing boats causes the number of marriages in Kentucky.

Part B – Long Answer

7) Consider the population model:

$$Wage = \beta_0 + \beta_1 Education + \epsilon$$

where *Wage* and *Education* are the hourly wage, and the number of years of education of a worker, respectively. The model, estimated by least squares, is:

```
summary(lm(Wage ~ Education))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    20.42       3.23   52.13   <2e-16 ***
Education      12.63       0.79   ?????   ?????? ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) [2.5 marks] Fill in the missing t-statistic and p-value.

b) [2.5 marks] Test the hypothesis that *Education* has no effect on *Wage*.

Now, consider the population model:

$$Wage = \beta_0 + \beta_1 Education + \beta_2 IQ + \epsilon$$

where *IQ* is the intelligence quotient score of the worker. The model, estimated by least squares, is:

```
summary(lm(Wage ~ Education + IQ))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.23       5.00    1.65   0.0990 .
Education        2.96       3.38   ?????   ??????
IQ               5.09       0.23   21.68  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) [5 marks] Compared to the first model, do you think the R^2 for this model will be larger or smaller? Explain.

d) [2.5 marks] Calculate the missing t-statistic and p-value.

e) [2.5 marks] Test the hypothesis that *Education* has no effect on *Wage*.

f) [5 marks] Why has the estimated value for β_1 changed, compared to the first model? Why do these two estimated models have different implications about the effect of *Education* on *Wage*?

g) [5 marks] Try to tell a story about why the inclusion of *IQ* makes the *Education* variable insignificant in the second model compared to the first (think about the Fireplaces example from chapter 6).

You will need the following table for question 8:

Table 7.1: χ^2 critical values for the F-test statistic.

q	5% critical value
1	3.84
2	3.00
3	2.60
4	2.37
5	2.21

8) [5 marks each] The following question uses data from a version of the CPS dataset. The dependent variable is *ahe* – the hourly earnings of a worker. *female* is a dummy variable indicating gender; *age* is self explanatory; *yrseduc* is the number of years of education of the worker. In addition, there is a *location* variable which indicates the region in which the worker lives: *northeast, south, west,* or *midwest* (the "base" group). The sample size is $n = 61395$.

	Model number				
Regressor	1	2	3	4	5
female	-4.17* (0.07)	-4.18* (0.07)	-4.19* (0.07)	-4.24* (0.25)	-4.18* (0.07)
age	1.89* (0.65)	1.79* (0.65)	0.15* (0.00)	0.16* (0.00)	0.98* (0.02)
age^2	-0.03 (0.05)	-0.03 (0.03)			-0.01* (0.00)
age^3	0.00 (0.00)	0.00 (0.00)			
age^4	-0.00 (0.00)	-0.00 (0.00)			
yrseduc	0.40* (0.10)	0.38* (0.10)	0.39* (0.10)	1.74* (0.02)	0.39* (0.10)
$yrseduc^2$	0.05* (0.00)	0.05* (0.00)	0.05* (0.00)		0.05* (0.00)
northeast	1.21* (0.11)		1.24* (0.11)	1.29* (0.11)	1.21* (0.11)
south	-0.02 (0.10)		-0.02 (0.10)	0.04 (0.10)	-0.01 (0.10)
west	0.76* (0.40)		0.74* (0.10)	0.81* (0.10)	0.76* (0.10)
intercept	-28.30* (6.27)	26.75* (6.27)	-1.22 (0.73)	-10.37* (0.25)	-17.03 (0.86)
R^2	0.2676	0.2651	0.2536	0.2514	0.2672
$\bar{R}^2$	0.2675	0.2650	0.2535	?????	0.2671

*significant at the 1% level.

a) How much more do workers in the *northeast* make compared to those in the *west*?

b) Using model (1) as the unrestricted model, test the hypothesis that *age* has a linear effect on *ahe*.

c) Test the hypothesis that *location* has no effect on *ahe*.

d) Using model (5), interpret the effect of *yrseduc* on *ahe*. Be sure to illustrate how this effect depends on the value of *yrseduc* itself.

e) Calculate the missing value for adjusted R-square.

END

Table 3.2: Area under the standard normal curve, to the right of z.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002