Econ 3040 Final Exam

Ryan T. Godwin

The exam is 120 minutes long, and consists of 120 marks (**1 mark per minute**). **There are 8 questions**. The number of marks allocated to each question is in [red]. For example, if a questions is worth 3 marks, a [3] will appear at the beginning of the question. You have an extra 20 minutes to upload your answers to the UM Learn dropbox. You may quickly submit a low-quality version of your exam, and then upload a higher quality version after the 20 minutes, as long as there are no substantial differences between the versions. There is a table of critical values for the F-statistic, and a table of standard Normal probabilities, at the end of the exam. **Do not collaborate with anyone on this exam.**

1. [**10 marks total**] The probability function for random variable X is:

Pr(X)	X
0.2	1
0.5	2
0.3	3

   a) [3] What is $E[X]$?

   b) [3] What is $var[X]$?

   c) [4] The probability function for random variable Z is:

Pr(Z)	Z
0.2	2
0.5	4
0.3	6

   Notice the relationship between X and Z. Doing as little work as possible, determine $E[Z]$ and $var[Z]$.

2. [5] Explain the concept of *unbiasedness*, as it applies to the statistical properties of an estimator.

3. [5] Suppose that you have a categorical variable in your data set, called *education*. It measures the education level of the individual. It takes on the values "highschool", "university", "Masters" and "PhD". Explain how this *education* variable could be used (as regressors) in a model that is estimated by OLS.

4. [**20 marks total**] Consider the population model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

   a) [5] Explain the interpretation of β_1 and β_2.

   b) [5] What makes the LS estimator, b_1, best?

   c) [5] Is it better to have a high variance for X_1, or a low variance? Explain.

   d) [5] Could the model get "worse" if we added some other variable (X_3)? Explain using R^2 and $\bar{R}^2$.

5. [**20 marks total**] The estimated model is:

$$\begin{aligned} \hat{wage} =& 11.00 - 2.34 \times female \\ & (0.30) \quad (0.44) \end{aligned}$$

$wage$ is the workers hourly wage, measured in dollars. $female$ is a dummy variable taking on the value 1 if the individual is female, and 0 if the individual is male.

a) [5] Do men and women have the same wages? Answer using a hypothesis test.

b) [5] What is the sample average wage for women?

c) [5] Using a 95% confidence interval, test the null hypothesis that men and women earn the same wages.

d) [5] Suppose instead that the dummy variable is defined in the opposite way, and the following model is estimated:

$$wage = \beta_0 + \beta_1 male + \epsilon$$

What will be the estimated value for β_1?

6. [**2 bonus**] Explain what heteroskedasticity is, and how it creates a problem in OLS estimation.

SEE NEXT PAGE

7. [27 marks total] This question uses the video game data from Lab 1 / Assignment 1. The population model is:

$$Sales = \beta_0 + \beta_1 Score + \beta_2 Score^2 + \beta_3 Nintendo + \beta_4 GenreParty + \beta_5 GenrePuzzle + \beta_6 GenreSports + \beta_7 GenreStrategy + \beta_8 RatedE + \beta_9 newgame + \epsilon$$

The variables in the model are:

- Sales - sales of the video game in millions of dollars
- Score - an average critic score of the game, taking values 0 to 10, with 10 being highest
- Nintendo - dummy variable: 1 if a Nintendo game, 0 otherwise
- GenreParty, GenrePuzzle, GenreSports, GenreStrategy - dummy variables indicating the genre of the video game
- RatedE - dummy indicating whether the video game is rated "E" for "everyone"
- newgame - dummy indicating whether the game was published after 2008

The R estimation results are provided below:

```
Call:
lm(formula = Sales ~ Score + Score2 + Nintendo + GenreParty +
    GenrePuzzle + GenreSports + GenreStrategy + RatedE + newgame,
    data = mydata)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.42732    0.55938   7.915 3.07e-15 ***
Score         -1.74690    0.16950 -10.306  < 2e-16 ***
Score2         0.16717    0.01263       ?        ?     ?
Nintendo       2.22960    0.13589  16.407  < 2e-16 ***
GenreParty     1.34413    0.67921   1.979   0.0479 *
GenrePuzzle   -1.16689    0.22983  -5.077 3.98e-07 ***
GenreSports   -0.14913    0.12563  -1.187   0.2352
GenreStrategy -0.78873    0.18073  -4.364 1.30e-05 ***
RatedE         0.36815    0.09162   4.018 5.96e-05 ***
newgame        0.08360    0.07564   1.105   0.2691
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.51 on 4696 degrees of freedom
Multiple R-squared:  0.1736,    Adjusted R-squared:  0.172
F-statistic: 109.6 on 9 and 4696 DF,  p-value: < 2.2e-16
```

---

a) [5] Does "Score" have a non-linear effect on "Sales"? Explain.

b) [3] Using one or two sentences, explain what it means for one variable to have a non-linear effect on another.

c) [2 bonus] What is the derivative of "Sales" with respect to "Score"?

d) [2] Which variables are insignificant?

e) [2] What is the value of the F-test statistic, for the null hypothesis that none of the "x" variables have any effect on "Sales"?

f) [2] Interpret the value of $\bar{R}^2$.

g) [4] Predict the sales for a Nintendo game, in the "Party" genre, rated "E" for everyone, and that receives a critic score of 9.

h) [9] What is the effect of "Score" on "Sales"? (This is a polynomial regression model. Recall that there is a special method for interpreting marginal effects.)

8. [**33 marks total**] Use some of the following 6 estimated models:

Table 1: Estimation results for question 8

	Dependent variable:					
	log(wage)					
	(1)	(2)	(3)	(4)	(5)	(6)
education	0.047***	0.056***	0.046***	0.058***	0.047***	0.044***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)
experience	0.014***	0.016***	0.014***	0.015***	0.015***	0.014***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
age	0.019***	0.019***	0.020***	0.020***	0.019***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
female	−0.259**	0.082***	−0.275**		−0.189	−0.284**
	(0.125)	(0.030)	(0.121)		(0.120)	(0.125)
Manitoba	−0.099***	−0.102***	−0.064***	−0.066***	−0.103***	
	(0.032)	(0.032)	(0.022)	(0.022)	(0.031)	
Saskatchewan	0.129***	0.131***	0.103***	0.101***	0.127***	
	(0.030)	(0.030)	(0.021)	(0.022)	(0.030)	
female × education	0.019**		0.020***		0.018**	0.021**
	(0.008)		(0.008)		(0.008)	(0.008)
female × experience	0.002*		0.003**			0.003**
	(0.001)		(0.001)			(0.001)
female × Manitoba	0.066	0.065			0.068	
	(0.044)	(0.044)			(0.044)	
female × Saskatchewan	−0.052	−0.061			−0.053	
	(0.043)	(0.043)			(0.043)	
Constant	1.716***	1.555***	1.724***	1.564***	1.685***	1.775***
	(0.087)	(0.065)	(0.086)	(0.065)	(0.086)	(0.087)
Observations	1,000	1,000	1,000	1,000	1,000	1,000
R^2	0.765	0.762	0.763	0.756	0.764	0.749
Adjusted R^2	0.762	0.760	0.761	0.754	0.762	0.747
Note:						*p<0.1; **p<0.05; ***p<0.01

The sample size is 1000. The variables in the data are:

- wage - yearly wage of the worker, measured in thousands of dollars
- experience - years of work experience
- age - the age of the worker in years
- female - a dummy variable indicating gender
- Manitoba - a dummy variable equal to 1 if the worker lives in Manitoba, 0 otherwise
- Saskatchewan - a dummy variable equal to 1 if the worker lives in Saskatchewan, 0 otherwise

---

Use a 5% significance level for all questions.

a) [2] Which of the six models do you think is "best", and why?

b) [3] Why is it a good idea to use "log(wage)" as the dependent variable, instead of just "wage"?

c) [4] Do the wages of workers depend on province (location)? Use a hypothesis test.

d) [4] Is there a difference in the wages of males and females? Use a hypothesis test.

e) [4]Does the effect of gender on wages depend on province (location)? Use a hypothesis test.

f) [4] How much more do workers in Saskatchewan make, compared to workers in Manitoba? Use model (1) to answer this question, and be careful - this is a tricky question.

g) [3] Is there a different effect of education on wages for men vs. women?

h) [9] Suppose that there is an unobservable "ability" variable (for example, the IQ score of the worker). Suppose that this variable partly determines both "education" and "wage". What will be the problem with estimating the effect that "education" has on "wage"?

END

Table 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

Table 3: 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