Econ 3040 – Final Exam, Apr. 18th, 2019

Professor: Ryan Godwin

You may use a calculator. Answer all questions in the answer book provided. The exam is 3 hours long and consists of 100 marks.

A formula sheet, and a table of probabilities from the standard Normal distribution, are provided at the back of the exam booklet.

YOU MUST SUBMIT ALL EXAM MATERIAL AT THE END OF THE EXAM.

DO NOT OPEN THE EXAM UNTIL YOU ARE INSTRUCTED TO DO SO.

NAME:	
STUDENT #:	

Part A – Multiple Choice – each question is worth <u>1 mark</u>

1) A large p-value implies
a. rejection of the null hypothesis.
b. a large test statistic.
c. a large estimate.
d. that the estimated values are consistent with the null hypothesis.

2) The formula for the OLS estimator, b_1, when adding a second regressor to the model,
a. stays the same.
b. changes, unless the second regressor is a dummy variable.
c. changes, unless the second variable is uncorrelated with the first variable.
d. changes.

3) Under imperfect multicollinearity
a. the OLS estimator cannot be computed.
b. two or more of the regressors are highly correlated.
c. the OLS estimator is biased even in samples of $n > 100$.
d. the error terms are highly, but not perfectly, correlated.

4) A type I error is
a. always the same as (1-type II) error.
b. the error you make when rejecting the null hypothesis when it is true.
c. the error you make when rejecting the alternative hypothesis when it is true.
d. always 5%.

5) In the multiple regression model, the least squares estimators are derived by
a. minimizing the RSS.
b. minimizing the ESS.
c. maximizing $\bar{R}^2$.
d. minimizing the sum of the squared horizontal distances between the regression line and the data points.

6) If you want to run a simple OLS regression of *Y* on *X* in *R*, you should type:
a. `regress(Y ~ X)`
b. `lm(Y ~ X)`
c. `ols(Y ~ X)`
d. `Y ~ beta0 + beta1*X`

7) The interpretation of the slope coefficient in the model: $\log(Y_i) = \beta_0 + \beta_1 X_i + \epsilon_i$, is as follows:
a. a 1% change in *X* is associated with a β_1 % change in *Y*.
b. a 1% change in *X* is associated with a change in *Y* of 0.01 β_1.
c. a change in *X* by one unit is associated with a 100 β_1 % change in *Y*.
d. a change in *X* by one unit is associated with a β_1 change in *Y*.

8) The OLS residuals
a. can be calculated using the errors from the regression function.
b. can be calculated by subtracting the fitted values from the actual values.
c. are unknown since we do not know the population regression function.
d. should not be used in practice since they indicate that your regression does not run through all your observations.

9) A type II error is
a. the error you make when not rejecting the null hypothesis when it is false.
b. typically smaller than the type I error.
c. the error you make when rejecting the alternative hypothesis when it is true.
d. the error you make when choosing type I error.

10) To standardize a variable you
a. subtract its mean and divide by its standard deviation.
b. allow for non-linear effects in the regression model.
c. account for heteroskedasticity.
d. add and subtract 1.96 times the standard deviation to the variable.

Part B - Short Answer – <u>5 marks</u> each

1) Describe why you would use an *F*-test instead of a *t*-test, when you are testing multiple restrictions.

2) Explain why it is dangerous to assume that the random errors (ϵ_i) are homoskedastic, when they might actually be heteroskedastic.

3) Give a common example of how the assumption "A.2: no perfect multicollinearity" can be violated, and explain the consequence of perfect multicollinearity.

4) Use the following data for this question:

$$Y = \{7, 8, 15\} \qquad X = \{2, 4, 6\}$$

The population model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

What are the OLS estimates for β_0 and β_1?

5) Using your answer to question (4) above, what are the OLS residuals?

6) Explain the concepts of unbiasedness, efficiency, and consistency, as they relate to the properties of an estimator.

7) Explain why it is important to use adjusted-R-square ($\bar{R}^2$) instead of R-square (R^2) in a multiple regression model.

8) Suppose that you obtain the following output from an OLS regression in *R*:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.05260    0.10561   0.498    0.620
x1          -0.04139    0.10713  -0.386    0.700
x2           0.10142    0.10928   0.928    0.356
x3          -0.02244    0.10472  -0.214    0.831
x4          -0.13196    0.12051  -1.095    0.276

Residual standard error: 1.04 on 95 degrees of freedom
Multiple R-squared:  0.02083, Adjusted R-squared:  -0.0204
F-statistic: 0.5052 on 4 and 95 DF,  p-value: 0.732
```

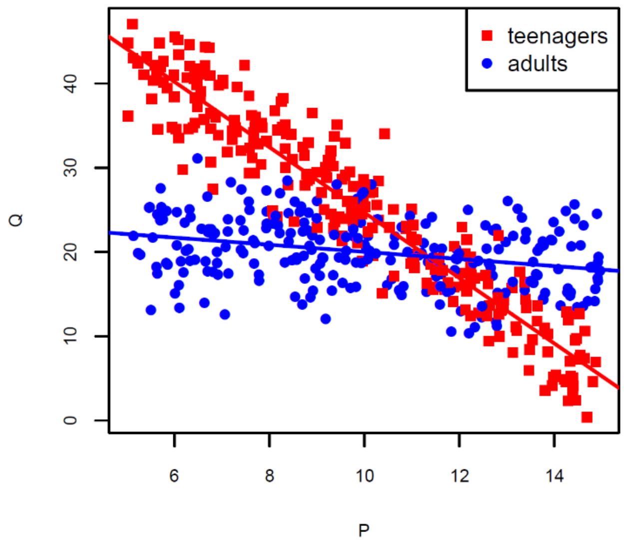
Test the null hypothesis that all βs (except the intercept) are equal to zero.

9) In the polynomial regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \cdots + \beta_r X_1^r + \epsilon,$$

explain how you could determine the appropriate *r* (the highest power in X_1 needed, in order to capture the non-linear effect).

10) This question uses the (hypothetical) demand for marijuana data:



The variables in the data are:

Q – quantity of marijuana consumed by the individual (grams / month)

P – the average price per / gram in the individual's location

adult – a dummy variable equal to 1 if individual is an adult, equal to 0 if the individual is a teenager

In the OLS regression:

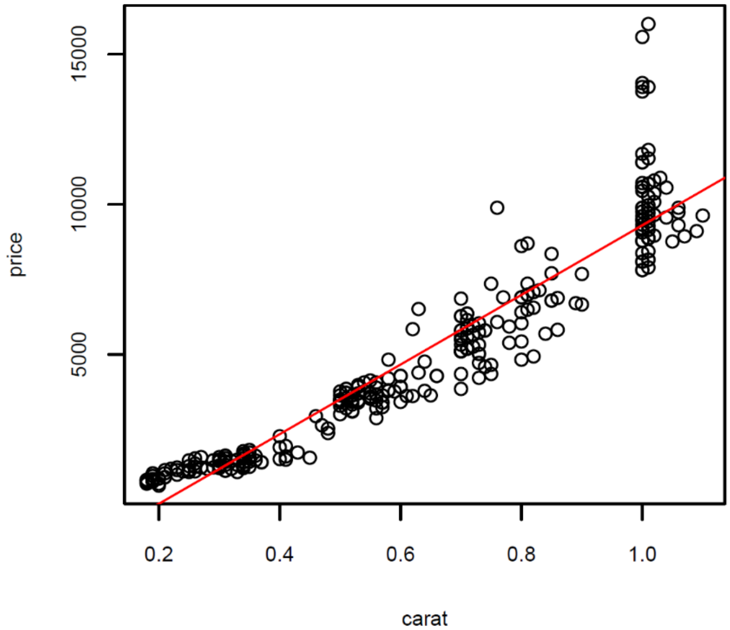
```
summary(lm(Q ~ P + adult + adult_P))
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  63.48944    0.85166   74.55   <2e-16 ***
P            -3.88168    0.08339  -46.55   <2e-16 ***
adult       -39.25222    1.21030  -32.43   <2e-16 ***
adult_P       3.45993    0.11695   29.58   <2e-16 ***
```

interpret the estimated value of `3.45993`.

Part C – Long Answer – each part is worth 4 marks

11) Below is a plot of the diamond data (size of the diamond in *carats* vs. the *price* of the diamond):



Included in the plot is the estimated equation:

$$\widehat{price} = -2298.4 + 11598.9 \times carat$$

a) It was discussed in class that the relationship between *carat* and *price* might be non-linear. What are the consequences of ignoring the nonlinear relationship above?

b) Instead of the linear model, a log-log model is estimated:

```
summary(lm(log(price) ~ log(carat)))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.12775    0.01440  633.99   <2e-16 ***
log(carat)   1.53726    0.01854   82.92   <2e-16 ***
```

What is the interpretation of the estimated coefficient (the estimated β) on *log(carat)*?

c) Describe two other ways in which you could capture the nonlinear effect of *carat* on *price* in a regression model.

12) This question uses a variant of the Current Population Survey (CPS) data. The variables in the data set are:

AHE – average hourly earnings, in $/week.

bachelor – a dummy variable equal to 1 if the worker has a university degree, or equal to 0 if the worker has a high school degree.

female – a dummy variable equal to 1 if worker is female, 0 otherwise.

age – the age, in years, of the worker.

The sample size is n = 7986.

You may need the following table for this question:

Critical values for *F*-statistics in large samples:

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

Model number:	1	2	3	4	5	6	7
Regressor	***AHE***	***AHE***	***log(AHE)***	***log(AHE)***	***log(AHE)***	***log(AHE)***	***log(AHE)***
age	0.439** (0.031)	2.068** (0.716)	0.024** (0.002)	0.147** (0.042)	0.146** (0.042)	0.191** (0.054)	0.160* (0.064)
age^2		-0.028* (0.012)		-0.002** (0.001)	-0.002** (0.001)	-0.003** (0.001)	-0.002* (0.001)
female×age						-0.097 (0.084)	-0.123 (0.085)
$female \times age^2$						0.002 (0.001)	0.002 (0.001)
bachelor×age							0.091 (0.084)
$bachelor \times age^2$							-0.001 (0.001)
female	-3.158** (0.180)	-3.149** (0.180)	-0.181** (0.011)	-0.180** (0.011)	-0.210** (0.014)	1.358 (1.238)	1.764 (1.251)
bachelor	6.865** (0.178)	6.863** (0.178)	0.405** (0.010)	0.405** (0.010)	0.378** (0.014)	0.377** (0.014)	-1.186 (1.236)
female×bachelor					0.064** (0.021)	0.063** (0.021)	0.066** (0.021)
intercept	1.884* (0.920)	-22.006* (10.532)	1.857** (0.054)	0.059 (0.611)	0.078 (0.610)	-0.633 (0.799)	-0.095 (0.939)
R^2	0.1900	0.1905	0.1924	0.1933	0.1942	0.1950	0.1968
$\bar{R}^2$	0.1897	0.1901	0.1921	0.1929	0.1937	0.1943	0.1959

Significance at the *5% and **1% significance level.

- **(a)** Using model (1), construct a 95% confidence interval around the estimated coefficient *age*.
- **(b)** Using model (2) as the unrestricted model, determine whether *age* has a linear or non-linear effect on *AHE*.
- **(c)** Using model (2), determine the effect of an additional year of *age* on *AHE*, when the worker is 20 years old, and when the worker is 60 years old.

The remaining questions refer to the log-linear models (models 3 – 7).

- **(d)** Explain the interpretation of the coefficient (the β) on *female×bachelor*.
- **(e)** Using any relevant models, determine if there is a different effect of education on wages, for men and for women.
- **(f)** Determine if there is a different effect of *age* on *AHE*, for men and for women.
- **(g)** Determine if *age* has a different effect on *AHE*, for individuals with a *bachelor* degree and for individuals without a *bachelor* degree.

Econ 3040 - Final Formula Sheet

expected value of Y (mean of Y)	μ_Y
variance of Y	$\sigma_Y^2 = E(Y-\mu_Y)^2 = E(Y^2) - (\mu_Y)^2$
standard deviation of Y	$\sigma_Y = \sqrt{\sigma_Y^2}$
covariance between X and Y	$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$
correlation coefficient (between X and Y)	$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
expected value of the sample average, $\bar{Y}$	$E(\bar{Y}) = \mu_Y$
variance of the sample average, $\bar{Y}$	$\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$
t-statistic for testing μ_Y (for large n, and when σ_Y^2 is *known*)	$t = \frac{\bar{Y}^{act} - \mu_{Y,0}}{\sigma_{\bar{Y}}} \sim N(0,1)$
sample variance (estimator for σ_Y^2)	$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$
sample covariance (estimator for covariance)	$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$
sample correlation (estimator for correlation)	$r_{xy} = \frac{s_{xy}}{s_x s_y}$
standard error of $\bar{Y}$ (estimator for the standard deviation of $\bar{Y}$)	$s_{\bar{Y}} = \sqrt{\frac{s_Y^2}{n}}$
t-statistic for testing μ_Y (for large n, and when σ_Y^2 is *unknown*)	$t = \frac{\bar{Y} - \mu_{Y,0}}{s_{\bar{Y}}} \sim N(0,1)$
95% confidence interval for μ_Y (for large n)	$conf.int. = \bar{Y} \pm 1.96 \times s_{\bar{Y}}$
population linear regression model with one regressor	$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \qquad i = 1, \dots, n$
OLS estimator of the slope (β_1)	$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$
OLS estimator of the intercept (β_0)	$b_0 = \bar{Y} - b_1 \bar{X}$
OLS predicted values	$\hat{Y}_i = b_0 + b_1 X_i$
OLS residuals	$e_i = Y_i - \hat{Y}_i$

explained sum of squares	$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$
total sum of squares	$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$
sum of squared residuals	$RSS = \sum_{i=1}^{n} e_i^2$
regression R^2	$R^2 = \frac{ESS}{TSS}$
t-statistic for testing β_1	$t = \frac{b_1 - \beta_{1,0}}{s.e.(b_1)}$
95% confidence interval for β_1 (for large n)	$conf.int. = b_1 \pm 1.96 \times s.e.(b_1)$
alternative regression R^2	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
adjusted R-square ($\bar{R}^2$)	$\bar{R}^2 = 1 - \frac{RSS}{TSS}\left(\frac{n-1}{n-k-1}\right)$
F-statistic	$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(n - k_U - 1)}$
F-statistic	$F = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(n - k_U - 1)}$

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