# Econ 3040 Final Exam

Ryan T. Godwin

The exam is 3 hours long, and consists of 100 marks. **There are 15 questions**. There is a table of critical values for the F-statistic, a table of standard Normal probabilities, and a formula sheet, at the end of the exam.

**Short answer - each question worth 4 marks - 40 marks total**

1. A random variable $X$ is equal to 1 with probability 0.4, and equal to 4 with probability 0.6. What is the mean and variance of $X$?

2. How is the least-squares estimator derived? (Where does the equation for $b_0$, $b_1$, etc. come from?) Don't try to derive the formula, just set-up the problem, or describe the process.

3. What does it mean for least-squares to be the most "efficient" estimator?

4. Why does $R^2$ always increase when a variable is added to the model? How does $\bar{R}^2$ fix the problem?

5. Explain the main problem with the following population model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 male + \beta_3 female + \epsilon$$

6. This question uses the diamond price data:

```
summary(lm(price ~ carat + I(carat^2), data=diam)
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    -42.51     316.37  -0.134   0.8932
carat         2786.10    1119.61   2.488   0.0134 *
I(carat^2)    6961.71     868.83   8.013  2.4e-14 ***
```

What is the predicted increase in price due to an increase in carats? Your answer should include several numbers.

7. For the model in question 6, how would you go about determining the appropriate degree ($r$) of the polynomial?

**Long answer - each part worth 3 marks - 60 marks total**

8. Two models are estimated to explain the effect of installing a fireplace on the selling price of a house (in dollars). The R output for the regression results are given below:

```
house.mod1 <- lm(Price ~ Fireplaces + Bathrooms, data=house)
summary(house.mod1)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    44771       5743   7.796 1.10e-14 ***
Fireplaces     25414       3749   6.778 1.67e-11 ***
Bathrooms      79940       3167  25.241  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 77970 on 1725 degrees of freedom
Multiple R-squared:  0.3734,    Adjusted R-squared:  0.3727
F-statistic:   514 on 2 and 1725 DF,  p-value: < 2.2e-16
```

```
house.mod2 <- lm(Price ~ Fireplaces + Living.Area + Bathrooms, data=house)
summary(house.mod2)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -118.217   5369.069  -0.022    0.982
Fireplaces   5232.053   3384.481   1.546    0.122
Living.Area    91.431      3.928  23.276  < 2e-16 ***
Bathrooms   25511.611   3620.039   7.047 2.63e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 68030 on 1724 degrees of freedom
Multiple R-squared:  0.5232,    Adjusted R-squared:  0.5224
F-statistic: 630.6 on 3 and 1724 DF,  p-value: < 2.2e-16
```

a) What is the *main* difference between the two models? (If you had to focus on *just one* difference, what would it be?)

b) What is the problem with the first model? (Why is it worse than the second model?)

c) Using the second model: how much do you *predict* a 2000 square foot house with 2 bathrooms and 1 fireplace would sell for?

9. When estimating the model:

$$wage = \beta_0 + \beta_1 education + \beta_2 gender + \beta_3 age + \beta_4 experience + \epsilon$$

the results indicate that `age` and `experience` are *insignificant*:

```
summary(lm(wage ~ education + gender + age + experience, data=cps))
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.9574     6.8350  -0.286    0.775
education      1.3073     1.1201   1.167    0.244
genderfemale  -2.3442     0.3889  -6.028 3.12e-09 ***
age           -0.3675     1.1195  -0.328    0.743
experience     0.4811     1.1205   0.429    0.668
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.458 on 529 degrees of freedom
Multiple R-squared:  0.2533,	Adjusted R-squared:  0.2477
F-statistic: 44.86 on 4 and 529 DF,  p-value: < 2.2e-16
```

so, the variables `age` and `experience` are dropped from the model, and we get:

```
summary(lm(wage ~ education + gender, data=cps))
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.21783    1.03632   0.210    0.834
education      0.75128    0.07682   9.779  < 2e-16 ***
genderfemale  -2.12406    0.40283  -5.273 1.96e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.639 on 531 degrees of freedom
Multiple R-squared:  0.1884,	Adjusted R-squared:  0.1853
F-statistic: 61.62 on 2 and 531 DF,  p-value: < 2.2e-16
```

a) What are the benefits to "dropping" variables from a model?

b) Why shouldn't we use t-tests to determine if these two variables can be dropped?

c) Test the null hypothesis:

$$H_0 : \beta_3 = 0 \text{ and } \beta_4 = 0$$

What do you conclude?

10. The following population model:

$$\log(CO_2) = \beta_0 + \beta_1 \log(GDP) + \epsilon$$

is estimated in R:

```
co2mod <- lm(log(co2) ~ log(gdp.per.cap), data = co2)
summary(co2mod)
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -9.94045    0.36806  -27.01   <2e-16 ***
log(gdp.per.cap)  1.20212          ?   28.39   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6642 on 132 degrees of freedom
Multiple R-squared:  0.8593,	Adjusted R-squared:  0.8582
F-statistic: 806.1 on 1 and 132 DF,  p-value: < 2.2e-16
```

$CO_2$ is per capita carbon dioxide emissions, and $GDP$ is GDP per capita, for 134 different countries.

a) What is the interpretation of the estimated value of 1.20212?

b) What is the value for the missing `Std. Error`?

c) What is the F-statistic of 806.1 for? Why do you think that the square of the t-statistic is equal to this F-statistic ($28.39^2 = 806.1$)?

11. This question involves *heteroskedasticity*. First, a wage model is estimated using least squares (and the `summary()` command):

```
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.53764    0.70887   0.758 0.448521
education              0.18311    0.11333   1.616 0.106753
gendermale             0.69499    0.20315   3.421 0.000672 ***
age                   -0.06472    0.11345  -0.570 0.568616
experience             0.07754    0.11355   0.683 0.494959
education:gendermale  -0.03362    0.01531  -2.196 0.028545 *
```

then, *heteroskedastic* robust standard errors are calculated (using the "sandwich" and "lmtest" packages like you did in assignment 4):

```
                       Estimate Std. Error t value  Pr(>|t|)
(Intercept)            0.537643   0.194521  2.7639 0.0059104 **
education              0.183114   0.011411 16.0471 < 2.2e-16 ***
gendermale             0.694988   0.191017  3.6384 0.0003013 ***
age                   -0.064716   0.013117 -4.9339 1.082e-06 ***
experience             0.077542   0.014099  5.4997 5.936e-08 ***
education:gendermale  -0.033616   0.014731 -2.2819 0.0228902 *
```

a) What are homoskedasticity and heteroskedasticity?
b) What is wrong with assuming homoskedasticity, when there is actually heteroskedasticity?
c) How could you use the first estimated model to test for heteroskedasticity?
d) Point out the importance of using robust standard errors by using the output above.

12. This question involves *instrumental variables*. Consider the simple model:

$$y = \beta_0 + \beta_1 x + \epsilon$$

a) Suppose that there is a missing variable $m$ that is correlated with both the dependent variable $y$, and a regressor $x$. In this case, what happens to the least-squares estimator $b_1$? (What are the properties of $b_1$?)
b) If the missing variable $m$ cannot be found and included in the model, what is one solution to the problem?
c) What properties must an instrument $z$ have, in order to be "valid"? (In order for it to work in instrumental variables estimation?)

Now, consider the *wage*, *education*, and *distance from college* data. First a model is estimated by LS:

```
college <- read.csv("https://rtgodwin.com/data/collegedist.csv")
ls <- lm(wage ~ education + urban + gender + ethnicity + unemp, data=college)
summary(ls)
```

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       8.000192   0.156928  50.980   <2e-16 ***
education         0.005369   0.010362   0.518   0.6044
urbanyes          0.070117   0.044727   1.568   0.1170
gendermale        0.085242   0.037069   2.300   0.0215 *
ethnicityhispanic 0.012048   0.062385   0.193   0.8469
ethnicityother    0.556056   0.052167  10.659   <2e-16 ***
unemp             0.133101   0.006711  19.834   <2e-16 ***
```

and then by instrumental variables (IV) estimation, using *distance from college* as the instrument:

```
iv <- ivreg(wage ~ education + urban + gender + ethnicity + unemp |
                    distance + urban + gender + ethnicity + unemp,
            data=college)
summary(iv)
```

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.65702    1.83641  -0.358   0.7205
education          0.64710    0.13594   4.760 1.99e-06 ***
urbanyes           0.04614    0.06039   0.764   0.4449
gendermale         0.07075    0.04997   1.416   0.1569
ethnicityhispanic -0.12405    0.08871  -1.398   0.1621
ethnicityother     0.22724    0.09863   2.304   0.0213 *
unemp              0.13916    0.00912  15.259  < 2e-16 ***
```

d) Describe the major important difference between the two estimated models.

e) How does the two-stage least squares (2SLS) procedure work? Explain the steps using the above example.

13. This question uses a table of estimated models (see Table 1 on the next page), estimated from the CPS data set. **Note that the sample size** is $n = 534$.

   a) Test the hypothesis that the effect of education on wage is the same for women and men.

   b) Do you think that the effects of education **or** experience on log(wage) are linear or non-linear?

   c) Does ethnicity effect wages? (Do a hypothesis test.)

   d) What is the interaction term gendermale $\times$ marriedyes measuring?

   e) What are the estimated differences between the wages of married vs. unmarried workers, and men vs. women?

**END**

Table 1: Wage equations estimated from CPS data for question 13.

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | log(wage) | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| education | −0.017 | 0.023 | 0.013 | 0.023 | 0.091*** |
| | (0.091) | (0.046) | (0.048) | (0.046) | (0.008) |
| education$^2$ | 0.005 | 0.002 | 0.003 | 0.002 | |
| | (0.003) | (0.002) | (0.002) | (0.002) | |
| experience | 0.034*** | 0.033*** | 0.034*** | 0.033*** | 0.011*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.002) |
| experience$^2$ | −0.001*** | −0.001*** | −0.001*** | −0.001*** | |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) | |
| ethnicityhispanic | −0.078 | | −0.074 | | |
| | (0.091) | | (0.091) | | |
| ethnicityother | −0.072 | | −0.075 | | |
| | (0.059) | | (0.058) | | |
| regionsouth | −0.118*** | −0.125*** | −0.118*** | −0.125*** | −0.134*** |
| | (0.043) | (0.042) | (0.043) | (0.042) | (0.043) |
| gendermale | 0.043 | 0.121* | 0.132** | 0.121* | 0.123* |
| | (0.668) | (0.066) | (0.066) | (0.066) | (0.067) |
| marriedyes | −0.038 | −0.054 | −0.049 | −0.054 | −0.013 |
| | (0.060) | (0.060) | (0.061) | (0.060) | (0.061) |
| education × gendermale | 0.052 | | | | |
| | (0.103) | | | | |
| education$^2$ × gendermale | −0.003 | | | | |
| | (0.004) | | | | |
| gendermale × marriedyes | 0.178** | 0.200** | 0.184** | 0.200** | 0.202** |
| | (0.082) | (0.081) | (0.082) | (0.081) | (0.082) |
| Constant | 1.047* | 0.920*** | 0.996*** | 0.920*** | 0.590*** |
| | (0.610) | (0.315) | (0.327) | (0.315) | (0.133) |
| $R^2$ | 0.330 | 0.320 | 0.322 | 0.320 | 0.296 |
| Adjusted $R^2$ | 0.314 | 0.309 | 0.309 | 0.309 | 0.288 |
| *Note:* | | | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 2:

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | log(wage) | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| education | $0.045^{***}$ | $0.056^{***}$ | $0.045^{***}$ | $0.057^{***}$ | $0.046^{***}$ | $0.044^{***}$ |
| | (0.006) | (0.005) | (0.006) | (0.005) | (0.006) | (0.007) |
| experience | $0.015^{***}$ | $0.016^{***}$ | $0.014^{***}$ | $0.015^{***}$ | $0.016^{***}$ | $0.014^{***}$ |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| age | $0.019^{***}$ | $0.019^{***}$ | $0.019^{***}$ | $0.019^{***}$ | $0.019^{***}$ | $0.019^{***}$ |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) |
| female | $-0.286^{**}$ | $0.051^{**}$ | $-0.264^{**}$ | | $-0.224^{*}$ | $-0.284^{**}$ |
| | (0.123) | (0.022) | (0.123) | | (0.117) | (0.125) |
| Manitoba | $-0.164^{***}$ | $-0.168^{***}$ | $-0.117^{***}$ | $-0.117^{***}$ | $-0.167^{***}$ | |
| | (0.028) | (0.028) | (0.020) | (0.020) | (0.028) | |
| fem_educ | $0.020^{**}$ | | $0.020^{**}$ | | $0.019^{**}$ | $0.021^{**}$ |
| | (0.008) | | (0.008) | | (0.008) | (0.008) |
| fem_exper | $0.002^{*}$ | | $0.002^{*}$ | | | $0.003^{**}$ |
| | (0.001) | | (0.001) | | | (0.001) |
| fem_Manitoba | $0.093^{**}$ | $0.096^{**}$ | | | $0.095^{**}$ | |
| | (0.039) | (0.039) | | | (0.039) | |
| Constant | $1.804^{***}$ | $1.643^{***}$ | $1.795^{***}$ | $1.638^{***}$ | $1.775^{***}$ | $1.775^{***}$ |
| | (0.086) | (0.063) | (0.086) | (0.064) | (0.084) | (0.087) |
| Observations | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| $R^2$ | 0.759 | 0.757 | 0.757 | 0.750 | 0.758 | 0.749 |
| Adjusted $R^2$ | 0.757 | 0.755 | 0.756 | 0.749 | 0.756 | 0.747 |
| *Note:* | | | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 3: 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 4: 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 |