Final Exam – 150 Points

You must answer all the questions. The exam is closed book and closed notes. You may use calculators, but they must not be graphing calculators. Do not use your own scratch paper.

You must show your work to receive full credit.

You have plenty of time to finish. Take your time and relax. And, have a safe and wonderful Summer!

Problem 1 (30 Points)

You roll two dice. Both dice are fair. The first one has THREE sides and the second one has SIX sides.

a.) Draw and label the Venn diagram describing all possible sample points. (5 Points)

Also possible to use a tree diagram.

b.) What is the probability that you will get a total of four or more points between the two dice? (5 Points)

\[
P(T \geq 4) = \frac{15}{18} = \frac{5}{6}
\]
e.) Given that you roll a two with one of the two dice what is the chance that the two dice together will total 4? (10 Points)

\[ P(T = 4 \mid 2) = \frac{P(T = 4 \cap 2)}{P(2)} \]

\[ P(2) = \frac{6}{18} + 2 \]

\[ P(T = 4 \mid 2) = \frac{1}{18} + 2 \]

\[ P(T = 4 \mid 2) = \frac{1}{18} + \frac{6}{18} = \frac{1}{8} + 2 \]


d.) Given that you roll a three with one of the two dice what is the chance that the two dice together will total a value greater than 4? (10 Points)

\[ P(T > 4 \mid 3) = \frac{P(T > 4 \cap 3)}{P(3)} \]

\[ P(3) = \frac{6}{18} + 2 \]

\[ P(T > 4 \mid 3) = \frac{6}{18} + 2 \]

\[ P(T > 4 \mid 3) = \frac{6}{18} + \frac{6}{18} = \frac{1}{8} \]

+ 2
Problem 2 (90 points)

Suppose that I run the following regression predicting the effects of classroom performance on students’ final exam grades:

\[ \text{final} = \beta_0 + \beta_1 \text{section} + \beta_2 \text{mtl} + \beta_3 \text{hwtotal} + u \]

Here, \( \text{final, mtl, hwtotal, section} \) are the percent scores on the final, midterm, homework, and section participation, respectively. The results from running this regression are below.

```
. regress final section mtl hwtotal

Source | SS        | df | MS          | Number of obs = 142
--------|-----------|----|-------------|------------------
Model   | 12155.6037| 3  | 4051.86791  | F( 3, 138) = 26.49
Residual| 21109.933 | 138| 152.970529  | Prob > F = 0.0000
--------|-----------|----|-------------|------------------
Total   | 33265.5367| 141| 235.925792  | R-squared = 0.3654
         |           |    |             | Adj R-squared = 0.3516
         |           |    |             | Root MSE = 12.368

| final | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-------|-------|-----------|-------|-----|----------------------|
| section | .0795122 | .058387 | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx |
| mtl     | .4671107  | .0669202 | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx |
| hwtotal | .235302   | .0720338 | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx |
| cons    | 14.46427  | 7.723413 | xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx |
```

a.) Please interpret the constant. (5 points)

On average, a student that earns \( 0\% \) in section, \( 0\% \) in the first midterm, and \( 0\% \) in homework will earn \( 14.46\% \) on the final exam.

b.) I claim that getting a higher grade on homework increases your predicted grade on the final. Conduct a one-sided hypothesis test at the 5% level for the coefficient on \( \text{hwtotal, } \beta_3 \). Please state your null and alternative hypotheses, and briefly interpret the result. (10 Points)

\[ H_0: \beta_3 = 0 \]
\[ H_a: \beta_3 > 0 \]

\[ t_{stat} = \frac{0.235 - 0}{0.0721} = 3.26 \]

\[ t_{crit} = 1.647 \]

\[ t_{stat} > t_{crit} \]

Reject null hypothesis.

Homework performance has a positive effect on grades which has a statistically significant difference from zero.
c.) Construct a 99% confidence interval for the coefficient on section, $\beta_1$. (10 Points)

$$0.0795 - 0.0584 < \beta_1 < 0.0795 + 0.0584 \cdot 2.57$$

$$-0.0706 < \beta_1 < 0.23$$

d.) I have reason to suspect that the variability of final exam scores changes with previous performance (homework, midterms, section). What is this called? What can be done about it? What Stata commands are necessary? (5 Points)

* Heteroskedasticity + 2

* Robust standard errors + 2

* Use 'robust' command, + 1
e.) I want to test the suspicion in ‘d’ rigorously. I run the following regression:

\[ \hat{y} = \delta_0 + \delta_1 \text{section} + \delta_2 \text{mtl} + \delta_3 \text{hwtotal} + \varepsilon \]

Here, \( \hat{u} \) is the residual from the regression in ‘a’. The estimates are as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 142</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7.2760e-12</td>
<td>3</td>
<td>2.4253e-12</td>
<td>F(3, 138) = 0.00</td>
</tr>
<tr>
<td>Residual</td>
<td>21040.4593</td>
<td>138</td>
<td>152.467096</td>
<td>Prob &gt; F = 1.0000</td>
</tr>
<tr>
<td>Total</td>
<td>21040.4593</td>
<td>141</td>
<td>149.223115</td>
<td>R-squared = 0.0000</td>
</tr>
</tbody>
</table>

\[ \text{Adj R-squared} = -0.0217 \]
\[ \text{Root MSE} = 12.348 \]

| what | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|------|-------|-----------|------|-------|----------------------|
| section | -3.84e-09 | 0.0582909 | -0.00 | 1.000 | -0.152588 to 0.1152588 |
| mtl     | -1.13e-08 | 0.06681   | -0.00 | 1.000 | -0.1321036 to 0.1321036 |
| hwtotal | 1.12e-08  | 0.0719152 | 0.00  | 1.000 | -0.1421981 to 0.1421982 |
| cons    | 1.08e-07  | 0.710694  | 0.00  | 1.000 | -15.24638 to 15.24638  |

The f-statistic for the full exclusionary test is very low (zero), which implies that the variables of the model tell us very little about the dependent variable. Does this address the assertion in ‘d’? If not, suggest an alternative. What assumption is at play here? (10 Points)

No, does not. \[ \hat{u}^2 \]

Since \( E(u|x) = 0 \), the residuals will not be independent of the explanatory variables.

Instead, you should regress:

\[ \hat{u}^2 = \delta_0 + \delta_1 \text{section} + \delta_2 \text{mtl} + \delta_3 \text{hwtotal} + \varepsilon \]

And run a full-exclusion test.
f) I suspect that the return to homework scores is dependent on whether or not you attend sections. To examine this possibility, I run the following regression:

\[ \text{final} = \beta_0 + \beta_1 \text{section} + \beta_2 \text{mtl} + \beta_3 \text{hwtotal} + \beta_4 \text{hwtotal} \times \text{section} + u \]

The results from estimating this equation are below:

```
. regress final section mtl hwtotal hwtotal*section

                         Number of obs =      142
                           F(  4,    137) =     19.90
                          Prob > F =    0.0000
                     R-squared =     0.3675
                      Adj R-squared =     0.3490
               Root MSE =        12.393

                      Source |       SS      df    MS
-------------------|-------------------------------
                Model |   12225.0771       4  3056.26926
               Residual |  21040.4597     137  153.579998
-------------------|-------------------------------
                   Total |  33265.5367     141  235.925792

                      final | Coef.  Std. Err.     t    P>|t|  [95% Conf. Interval]
-------------------|-----------------------------------------
                section |   .0842367     .0589235   1.43   0.155    -.0322803    .2007538
                mtl  |    .7592581     .4395156   1.73   0.086    -.1098538    1.62837
                hwtotal |    .4647203     .3486564   1.33   0.185    -.2247237    1.154164
hwtotal*section   |  -.0034208     .0050861  -.63   0.531  -.0133838    .0065360
                _cons |  -.5.333285    30.43569  -0.17   0.861    -.65.51776    54.85119
```

Derive the return to section attendance. Plug in the estimated coefficients where necessary. Please interpret briefly. (10 Points)

\[
\frac{\partial \text{final}}{\partial \text{section}} = \beta_1 + \beta_4 \text{hwtotal} \\
= 0.0842 - 0.0034 \text{hwtotal} + 3
\]

The returns to section attendance are smaller as you get better grades on homework + 3
g.) What is the homework score which yields a negative return to section attendance? Given that homework scores are between 0 and 100, is the return to section attendance always positive? (10 Points)

\[
\frac{d Final}{d Section} = 0 \quad \text{if} \quad \beta_i + \beta_y \text{untotal} = 0
\]

\[
\text{untotal} = -\frac{\beta_i}{\beta_y}
\]

For homework scores above 24.62, the returns to section are negative. 

\[
-\frac{(0.0842)}{(-0.00342)} = 24.62
\]

h.) Is there a significant interaction between homework and section attendance? Conduct a two-sided test at the 1% level, stating your null and alternative hypotheses, also briefly interpreting the result. (10 Points)

\[
H_0: \beta_y = 0 \quad H_1: \beta_y \neq 0
\]

\[
t_{stat} = \frac{-0.00342 - 0}{0.00509} = -0.672
\]

\[
t_{crit} = 2.57
\]

\[
|t_{crit}| < |t_{stat}| \Rightarrow \text{Fail to reject the null.}
\]

There exists no significant interaction between the returns to section and homework, and vice versa.
i.) Rather than using interactions as in ‘f’, I have added in squared terms of homework, *hwtotalsqr*, and section attendance, *sectionsqr*. Their coefficients are $\beta_5$ and $\beta_6$, respectively.

```
. regress final section mtl hwtotal hwtotalsqr sectionsqr
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 142</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>12244.9991</td>
<td>5</td>
<td>2448.99983</td>
<td>$F(5, 136) = 15.84$</td>
</tr>
<tr>
<td>Residual</td>
<td>21020.5376</td>
<td>136</td>
<td>154.562776</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R-squared = 0.3681</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.3449</td>
</tr>
<tr>
<td>Total</td>
<td>33265.5367</td>
<td>141</td>
<td>235.925792</td>
<td>Root MSE = 12.432</td>
</tr>
</tbody>
</table>

| final                  | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|------------------------|-------|-----------|------|------|---------------------|
| section                | -0.0044826 | .2068736 | -0.02| 0.983| -.4135877 to .4046226 |
| mtl                    | .4710029  | .0674799 | 6.98 | 0.000| .3375573 to .6044485 |
| hwtotal                | .4722948  | .3601857 | 1.31 | 0.192| -.2399944 to 1.184584 |
| hwtotalsqr             | -.0017249 | .0025807 | -0.67| 0.505| -.0068284 to .0033786 |
| sectionsqr             | .0006319  | .0015609 | 0.40 | 0.686| -.0024549 to .0037186 |
| _cons                  | 9.323664  | 13.5762  | 0.69 | 0.493| -17.52409 to 36.17142 |

Which model is preferred, the one in ‘i’, the one in ‘a’? Please justify your answer. If a test is required, state your null and alternative hypotheses, test it at the 5% level, and briefly interpret the result. (10 Points)

$$H_0: \beta_5 = \beta_6 = 0$$
$$H_a: \beta_5 \text{ not time} + 1$$

$$F_{stat} = \frac{21109 - 21020}{154.6} = 0.287$$

Fail to reject $H_0$. +2

j.) Suppose that natural ability is an unobserved variable, which does not change over time. I am worried that not including it may be causing omitted variable bias. What technique is appropriate for this problem, and why? (10 Points)

Differencing +5

Because natural ability is time invariant, taking the difference between two years yields a specification where ability has no effect.
Problem 4 (30 Points)

Professor Spearot is getting older. He is worried about a receding hair line. To analyze male hair patterns as a function of demographics, he estimates the following linear probability model using a sample of men:

\[ Bald = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Dad} + u \]

*Bald* takes on the value of 1 if the respondent is bald, and 0 otherwise. *Age* is the Age of the respondent, and *Dad* is an indicator variable taking the value of 1 if the respondent’s father is bald and 0 otherwise.

a.) Suppose that \( \beta_2 \) is positive. How do I interpret the estimate of the coefficient on *Dad*, \( \beta_2 \)? (5 Points)

Having a Dad that is bald increases the probability that I am bald by \( \beta_2 \).

b.) Suppose that I estimate the model, and I generate predictions for each respondent. Some predictions are negative. Is this sensible? What alternative estimation procedure could remedy this problem? Why? (10 Points)

No, since the predictions are probabilities and negative probabilities don't make sense.

Use Probit or Logit. Probability \( \in [0, 1] \) with probit and logit.

c.) Suppose that Stress, an unobserved variable, increases with age. Stress also lead to a higher likelihood of baldness. What is this called? In what direction is the bias? (5 Points)

Omitted variable bias. +3

Positive Bias. \( \beta_1 \) is over-estimated. +2
d.) Professor Spearot’s father is Bald (sorry Dad!). Professor Spearot is 29 years old. Please derive the estimating equation required to generate a prediction for somebody with Professor Spearot’s characteristics. Please also write the precise STATA commands required to run this regression. (10 points)

\[
\hat{\theta} = \beta_0 + \beta_1 \cdot 29 + \beta_2 \cdot 1 + \epsilon
\]

\[
\text{Build} = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Dad}
\]

\[
\hat{\theta} = \hat{\beta}_1 - \beta_1 \cdot 29 - \beta_2 \cdot 1 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Dad}
\]

\[
\hat{\text{Build}} = \hat{\beta}_0 + \hat{\beta}_1 \cdot (\text{Age} - 29) + \hat{\beta}_2 \cdot (\text{Dad} - 1)
\]

\[
\text{gen \ Dad} - 1 = \text{Dad} - 1
\]

\[
\text{gen \ Age} - 29 = \text{Age} - 29
\]

\[
\text{regress \ Build \ Age} - 29 \ \text{Dad} - 1
\]
Extra Credit: (10 Points)

Bob Baden was once a college hockey player (no joke here). Skilled and graceful, he was an offensive weapon.

Suppose that Bob takes three shots at the net. The probability of scoring on the first shot is 0.5. Each time he scores, the probability of scoring on the next shot goes up by 0.1. What is the probability of scoring on the 3rd shot?

\[ S = \text{score} \]
\[ M = \text{miss} \]

\[
\begin{align*}
&\text{S}\quad \text{S}\quad \text{S} \\
&\quad \text{M}\quad \text{S}\quad \text{S} \\
&\qquad \text{M}\quad \text{S}\quad \text{M} \\
&\qquad \quad \text{S}\quad \text{M}\quad \text{S} \\
&\qquad \quad \quad \text{S}\quad \text{S}\quad \text{S} \\
&\quad \text{A}\quad \text{B}\quad \text{C}\quad \text{D}\quad \text{E}\quad \text{F}\quad \text{G}\quad \text{H} \\
&\quad 0.21 \quad 0.09 \quad 0.12 \quad 0.06 \quad 0.15 \quad 0.1 \quad 0.125 \quad 0.125
\end{align*}
\]

\[
\Pr(\text{score on third shot}) = \Pr(A \text{ or C or E or H})
\]
\[
= \Pr(A) + \Pr(C) + \Pr(E) + \Pr(H)
\]
\[
= 0.21 + 0.12 + 0.15 + 0.125
\]
\[
= 0.685
\]