Exam 3 – 80 Points

You must answer all questions. Please write your name on every page. The exam is closed book and closed notes. You may use calculators, but no cell phones. Do not use your own scratch paper.

**You must show your work to receive full credit**

I have neither given nor received unauthorized aid on this examination, nor have I concealed any similar misconduct by others.

Signature

Problem 1 (40 Points)

We wish to predict college outcomes using the following regression:

\[
\text{college} = \beta_0 + \beta_1 \text{mom_college} + \beta_2 \text{dad_college} + u
\]

Here, college is a dummy variable taking on a value of 1 for respondents with 16 or more years of education, and zero otherwise. The dummy variables mom_college and dad_college take on a value of 1 if the mom and dad went to college, respectively, and zero otherwise.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 722</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>XXXXXXXX</td>
<td>2</td>
<td>6.40338444</td>
<td>F(  2,   719) = 33.13</td>
</tr>
<tr>
<td>Residual</td>
<td>XXXXXXXX</td>
<td>719</td>
<td>.193286522</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>XXXXXXXX</td>
<td>721</td>
<td>.210512869</td>
<td>Adj R-squared = 0.0818</td>
</tr>
</tbody>
</table>

| college | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|----------------------|
| mom_college | .331223 | .0724886  | XX | XXXXX | XXXXXXXXXX |
| dad_college | .2727932 | .0654208  | XX | XXXXX | XXXXXXXXXX |
| _cons | .2558639 | .0172718  | XX | XXXXX | XXXXXXXXXX |

a.) Please construct and interpret a 90% confidence interval for the intercept. Show your work! (10 Points)

\[
0.2558639 - 0.172718 \times 1.645 < B_0 < 0.2558639 + 0.172718 \times 1.645
\]

\[
0.227 < B_0 < 0.284 +4
\]

+2

For a respondent with parents that did not go to college, with 90% confidence, the probability of going to college is between 0.227 and 0.284. +4
b.) Please interpret the coefficient on `mom_college`. At the 99% confidence level, please test whether it is greater than zero using a one-sided test, and briefly interpret your result. Show your work!! (10 Points)

\[ \begin{align*}
\text{Ho: } B_1 &= 0 \ (<=0 \text{ also fine}) \\
\text{Ha: } B_1 &> 0 \\
T_{\text{crit}} &= 2.3 \\
T_{\text{Stat}} &= (0.331 - 0)/0.0725 = 4.5
\end{align*} \]

\[ 4.5 > 2.3 \implies \text{Reject the null!} \]

*Holding father’s college outcome constant, at the 99% level of confidence, having a mom that went to college has a significant and positive effect on the respondent going to college.*

\[ +3 \]

c.) It appears that having a mother who went to college has a larger effect on college outcomes than having a father who went to college. Please derive an equation that allows me to test whether the effect of the mother’s college outcome is the same as the father’s college outcome. Along with the derivation, please state the null and alternative hypotheses, and write down any Stata commands required to generate new variables and run the regression. Show your work! (10 Points)

\[ \begin{align*}
\text{Ho: } B_1 - B_2 &= 0 \\
\text{Ha: } B_1 - B_2 &\neq 0
\end{align*} \]

\[ \theta = B_1 - B_2 \] Solving for \( B_2 \)

\[ B_2 = B_1 - \theta \]

*Substituting for \( B_2 \) in the regression equation, we get:*

\[ \begin{align*}
\text{college} &= \beta_0 + \beta_1 \text{mom\_college} + (\theta + \beta_1) \text{dad\_college} + u \\
\text{college} &= \beta_0 + \beta_1 (\text{mom\_college} + \text{dad\_college}) + \theta \text{dad\_college} + u
\end{align*} \]

*Stata Commands:

\[ \text{gen parent\_college = mom\_college+dad\_college} \]

\[ \text{regress college parent\_college dad\_college} \]
For the next few regressions, we add an effect of siblings, \textit{sibs}, which is the number of siblings of the respondent. Specifically, we estimate the following:

\[ \text{college} = \beta_0 + \beta_1 \text{mom\_college} + \beta_2 \text{dad\_college} + \beta_3 \text{sibs} + u \]

The results are the following:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 722</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>16.6200171</td>
<td>3</td>
<td>5.5400057</td>
<td>F( 3, 718) = 29.43</td>
</tr>
<tr>
<td>Residual</td>
<td>135.159761</td>
<td>718</td>
<td>.188244793</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>151.779778</td>
<td>721</td>
<td>.210512869</td>
<td>Adj R-squared = 0.1058</td>
</tr>
</tbody>
</table>

| college | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|----------------------|
| mom\_college | .3245219 | .0715525 | XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
| dad\_college | .2621026 | .0646056 | XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
| sibs     | -.0323631 | .0071906 | XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
| _cons    | .3497149 | .0269324 | XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |

Which regression is preferred, the regression in ‘1a’ or ‘1d’? Please test this hypothesis at the 95% level, stating your null and alternative hypotheses. Briefly interpret your result, and show your work!! \textbf{(10 points)}

(I know this is different and that I X’d out something that you want to use. But think about it and you will get it!)

\textbf{F tests and t tests are the same with one variable restrictions. So use a two-sided t-test.}

\textit{Ho:} \beta_3=0 \; +1

\textit{Ha:} \beta_3 \neq 0 \; +1

\[ T_{\text{crit}} = 1.96 \; +1 \]

\[ T\text{-Stat} = (-.0323631-0)/.0071906= -4.50 \; +3 \]

\[ |tstat| > tcrit \; Reject the null!! \; +1 \]

Regression in part ‘d’ is preferred. Sibs has a significant effect on college attendance. \textbf{+3}

\textbf{(4 points max if adj R2 was used instead of the correct approach)}
e.) You’re unhappy with the regression in ‘d’, and produce an interaction between sibs and parental education.

\[ \text{college} = \beta_0 + \beta_1 \text{mom}_{-} \text{college} + \beta_2 \text{dad}_{-} \text{college} + \beta_3 \text{sibs} + \beta_4 \text{mom}_{-} \text{college} \cdot \text{sibs} + \beta_5 \text{dad}_{-} \text{college} \cdot \text{sibs} + u \]

The results are below:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 722</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>17.305119</td>
<td>5</td>
<td>3.46102379</td>
<td>F(  5,   716) = 18.43</td>
</tr>
<tr>
<td>Residual</td>
<td>134.474659</td>
<td>716</td>
<td>.18781377</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>151.779778</td>
<td>721</td>
<td>.210512869</td>
<td>R-squared = 0.1140</td>
</tr>
</tbody>
</table>

| college     | Coef.    | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-------------|----------|-----------|-------|-----|---------------------|
| mom_{-}college | .112594  | .1339742  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |
| dad_{-}college | .3460208 | .1252459  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |
| mom_{-}college\_sibs | .0951133 | .0523259  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |
| dad_{-}college\_sibs | -.0467353| .0497364  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |
| sibs        | -.0343139| .0073956  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |
| _cons       | .3561122 | .0274761  | XXXXX|     | XXXXXXXXXXXXXXXXXXXXX |

Which regression is preferred, the regression in ‘1d’ or ‘1e’? Please test this hypothesis at the 95% level, stating your null and alternative hypotheses. Briefly interpret your result, and show your work!! (10 Points)

Ho: B4=0,B5=0 +.5
Ha: Ho not true +.5

q=2 +.5
df_{u}=716 +.5
SSR_{u}=134.47 +.5
SSR_{r}=135.16 +.5

F_{crit}=3 +1
F_{stat}=((135.16-134.47)/2) / (134.47/716) =1.83 +3
F_{stat}<F_{crit}
Fail to reject the null! +1

The interaction terms do not have a significant effect on college choices. Model in ‘d’ is preferred to the model in ‘e’ +2
f.) Suppose I claim that having a mother who attended college affects the relationship between siblings and the respondent’s college outcome. What is the probability that I’m wrong? (10 Points)

\[ \text{Tstat} = \frac{.0951133-0}{.0523259} = 1.8177 +4 \]

\[ \text{P-value} = 2\times(1 - \text{Pr}(Z<1.8177)) \]
\[ = 2\times(1 - 0.9656) \]
\[ = 0.0688 +6 \]

(4 points max if Pvalue calculated correctly but for incorrect coefficient)

Problem 2 (25 Points)

a.) For this problem, we wish to associate wages with education, location, and age:

\[ \ln(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{urban} + \beta_3 \text{age} + \beta_4 \text{age}^2 + u \]

Here, \text{wage} is the monthly wage in dollars, \text{urban} is a dummy variable identifying respondents that live in metropolitan areas, \text{educ} is years of schooling, and \text{age} is the age of the respondent. Results:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 935</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>26.2161887</td>
<td>4</td>
<td>6.55404717</td>
<td>F(4, 930) = 43.71</td>
</tr>
<tr>
<td>Residual</td>
<td>139.440095</td>
<td>930</td>
<td>.149935586</td>
<td>R-squared = 0.1583</td>
</tr>
<tr>
<td>Total</td>
<td>165.656283</td>
<td>934</td>
<td>.177362188</td>
<td>Adj R-squared = 0.1546</td>
</tr>
</tbody>
</table>

\[ \text{Root MSE} = 0.38722 \]

| \text{ln_wage} | Coef.    | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------------|----------|-----------|-------|--------|-----------------------|
| educ           | .05774   | .0058051  | xxxxxxxxx | xxxxxxxxx |
| urban          | .1714169 | .0282304  | xxxxxxxxx | xxxxxxxxx |
| age            | .0137273 | .1028383  | xxxxxxxxx | xxxxxxxxx |
| age2           | .0001333 | .0015446  | xxxxxxxxx | xxxxxxxxx |
| _cons          | 5.277044 | 1.69378   | xxxxxxxxx | xxxxxxxxx |

Is there an age at which wages are maximized? If so, solve for this age. If not, tell me why. Show your work!! (10 Points)

No, there is not. Differentiating

\[ \frac{d\ln(\text{wage})}{d\text{age}} = 0.0137273 + 2\times\text{age} \times 0.0001333 \]
\[ \frac{d^2\ln(\text{wage})}{d\text{age}^2} = 2\times 0.0001333 = 0.0002666 \]

Second derivative is positive. Therefore, the function has an age at which wages are minimized, but no maximum. +6
b.) Please precisely interpret the coefficient on urban. Show your work! (5 Points)

Note that the left hand side variable is the natural log of wages. So, we exponentiate the coefficient on urban to get the precise effect

\[
\%\Delta \text{wage}/\Delta \text{urban} = 100^* (\exp(B_2 \Delta \text{urban})-1)
\]

Note that moving to urban implies \( \Delta \text{urban} = 1 \)

\[
\%\Delta \text{wage}/\Delta \text{urban} = 100^* (\exp(.1714)-1)=18.69 \quad +2
\]

+1 +1 +1
Holding education and age constant, living in an urban area increases the wage by 18.69% relative to living in a rural area.

---

c.) What is the precise effect of going from 12 years of education (HS) to 22 years of education (PhD)? Show your work! (5 Points)

Note that the left hand side variable is the natural log of wages. So, we exponentiate the change in education to get the precise effect.

\[
\%\Delta \text{wage}/\Delta \text{urban} = 100^* (\exp(B_1 \Delta \text{educ})-1)
\]

Note that moving to urban implies \( \Delta \text{educ} = 10 \)

\[
\%\Delta \text{wage}/\Delta \text{urban} = 100^* (\exp(.05774*10)-1)=78.14 \quad +3
\]

+1 +1
Holding age and urban constant, going from a high school education to a PhD increases your wage by 78.14%

---

Have a great weekend!
Normal Distribution
from -oo to Z

5 | 0.00  0.01  0.02  0.03  0.04  0.05  0.06  0.07  0.08  0.09
---|-------------------
 0.0| 0.5000 0.5040 0.5080 0.5120 0.5160 0.5199 0.5239 0.5279 0.5319 0.5359
 0.1| 0.5398 0.5438 0.5478 0.5517 0.5557 0.5596 0.5636 0.5675 0.5714 0.5753
 0.2| 0.5793 0.5832 0.5871 0.5910 0.5948 0.5987 0.6026 0.6064 0.6103 0.6141
 0.3| 0.6179 0.6217 0.6255 0.6293 0.6331 0.6368 0.6406 0.6443 0.6480 0.6517
 0.4| 0.6554 0.6591 0.6628 0.6664 0.6700 0.6736 0.6772 0.6808 0.6844 0.6879
 0.5| 0.6915 0.6950 0.6985 0.7019 0.7054 0.7088 0.7123 0.7157 0.7190 0.7224
 0.6| 0.7257 0.7291 0.7324 0.7357 0.7389 0.7422 0.7454 0.7486 0.7517 0.7549
 0.7| 0.7580 0.7611 0.7642 0.7673 0.7704 0.7734 0.7764 0.7794 0.7823 0.7852
 0.8| 0.7881 0.7910 0.7939 0.7967 0.7995 0.8023 0.8051 0.8078 0.8106 0.8133
 0.9| 0.8159 0.8186 0.8212 0.8238 0.8264 0.8289 0.8315 0.8340 0.8365 0.8389
 1.0| 0.8413 0.8438 0.8461 0.8485 0.8508 0.8531 0.8554 0.8577 0.8599 0.8621
 1.1| 0.8643 0.8665 0.8686 0.8708 0.8729 0.8749 0.8770 0.8790 0.8810 0.8830
 1.2| 0.8849 0.8869 0.8888 0.8907 0.8925 0.8944 0.8962 0.8980 0.8997 0.9015
 1.3| 0.9032 0.9049 0.9066 0.9082 0.9099 0.9115 0.9131 0.9147 0.9162 0.9177
 1.4| 0.9192 0.9207 0.9222 0.9236 0.9251 0.9265 0.9279 0.9292 0.9306 0.9319
 1.5| 0.9332 0.9345 0.9357 0.9370 0.9382 0.9394 0.9406 0.9418 0.9429 0.9441
 1.6| 0.9452 0.9463 0.9474 0.9484 0.9495 0.9505 0.9515 0.9525 0.9535 0.9545
 1.7| 0.9554 0.9564 0.9573 0.9582 0.9591 0.9599 0.9608 0.9616 0.9625 0.9633
 1.8| 0.9641 0.9649 0.9656 0.9664 0.9671 0.9678 0.9686 0.9693 0.9699 0.9706
 1.9| 0.9713 0.9719 0.9726 0.9732 0.9738 0.9744 0.9750 0.9756 0.9761 0.9767
 2.0| 0.9772 0.9778 0.9783 0.9788 0.9793 0.9798 0.9803 0.9808 0.9812 0.9817
 2.1| 0.9821 0.9826 0.9830 0.9834 0.9838 0.9842 0.9846 0.9850 0.9854 0.9857
 2.2| 0.9861 0.9864 0.9868 0.9871 0.9875 0.9878 0.9881 0.9884 0.9887 0.9890
 2.3| 0.9893 0.9896 0.9898 0.9901 0.9904 0.9906 0.9909 0.9911 0.9913 0.9916
 2.4| 0.9918 0.9920 0.9922 0.9925 0.9927 0.9929 0.9931 0.9932 0.9934 0.9936
 2.5| 0.9938 0.9940 0.9941 0.9943 0.9945 0.9946 0.9948 0.9949 0.9951 0.9952
 2.6| 0.9953 0.9955 0.9956 0.9957 0.9959 0.9960 0.9961 0.9962 0.9963 0.9964
 2.7| 0.9965 0.9966 0.9967 0.9968 0.9969 0.9970 0.9971 0.9972 0.9973 0.9974
 2.8| 0.9974 0.9975 0.9976 0.9977 0.9977 0.9978 0.9979 0.9979 0.9980 0.9981
 2.9| 0.9981 0.9982 0.9982 0.9983 0.9984 0.9984 0.9985 0.9985 0.9986 0.9986
 3.0| 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9989 0.9989 0.9990 0.9990

TABLE G.3b
5% Critical Values of the F Distribution

<table>
<thead>
<tr>
<th>Numerator Degrees of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>∞</td>
</tr>
</tbody>
</table>

**Example:** The 5% critical value for numerator $df = 4$ and large denominator $df (∞)$ is 2.37.

**Source:** This table was generated using the Stata® function invFtail.