Final Exam Version B – 80 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 Happy Holiday!

1. You have four kids that that weigh 50, 60, 70, and 80 pounds. Their respective heights are 2, 3, 4, and 3 ft.

a. What is the covariance between height and weight? (5 points)

\[
\mu_w = \frac{50 + 60 + 70 + 80}{4} = \frac{260}{4} = 65
\]

\[
\mu_H = \frac{2 + 3 + 4 + 3}{4} = \frac{12}{4} = 3
\]

\[
\sigma_{hw} = \frac{1}{(4-1)} \left( (50-65)^2 \cdot 2 - 3 + (60-65)^2 \cdot 3 - 3 + (70-65)^2 \cdot 4 - 3 + (80-65)^2 \cdot 3 - 3 \right)
\]

\[
= \frac{1}{3} \left( 15 + 5 \right) = \frac{20}{3}
\]

+ 1

+ 1

+ 1 for some work

\( N \)
b. Suppose I estimate \( \text{Height} = \beta_0 + \beta_1 \text{Weight} + u \) using a different sample (not your answer from a). The sample covariance of \( \text{Height} \) and \( \text{Weight} \) is 10. The sample variance of \( \text{Weight} \) is 2. What is the estimate of \( \beta_1 \)? (5 points)

\[
\hat{\beta}_1 = \frac{\sigma_{\text{hw}}}{\sigma_{\text{w}}^2} = \frac{10}{2} = 5
\]

2. You wish to predict the effects of education, experience, and tenure on wage outcomes. Specifically, you estimate the following specification:

\[
\log(\text{wage}) = \beta_0 + \beta_{\text{educ}} + \beta_{\text{tenure}} + \beta_{\text{exper}} + u
\]

The results from running this regression are below:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 935</th>
<th>F(3, 931) = 56.97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>25.6953278</td>
<td>3</td>
<td>8.56510927</td>
<td></td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>139.960966</td>
<td>931</td>
<td>.150334013</td>
<td></td>
<td>0.1551</td>
</tr>
<tr>
<td>Total</td>
<td>165.656294</td>
<td>934</td>
<td>.177362199</td>
<td></td>
<td>0.1524</td>
</tr>
</tbody>
</table>

| lwage | Coef. | Std. Err. | t     | P>|t|       | [95% Conf. Interval] |
|-------|-------|-----------|-------|----------|----------------------|
| educ  | .0748638 | .0065124 | xxxx | xxxxx    | xxxxxxxxx - xxxxxxxxx |
| tenure| .0133748 | .0025872 | xxxx | xxxxx    | xxxxxxxxx - xxxxxxxxx |
| exper | .0153285 | .0033696 | xxxx | xxxxx    | xxxxxxxxx - xxxxxxxxx |
| cons  | 5.496696 | .1105282 | xxxx | xxxxx    | xxxxxxxxx - xxxxxxxxx |

a. Do a two-sided t-test at the 5% level to determine if experience (\( \text{exper} \)) is a statistically significant determinant of the log wage. Please state the null and alternative hypotheses, and interpret the result. (5 points)

\[
H_0: \beta_3 = 0 \quad \text{versus} \quad H_a: \beta_3 \neq 0
\]

\[
\text{t}_{\text{stat}} = \frac{0.01532 - 0}{0.003396} = 4.55
\]

\[
\text{t}_{\text{crit}} = 1.96
\]

\( |\text{t}_{\text{stat}}| > |\text{t}_{\text{crit}}| \Rightarrow \text{Reject } H_0 \)

Experience is significant in determining the wage.
b. Perhaps you’ve heard the phrase, “I was trained in the school of hard knocks...it’s just as good as school”. Write down the hypothesis that states that education (educ) and experience (exper) have equal effects on the log wage. Also provide a two-sided alternative. (2 Points)

\[ H_0 : \beta_1 = \beta_3 \quad + 1 \]

\[ H_a : \beta_1 \neq \beta_3 \quad + 1 \]

If the state the hypothesis as \( \theta = \beta_1 - \beta_3 = 0 \) that \( \beta_3 \neq 0 \) also line

c. Please manipulate the regression equation in (a) so that your null hypothesis in (b) can be tested using a t-test. Show your work!! (8 Points)

\[ \log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{tenure} + \beta_3 \text{exper} + u \]

\[ = \beta_0 + (\theta + \beta_3) \text{educ} + \beta_2 \text{tenure} + \beta_3 \text{exper} + u \]

\[ + 1 \]

\[ = \beta_0 + (\theta + \beta_3) \text{educ} + \beta_2 \text{tenure} + \beta_3 (\text{educ} \times \text{exper}) + u \]

\[ + 2 \text{for some work} \]
Suppose that I adjust the specification in (a) as follows,

\[
\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{tenure} + \beta_3 \text{exper} + \beta_4 \text{exper}^2 + \beta_5 \text{age} + \beta_6 \text{age}^2 + u
\]

where \text{age} is the age of the respondent (in years).

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>26.2843766</td>
<td>6</td>
<td>4.38072944</td>
</tr>
<tr>
<td>Residual</td>
<td>139.371918</td>
<td>928</td>
<td>.150185256</td>
</tr>
<tr>
<td>Total</td>
<td>165.656294</td>
<td>934</td>
<td>.177362199</td>
</tr>
</tbody>
</table>

Number of obs = 935
F( 6, 928) = 29.17
Prob > F = 0.0000
R-squared = 0.1587
Adj R-squared = 0.1532
Root MSE = 0.38754

| lwage | Coef.    | Std. Err. | t     | P>|t|   | [99% Conf. Interval] |
|-------|----------|-----------|-------|-------|----------------------|
| educ  | 0.0703793| 0.0069319 | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| tenure| 0.0127327| 0.0026224 | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| exper | 0.0230826| 0.013667  | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| exper2| -0.0005216| 0.0006   | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| age   | 0.0069779| 0.1034685 | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| age2  | 0.000507 | 0.0015552 | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |
| cons  | 5.26504  | 1.697521  | xxxxxx| xxxxxx| xxxxxxxxxxxxxxxxxxxxxx |

d. Please construct a 99% confidence interval for the coefficient on tenure, \( \beta_2 \). (5 points)

\[
\begin{align*}
CI &= \hat{\beta}_2 \pm se(\hat{\beta}_2) \cdot t_{cr,1} + 1 \\
C_{cr,t} &= 2.576 \\
CI &= 0.0127 \pm 0.00262 \cdot (2.576) + 1 \\
\beta_{2C} &\in [0.0060, 0.0194] + 1 \\
&\text{for any work}
\end{align*}
\]
e. What is the p-value for the linear experience term (exper)? Please draw the distribution under the null and compute the two-sided p-value (please give the range and an approximate value). (5 points)

\[ t_{\text{stat}} = \frac{0.02306 - 0}{0.01367} = 1.688 \]

\[ 0.05 \leq P \leq 0.10 \]

\[ \text{If they have one, OK } P \approx 0.09 \]

\[ +1 \text{ for reasonable work} \]

f. Please test the hypothesis that adding age, age², and exper² makes no difference in predicting the log wage. That is, please test whether these variables are jointly insignificant. Do this at the 95% level, stating the null and alternative hypotheses. (10 points)

\[ F_{\text{stat}} = \frac{\frac{SSR_{R} - SSR_{UR}}{\# \text{ restrict}}}{\frac{SSR_{UR}}{(n-k-1) \text{ re}} = \frac{139.96 - 139.37}{3} \]

\[ = \frac{139.37}{928} = 1.31 \]

\[ F_{\text{crit}} = 3.07 \]

\[ F_{\text{stat}} < F_{\text{crit}} + 2 \]

\[ \text{Fail to reject } H_0 \]

\[ H_0: \beta_4 = 0, \beta_5 = 0, \beta_6 = 0 \]

\[ H_A: H_0 \text{ not true} \]

\[ +2 \]
g. Professor Spearot will be 29 in January. He feels old, though is shamelessly hoping that his wage makes up for it. Assuming that he gains no additional experience and no additional tenure, what is the predicted effect on the wage in going from age 28 to age 29? Interpret briefly. (5 points)

\[
\frac{d\log(wage)}{dwage} = \beta_3 + 2\beta_4 \text{age} \\
= 0.00698 + 2 \times 0.0000507 \times 28 \\
= 0.0098171 + 3
\]

Wage increases by 0.0098171% + 3

h. At what experience level is the wage maximized? Show your work! (5 points)

\[
F(\text{exp}) = \beta_3 \text{exp} + \beta_4 \text{exp}^2
\]

\[
\frac{dF}{d\text{exp}} = \beta_3 + 2\beta_4 \text{exp} = 0 + 2
\]

\[
\text{exp} = -\frac{\beta_3}{2\beta_4} = \frac{0.0231}{2 \times (-0.000522)} = 22.12
\]

\boxed{\text{exp} = 22.12 + 2}
3. Suppose that I run the following regression predicting the effects of candidate expenditures and other factors on election outcomes:

\[
\text{vote}_A = \beta_0 + \beta_1 \text{lexpend}_A + \beta_2 \text{lexpend}_B + \beta_3 \text{prtyst}_A + u
\]

In the regression equation, lexpendA is the log expenditures of candidate A, lexpendB is the log expenditures of candidate B, and prtystr is the relative strength of party A. The results from running this regression are below:

```
. regress votea lexenda lexemb prtystra
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 173</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>38405.1089</td>
<td>3</td>
<td>12801.703</td>
<td>F( 3, 169) = 215.23</td>
</tr>
<tr>
<td>Residual</td>
<td>10052.1396</td>
<td>169</td>
<td>59.4801161</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>48457.2486</td>
<td>172</td>
<td>281.728189</td>
<td>R-squared = 0.7926</td>
</tr>
</tbody>
</table>

Adj R-squared = 0.7889
Root MSE = 7.7123

| votea | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------|-------|-----------|-------|-----|---------------------|
| lexenda | 6.083316 | .38215 | 15.92 | 0.000 | 5.328914 | 6.837719 |
| lexemb  | -6.615417 | .3788203 | -17.46 | 0.000 | -7.363247 | -5.867588 |
| prtystra | .1519574 | .0620181 | 2.45 | 0.015 | .0295274 | .2743873 |
| cons   | 45.07893 | 3.926305 | 11.48 | 0.000 | 37.32801 | 52.82985 |

a. Are the variables in the model (lexenda, lexemb, and prtystra) a significant determinant of the votes candidate A receives? Please test this hypothesis at the 95% level. Write the null hypothesis, the alternative, and briefly interpret the result. (5 Points)

\[ H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0 \]

\[ H_A: H_0 \text{ not true} \]

\[ F_{\text{stat}} = 215.23 \]

\[ F_{\text{crit}} = 2.60 \]

\[ F_{\text{stat}} > F_{\text{crit}} \]

Reject the null, \( H_0 \) contains poor restrictions.
b. Suppose that Candidate A is Mike Gravel and Candidate B is Mitt Romney. I want to predict the outcome of this hypothetical race, and produce confidence intervals for this prediction. Please write an equation for the prediction if lexpendA=1 (Mike), lexpendB=10 (Mitt), and prtysta=50 (citizens hate both parties equally!!). Derive a new estimating equation to generate the prediction and its standard error. Please also write the necessary commands to generate any new variables in STATA. (10 Points)

\[ \theta = \beta_0 + \beta_1 \cdot \text{lexpendA} + \beta_2 \cdot \text{lexpendB} + \beta_3 \cdot \text{prtystra} \]

\[ \beta_0 = \theta - \beta_1 \cdot \text{lexpendA} - \beta_2 \cdot \text{lexpendB} - \beta_3 \cdot \text{prtystra} \]

\[ \text{VoteA} = \theta - \beta_1 \cdot \text{lexpendA} - \beta_2 \cdot \text{lexpendB} - \beta_3 \cdot \text{prtystra} + \beta_1 \cdot \text{lexpendA} + \beta_2 \cdot \text{lexpendB} + \beta_3 \cdot \text{prtystra} + 2 \]

\[ \text{VoteA} = \theta + \beta_1 \cdot \text{lexpendA} - 1 + \beta_2 \cdot \text{lexpendB} - 10 + \beta_3 \cdot \text{prtystra} - 50 \]

\[ \text{VoteA} = \theta + \beta_1 \cdot \text{lexpendA} - 1 + \beta_2 \cdot \text{lexpendB} - 10 + \beta_3 \cdot \text{prtystra} - 50 + 3 \]

\[ \text{generate lexpendA}_1 = \text{lexpendA} - 1 \]
\[ \text{generate lexpendB}_1 = \text{lexpendB} - 10 \]
\[ \text{generate prtystra}_1 = \text{prtystra} - 50 \]
\[ \text{generate lexpendA}_2 = \text{lexpendA} - 1 + 1 \]
\[ \text{generate lexpendB}_2 = \text{lexpendB} - 10 + 1 \]
\[ \text{generate prtystra}_2 = \text{prtystra} - 50 + 1 \]
c. Suppose that as campaign expenditures rise, unobserved factors affecting voting outcomes tend to become more variable. What kind of problem is this? What should be done about it? (5 Points)

Heteroskedasticity + 3

Use Robust standard errors + 2

d. Suppose that candidate A spends more money because he/she has better ideas, and better ideas get you more votes. What assumption is violated in our current model? In what direction is $\beta_1$ biased? (5 Points)

$E(u|x) \neq 0$ + 2

$\beta_1$ will be biased upward + 3
Extra Credit:

Look back at the full model in problem #2. Age discrimination is perceived to be commonplace in American society. Age discrimination occurs if there exists an age above which wages go down, independent of other attributes (education, experience, tenure). Using the results in Problem #2, briefly discuss whether there is evidence of age discrimination. (2 points)

There is no positive age above which wage decreases.

No discrimination.

Could also solve for age. +2

Helpful formulas

\[
\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2 \\
\hat{\sigma}_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) \\
\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \\
\hat{\beta}_0 = \hat{\mu}_y - \hat{\beta}_1 \hat{\mu}_x \\
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2} \\
R^2 = 1 - \frac{SSR}{SST} \\
SSR = \sum_{i=1}^{n} (\hat{y}_i)^2 \\
SST = \sum_{i=1}^{n} (y_i - \hat{\mu}_y)^2 \\
Adj R^2 = 1 - \frac{SSR}{SST} - \frac{n-1}{n-k-1} \\
Fstat = \frac{SSR_{R} - SSR_{UIR}}{SSR_{UIR}} \frac{n-k-1}{q}