**Assignment 3**

Applied Econometrics and Time Series Analysis

Group Members

Jyotsna Nagaprasad Ambekar JXA210018

Pooja Akkaladevi PXA210024

Karthik Ram - KCR210004

5.1

1. True. The Classical linear regression model assumes the error term **Ui** to be normally distributed. The t-test is based on variables with a normal distribution and {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>^</mo></mover></mstyle></math>"}
   stack beta subscript 1 with hat on top and {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover></mstyle></math>"}
   stack beta subscript 2 with hat on topare linear combinations of the error term **Ui**
2. True, OLS estimators are unbiased as long as **E(Ui) = 0**
3. True,  
   Without the error term the OLS regression model can be represented as

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>Y</mi><mi>i</mi></msub><mo>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover><mo>&#xA0;</mo><msub><mi>X</mi><mi>i</mi></msub><mo>&#xA0;</mo><mo>+</mo><mo>&#xA0;</mo><mover><msub><mi>u</mi><mi>i</mi></msub><mo>^</mo></mover></mstyle></math>"}
Y subscript i space equals space stack beta subscript 2 with hat on top space X subscript i space plus space stack u subscript i with hat on top

The equation passes through the origin.

The normal equation that is obtained on minimizing the sum of squared error term over the observations does not include the intercept.  
  
Hence, in the regression, the model with no intercept {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mstyle displaystyle=\"false\"><munderover><mrow><mo>&#x2211;</mo><msub><mover><mi>u</mi><mo>^</mo></mover><mi>i</mi></msub><mo>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><mn>0</mn></mrow><mrow/><mrow/></munderover></mstyle></mstyle></math>"}
stack sum u with hat on top subscript i space equals space 0 with blank below and blank on top will not hold

1. False, The p-value and the size of the test always do not mean the same thing.  
     
   The size of a test is the probability of falsely rejecting the null hypothesis.   
   A test is said to have a significance level 𝛼 if its size is less than or equal to 𝛼.  
     
   The p-value is the probability of obtaining a test statistic at least as extreme as the one we observed from the sample if the null hypothesis were true. It may be less than or greater than the type I error.
2. **True**With the intercept the two equations derived from minimizing the sum of the squared error term over all the observations holds.
3. False, the Null hypothesis is only assumed to be valid until proven otherwise. Failing to reject the Null hypothesis doesn't make it true
4. False, The variance of the beta hat also depends on **{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><munderover><mo>&#x2211;</mo><mrow/><mrow/></munderover><msubsup><mi>x</mi><mi>i</mi><mn>2</mn></msubsup></mstyle></math>"}
   sum from blank to blank of x subscript i superscript 2**
5. False, Unconditional, and conditional mean of random variables is the same if the variables are independent
6. True, this can be obtained from the following equation

Text, letter

Description automatically generated

1. True, this is evident from the following equations. If **{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover></mstyle></math>"}
   stack beta subscript 2 with hat on top** is zero then **{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mstyle displaystyle=\"false\"><munderover><mrow><mo>&#x2211;</mo><msubsup><mi>y</mi><mi>i</mi><mn>2</mn></msubsup><mo>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><munderover><mrow><mo>&#x2211;</mo><msubsup><mover><mi>u</mi><mo>^</mo></mover><mi>i</mi><mn>2</mn></msubsup></mrow><mrow/><mrow/></munderover></mrow><mrow/><mrow/></munderover></mstyle></mstyle></math>"}
   stack sum y subscript i superscript 2 space equals space stack sum u with hat on top subscript i superscript 2 with blank below and blank on top with blank below and blank on top**

**Text, letter, whiteboard

Description automatically generated**

5.3

1. Missing Values
   1. **{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover></mstyle></math>"}
      stack beta subscript 2 with hat on top =** 0.069581
   2. t value - intercept = -0.0165
2. Mean hourly wage increases by 0.724097 dollars for every additional year in education
3. Hypothsis test to check if education has no effect whatsoever on wages is given by

Null hypothesis {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
beta subscript 2 = 0

Alternate hypothesis {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
beta subscript 2 not equal to zero

Since we are not aware of the population standard deviation of the wages, and have around 13 samples, t - test would suffice

The t value for {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
beta subscript 2 is 10.4065 and the probability of obtaining that value is very small.

The p value is 4.958e-07

Therefore, the null hypothesis that education has no effect on wage can be rejected.

1. X Performing ANOVA on the linear regression yields

RSS = 9.6928

ESS = 95.4255

Numerator DF = 1

Denominator DF = 11

F = 108.295

The p value of the F distributing for the null hypothesis that there is no relationship between the wage and income is 0. Hence null hypothesis is rejected.

1. If the value of {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msup><mi>r</mi><mn>2</mn></msup></mstyle></math>"}
   r squared was missing in the table the following formula can be used.

**{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msup><mi>r</mi><mn>2</mn></msup><mo>&#xA0;</mo><mo>=</mo><mo>&#xA0;</mo><mfrac><msup><mi>t</mi><mn>2</mn></msup><mrow><msup><mi>t</mi><mn>2</mn></msup><mo>&#xA0;</mo><mo>+</mo><mpadded lspace=\"+1px\"><mo>&#xA0;</mo><mfenced><mrow><mi>n</mi><mo>-</mo><mn>2</mn></mrow></mfenced><mo>&#xA0;</mo></mpadded></mrow></mfrac></mstyle></math>"}
r squared space equals space fraction numerator t squared over denominator t squared space plus space open parentheses n minus 2 close parentheses space end fraction**

t = 10.4065, n = 13

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msup><mi>r</mi><mn>2</mn></msup></mstyle></math>"}
r squared = 0.9077

5.9

1. Plotting the data and eyeballing the regression line

Chart, scatter chart

Description automatically generated

1. Estimates of parameters
   1. {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>1</mn></msub></mstyle></math>"}
      beta subscript 1 = 12129.3710
   2. {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
      beta subscript 2 = 3.3076
   3. {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>1</mn></msub></mstyle></math>"}
      beta subscript 1 SE = 1197.3508
   4. {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
      beta subscript 2 SE = 0.3117
   5. {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msup><mi>r</mi><mn>2</mn></msup></mstyle></math>"}
      r squared = 0.69059
   6. RSS = 264825250
   7. ESS = 608555015
2. Since **{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover></mstyle></math>"}
   stack beta subscript 2 with hat on top** is positive. Which means increase in spending leads to increase in salary for the teacher. Increase in one dollar on per pupil increases average salary to the teachers by 3.3076
3. Interval estimate can be calculated from

{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
beta subscript 2 ![{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mo>&#xB1;</mo></mstyle></math>"}
plus-or-minus](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAFoAAAA7CAYAAAD7AJ39AAAAAXNSR0IArs4c6QAAAARnQU1BAACxjwv8YQUAAAAJcEhZcwAADsQAAA7EAZUrDhsAAADXSURBVHhe7dqxasJQAIbRKOJ7FXwGn8y9Q3cHl0JfTi/ErXUzXyA9B35y548MIcm0suPY/cUuY5uxf15ZmNARoSNCR4SOCB0ROiJ0ROiI0BGhI0JHhI4IHRE6InRE6IjQKzuN/fV5aYu7ji3OHR0ROiJ0ROiI0BGh/wl/KvFeQkeEjggdEToidEToiNARoSNCR4SOCB0ROiJ0ROiI0BGhI0JHhI7snte1HMa+5uMvP2Of8xEA2Cw/or+Z5+iI0BGhI0JHhI68etfxMfY9HzfvNnaej0uZpgcpN4b7lNSI1gAAAABJRU5ErkJggg==) t(0.975,49)\*SE({"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
beta subscript 2)

= (2.6812004 , 3.9339696)

Since 3.0 lies between the interval estimate, it doesnt reject the null hypothesis

1. SALARY = {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>1</mn></msub></mstyle></math>"}
   beta subscript 1 + {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><msub><mi>&#x3B2;</mi><mn>2</mn></msub></mstyle></math>"}
   beta subscript 2\*SPENDING

= 12129.3710 + 3.3076 \* 5000

= 28667.371

The mean and individual forecast values are : 28667.371

Standard error of mean forecast value **Eq(5.10.2)** = 520.5117

**A picture containing text, clock, gauge

Description automatically generated**

And the standard error of the individual forecast using **Eq(5.10.6)** = 2382.337

A picture containing text, clock, gauge

Description automatically generated

Mean prediction with 95% confidence interval : [27626,29708]

Individual prediction with 95% confidence interval : [23902,33432]

6.1 The OLS formula to estimate the intercept is

**{"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>1</mn></msub><mo>^</mo></mover></mstyle></math>"}
stack beta subscript 1 with hat on top =** mean if the regressand - {"mathml":"<math style=\"font-family:stix;font-size:16px;\" xmlns=\"http://www.w3.org/1998/Math/MathML\"><mstyle mathsize=\"16px\"><mover><msub><mi>&#x3B2;</mi><mn>2</mn></msub><mo>^</mo></mover></mstyle></math>"}
stack beta subscript 2 with hat on top mean of the regressor

Change the alternate names of regressor and regressor  
 But if x and y are in the deviation format, both the terms are always zero.

Therefore, in this case, the estimated intercept is always zero.

**6.3**

1. The model is linear parameters, hence it is a linear regression model
2. The model can be estimated using the ordinary least squares (OLS) method. For simplicity, 1/ Yi and 1/Xi can be considered as Y’ and X’.
3. As Xi tends to infinity, the 1/Xi terms tend to zero, hence Yi tends to Beta1
4. The model can explain the relationship between low household consumption and income increase.

**6.4)** Given the log-linear model:

**Case1:**

When = 1

The equation becomes ln Yi =



Graph Y vs X

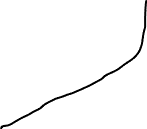
Y

X

**Case2:**

When > 1

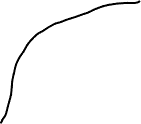
Y



X

**Case3:**

When < 1



**6.6) a)** The First equation can be written as

ln(w1Yi) = α1 + α2ln(w2Xi) + u\*i

lnw1 + lnYi = α1 + α2ln(w2) + α2ln(Xi) + ui\*

as w1  and w2 are constant the terms lnw1 and lnw2  are also constants

so bringing the constant terms together we get

lnYi = (-lnw1+ α1 + α2ln(w2)) + α2ln(Xi) + ui\*

replacing the constants with k we get

lnYi = k + α2ln(Xi) + ui\*

Comparing the above equation with the second model which is lnYi = β1 + β2lnXi+ ui

Only the intercepts are different, but the slope coefficients remain same.

And the standard errors of the slope coefficients and intercepts are also same.

b) The r2is same between these two models

**6.14)**

**Text

Description automatically generated**

From the above output from the R code

log( = -0.4526 + 1.3338logW

the standard error of elasticity (β2) is 0.4467

The hypothesis statements to verify that the elasticity 1.3338 is not significantly different from 1 are as follows

H0 : β2 = 1

Ha : β2 ≠ 1

the test statistic t0  = (1.3338-1)/0.4467 = 0.7472

For 95% confidence interval and for the 13 degrees of freedom (total sample points – number of parameters that we have to estimate= 15-2=13) the critical t value is t0.025,13 = 2.16

Hence 0.7472 lies in the acceptance region which is (-2.16,2.16) so need to accept the null hypothesis

Hence we can conclude that the elasticity β2  is not significantly different from 1.

**7.1)** R code file attached in the assignment zip file.

**Text, letter

Description automatically generated**

From the above output from the R code

The estimates of the intercepts and coefficients of the three equations are the following

1

1. No, α2 is not equal to β2 because the true model is represented by the third equation where X2 an X3 both effect the target variable Y. Hence is the biased estimator of β2
2. No,is not equal to β3 because the true model is represented by the third equation where X2 an X3 both effect the target variable Y. Hence is the biased estimator of β3

**7.8)** Given,

The equation : Yi = β1 + β2 Education i + β3 Years of experience + ui

If we leave out the years of experience variable the coefficient of education will be biased (which indicates that the Yi only depends on education which isn’t the case). Such type of bias is referred to as omitted variable bias. The standard error, the residual sum of squares,R2 will be affected as a result of this omission.

Question 7.14

1. The ln ui should have normal distribution.

Text, letter

Description automatically generated

By assuming the normal distribution, we can apply the classical normal linear regression model (CNLRM). After estimating the Cobb-Douglas model, we can take the residuals and perform normality tests like Anderson test or Shapiro test etc.

1. If ln ui should follow normal distribution with zero mean and constant variance, ui should follow log-normal distribution with mean e to the power of sigma squared end exponent and variance e to the power of sigma squared end exponent open parentheses e to the power of sigma squared end exponent minus 1 close parentheses

Question 7.15

Regression passing through origin –

Diagram, schematic

Description automatically generated

1. Estimating the unknows will be done by computing the normal equations.

sum Y subscript i X subscript 2 i end subscript equals beta subscript 2 sum X subscript 2 i end subscript superscript 2 space plus space beta subscript 3 sum for blank of X subscript 2 i end subscript space X subscript 3 i end subscript

stack sum Y subscript i X subscript 3 i end subscript with blank below space equals space beta subscript 2 sum for blank of X subscript 2 i end subscript X subscript 3 i end subscript space plus space beta subscript 3 sum for blank of X subscript 3 i end subscript squared

1. No, sum u with hat on top semicolonwill not be zero for this model. The prediction is not always same as that of actual. There is always some error in the model. Hence, we will not observe it to be zero.
2. Yes, sum u with hat on top subscript i X subscript 2 i end subscript equals sum stack u subscript i with hat on top X subscript 3 i end subscript space equals space 0 holds good for this model.
3. We can use this model anytime when the problem has regression line passing through origin.
4. We can generalize these equations to k-variable model.

Question 7.18

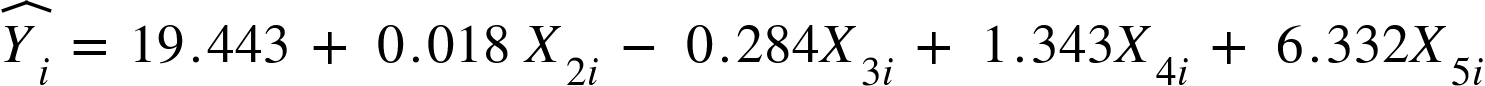
1. The multiple linear regression model was developed in R. Summary of the model is

Text

Description automatically generated

The summary gives us coefficients & standard errors for all the independent variables.

The equation for the model is



Standard errors are 3.406, 0.006, 0.457, 0.259, 3.029 respectively.

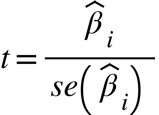
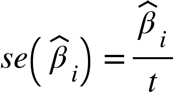
Multiple R squared = 0.9776

Adjusted R squared = 0.9716

1. All slope coefficients are expected to be positive except coefficient for US military sales. All other variables have statistical significance at the 5% level.
2. Federal outlays and some time series variable which shows the progress over a period of time may be included in the model.

Question 7.20

1. A 1 unit increase in natural logarithm value of adult male unemployment rate leads to 0.34 units decrease in natural logarithm value of quit rate. A 1 unit increase in natural logarithm value of % of employees younger than 25 leads to 1.22 units increase in natural logarithm value of quit rate. the same 1 unit increase in natural logarithm value of ratio of manufacturing employment in quarter (t-1) and quarter (t-4) leads to 1.22 units increase in natural logarithm value of quit rate. 1 unit increase in % of women employees is leading to 0.80 units increase in quit rate. Over the time period of study, quit rate declined at the rate of 0.53% per year.
2. Quit rate & unemployment rate are expected to be negatively related
3. As more people younger than 25 are joining the manufacturing unit, there is more quit rate because there is more turnover among the younger employees
4. The decline rate is 0.54% As there is better environment to work and better benefits are given to the employees over time period of the study, there might have been decrease in the quit rate.
5. R with bar on top squared is not so low. We can consider 0.5370 to be moderate. However R with bar on top squared value below 0.3 can be considered too low.
6. Since t values are given, we can easily calculate standard errors. Under null hypothesis that beta subscript i = 0, we have

 this implies 

Question 7.23

1. The variables are changed as per the question and the model is developed with those variables. The summary of the model is –

Text

Description automatically generated

The coefficients suggest that natural logarithmic value of education increases by 1 unit, then the wage also increases by 0.003934 units in natural logarithm scale.

1. The model has multi collinearity due to the logarithmic properties. log(x)2 = 2 log(x). Hence the same is observed in the education variable.