**Question One:**

1. Open the file PRICEINCOME

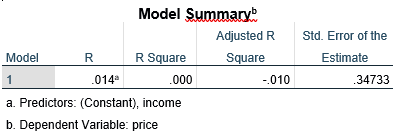
**This file contains data on the price of a good and the income of its consumers**

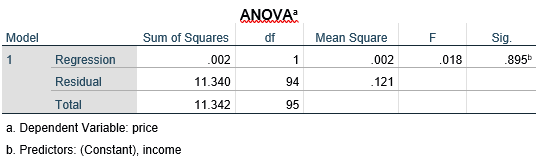
1. Regress PRICE on INC and report your results (2 marks) ***Report the sample regression function and R squared (write them out, not insert the output)***

Price = .827 + .014 x Income

R2= .000

The significance value of 0.895 (greater than 0.05) indicates that the regression model does not significantly predict the price. This is also indicated by the R squared value of .000 which indicates that about 0% of the variation in price is explained by income.

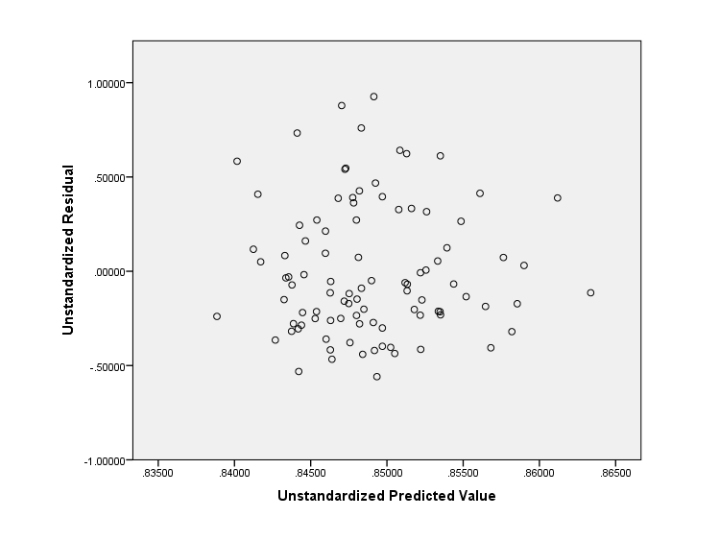




1. Does the model in a) exhibit heteroscedasticity? Explain your answer

(4 marks) **Quote appropriate regression output, carry out test and give result of test with explanation**

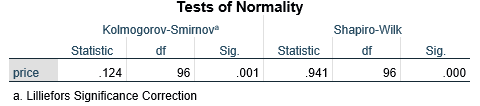
Based on the below scatterplot, the model does exhibit heteroscedasticity. This was interpreted as such because the points in the above graph are not scattered equally, as would be seen in homoscedasticity.

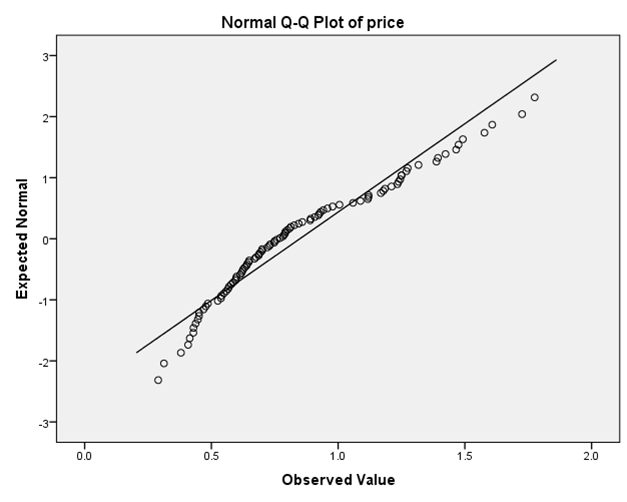
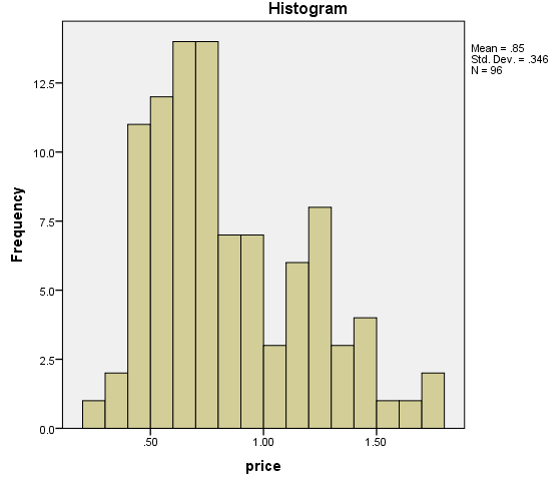
****

1. Are the variables PRICE and INC normally distributed? Justify your answer (4 marks) **Provide appropriate graphs/diagrams** **(can cut/paste) and explain your answer**

Price - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is not normally distributed. This is because p = 0.001, and thus p < 0.05, indicating that p is significant, and thus that the data is not normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is not distributed around a central value; the data has a right skew. Furthermore, in the Q-Q plot, the data points stray from the line of normality, indicating again that the data is not normal.

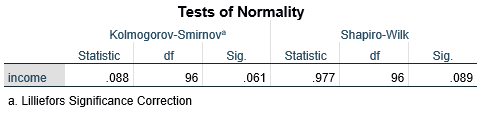
Alpha value = 0.05.

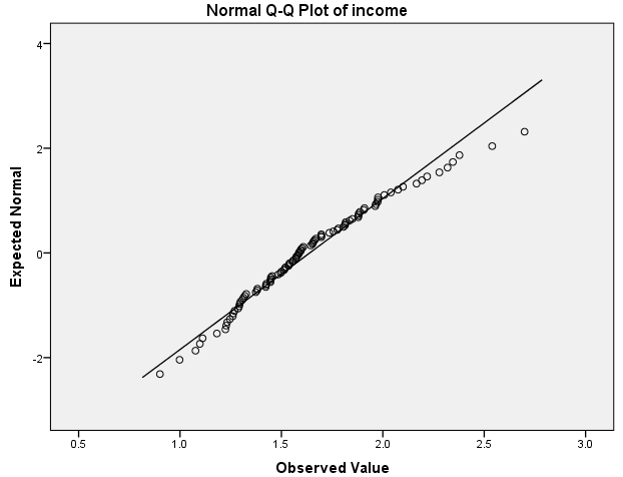
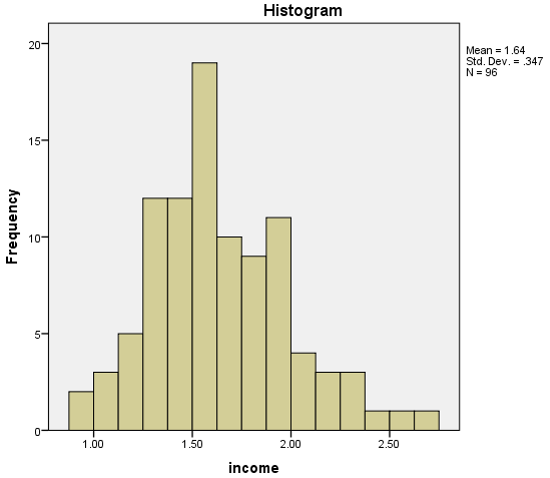
****

****

Income - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is normally distributed. This is because p = 0.061, and thus p > 0.05, indicating that p is insignificant, and thus that the data is normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is distributed around a central value. Furthermore, in the Q-Q plot, the data points remain relatively in line with the line of normality, indicating again that the data is normal.

Alpha value = 0.05.

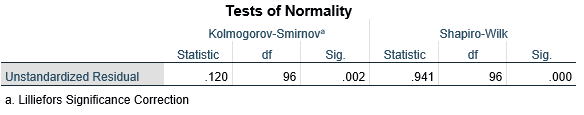


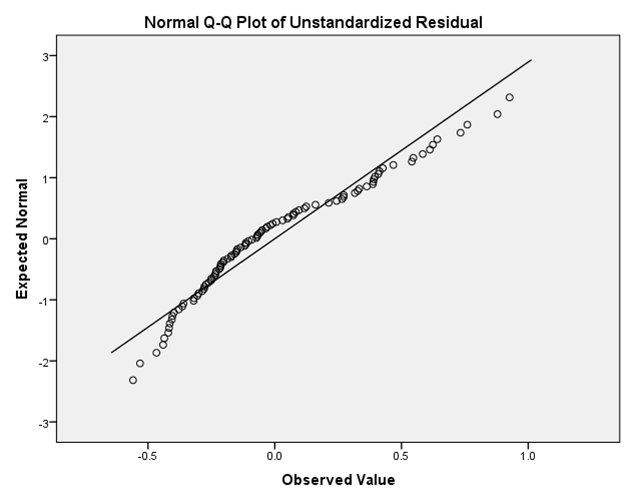
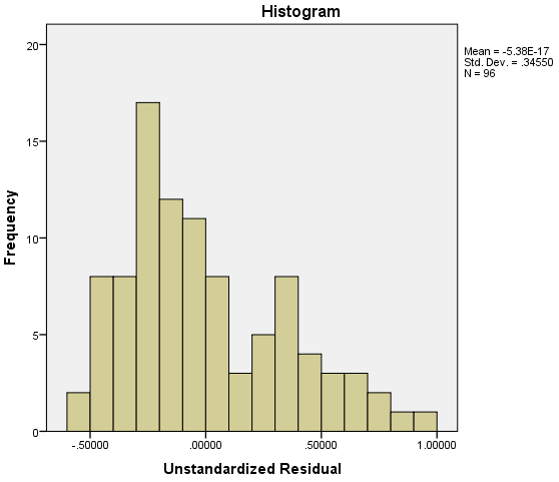


1. Are the residuals in your model normally distributed? Justify your answer (3 marks) **Provide appropriate graphs/diagrams (can cut/paste) and explain your answer**

Unstandardized Residual - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is not normally distributed. This is because p = 0.002, and thus p < 0.05, indicating that p is significant, and thus that the data is not normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is not distributed around a central value; the data has a right skew. Furthermore, in the Q-Q plot, the data points stray from the line of normality, indicating again that the data is not normal.

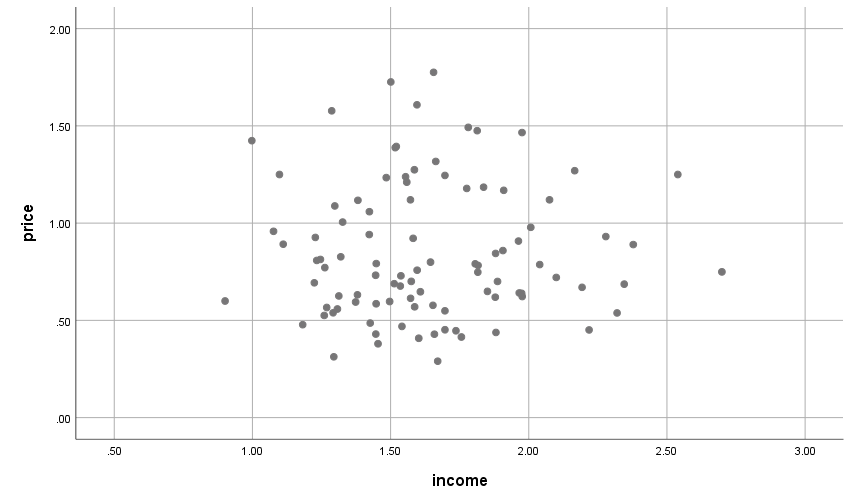
Alpha value = 0.05.

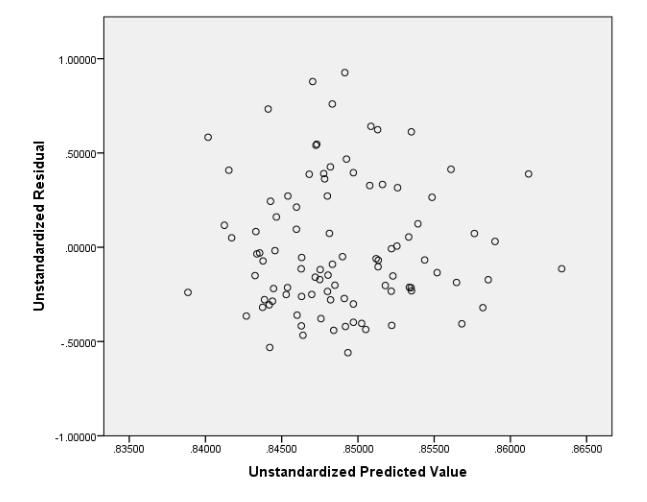




1. Is the model in a) mis-specified? Explain your answer (3 marks) **Provide appropriate graphs/diagrams (can cut/paste) and explain your answer**

The regression model is mis-specified. When the variables income and price are plotted, they do not appear to have a linear relationship (please see below in a plot made on spss). This can also be seen in the plot of residuals and predicted values, also pasted below. As the variables do not follow a linear relationship, this violates a key assumption of linear regression, and thus our model is mis-specified.



****

Total Marks: 16 marks

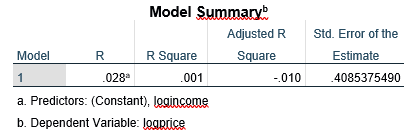
**Question 2:**

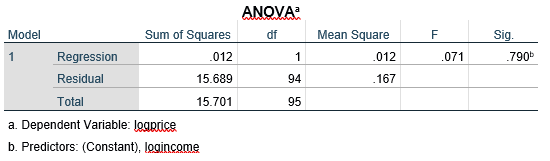
1. Open the file PRICEINCOME
2. Regress LOGPRICE on LOGINC and report your results (2 marks) ***Report the sample regression function and R squared (write them out, not insert the output)***

LogPrice = (-0.06)LogIncome + 0.192

R2= .001

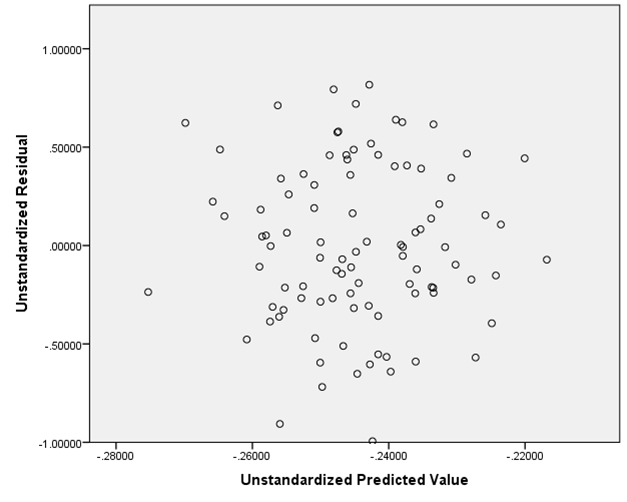
The significance value of 0.795 (greater than 0.05) indicates that the regression model does not significantly predict the price. This is also indicated by the R squared value of .001 which indicates that about 0.1% of the variation in price is explained by income.





1. Does the model in a) exhibit heteroscedasticity? Explain your answer (4 marks) **Quote appropriate regression output, carry out test and give result of test with explanation**

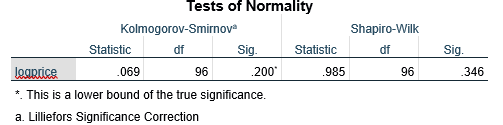
Based on the above scatterplot, the model does exhibit heteroscedasticity. This was interpreted as such because the points in the above graph are not scattered equally, as would be seen in homoscedasticity.

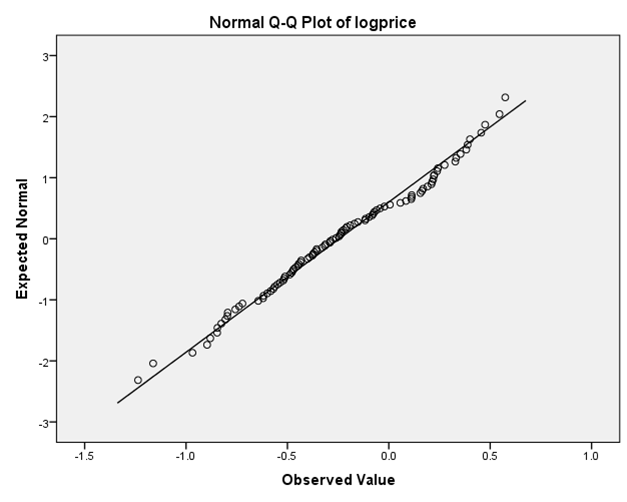
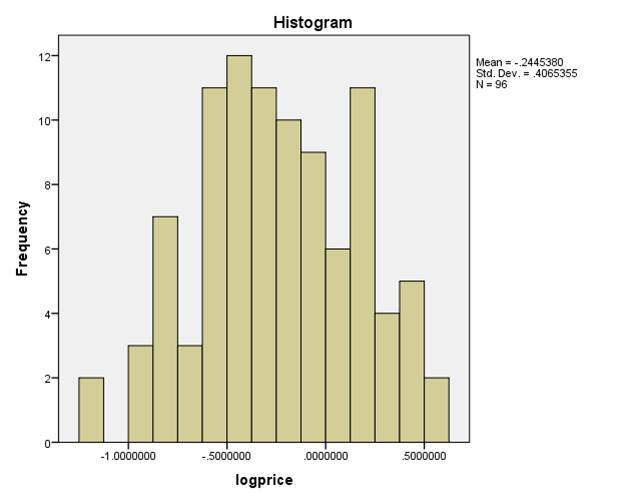
******

1. Are the variables LOGPRICE and LOGINC normally distributed? Justify your answer (4 marks) **Provide appropriate graphs/diagrams (can cut/paste) and explain your answer**

Logprice - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is normally distributed. This is because p = 0.200, and thus p > 0.05, indicating that p is insignificant, and thus that the data is  normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is distributed around a central value. Furthermore, in the Q-Q plot, the data points remain consistent with the line of normality, indicating again that the data is normal.

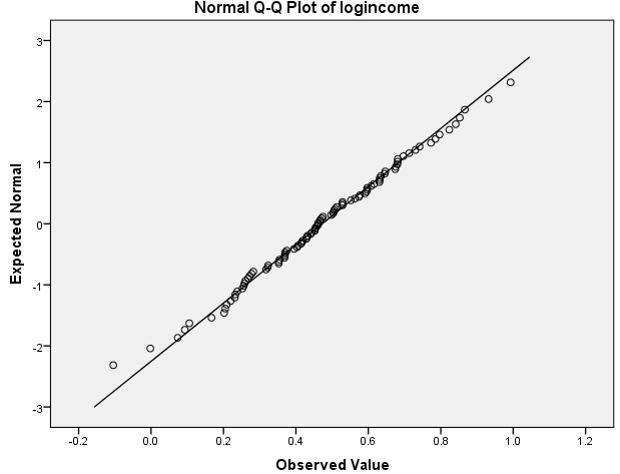
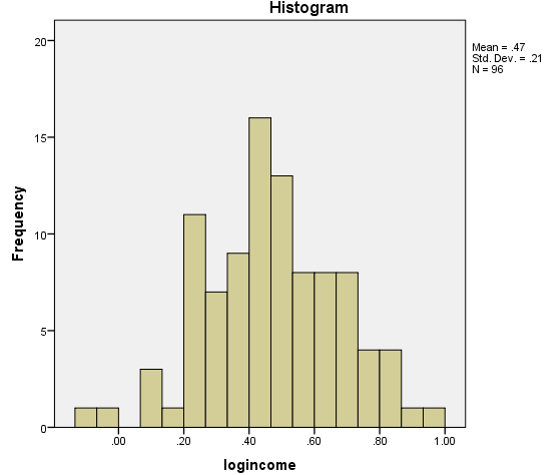
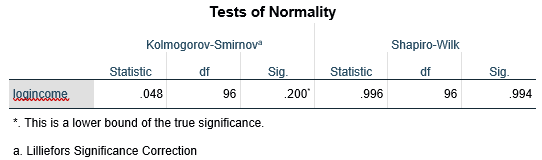
Alpha value = 0.05.





Logincome - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is normally distributed. This is because p = 0.200, and thus p > 0.05, indicating that p is insignificant, and thus that the data is  normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is distributed around a central value. Furthermore, in the Q-Q plot, the data points remain consistent with the line of normality, indicating again that the data is normal.

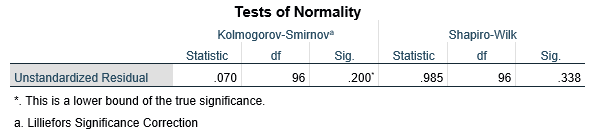
Alpha value = 0.05.

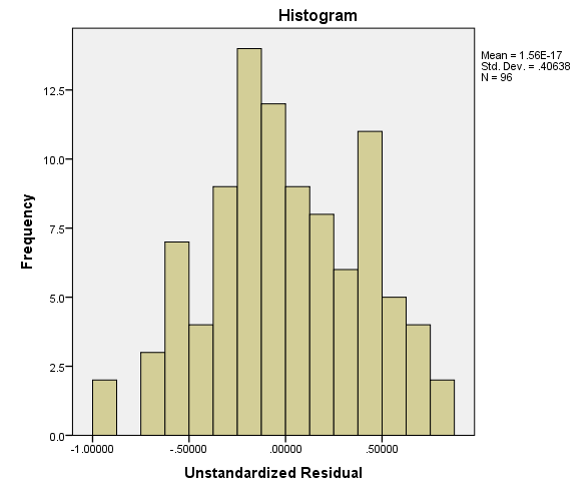
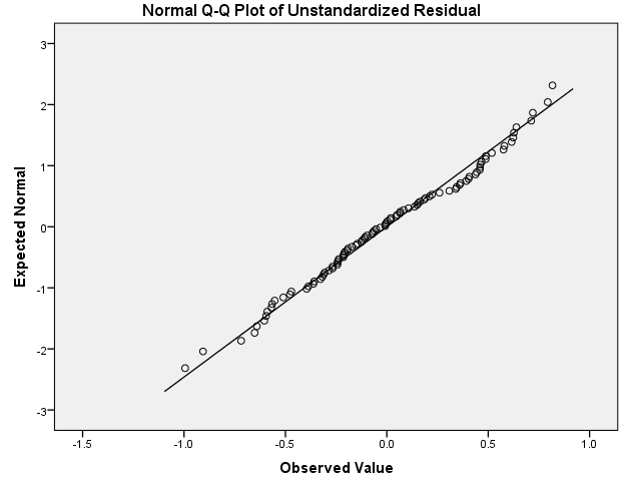


1. Are the residuals in your model normally distributed? Justify your answer (3 marks) **Provide appropriate graphs/diagrams (can cut/paste) and explain your answer**

Unstandardized Residual - Looking at the Tests of Normality, specifically the Kolmonogorov-Smirnov test (as our sample size is greater than 50), the p value indicates that the data is normally distributed. This is because p = 0.200, and thus p > 0.05, indicating that p is insignificant, and thus that the data is normally distributed. This may also be indicated by looking at the histogram and Q-Q plot. Looking at the distribution of the data in the histogram, it is evident that the data is distributed around a central value. Furthermore, in the Q-Q plot, the data points remain in line with the line of normality, indicating again that the data is normal.

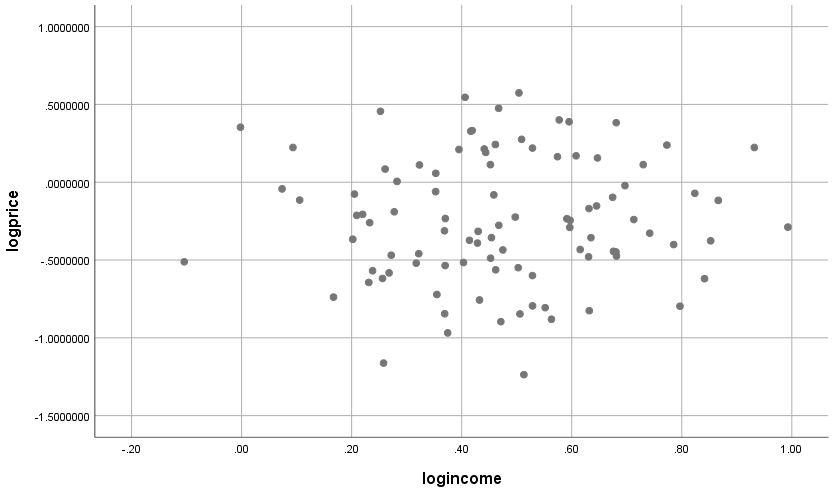
Alpha value = 0.05.

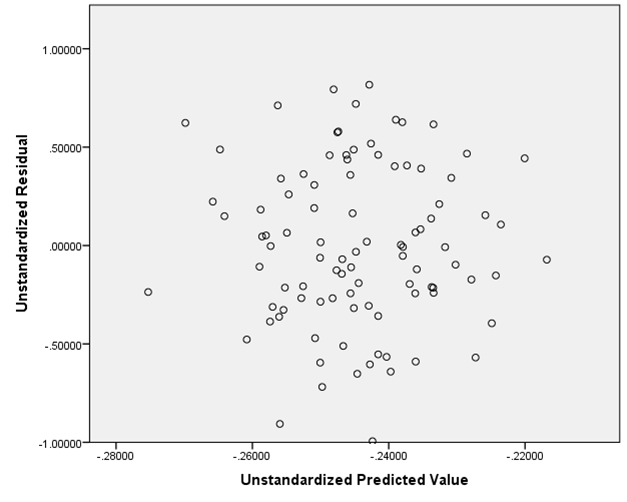


1. the model in a) mis-specified? Explain your answer (3 marks) **Provide appropriate graphs/diagrams (can cut/paste) and explain your answer**

The regression model is mis-specified. When the variables Logincome and logprice are plotted, they do not appear to have a linear relationship (please see below in a plot made on spss). This can also be seen in the plot of residuals and predicted values, also pasted below. As the variables do not follow a linear relationship, this violates a key assumption of linear regression, and thus our model is mis-specified.





Total Marks: 16 marks

**Question 3:**

1. Open the file SALARYASSESSMENT

**This file contains data on the salary of CEOs and performance indicators of their firms**

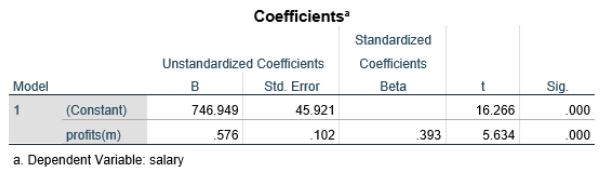
1. Regress SALARY on PROFITS and interpret your estimate of the coefficient on PROFITS  (1 mark) **Say what the number means for the estimated relationship between salary and profits.**

Salary = (.576)Profits + 746.949

Profits = (Salary - 746.969)/(0.576) = (salary-746.969)(1.74)

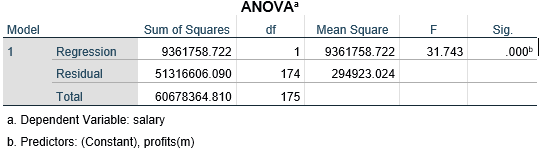
Profits=1.74(salary)-1299.73

The regression coefficient is 0.576. As it is a positive coefficient, that indicates a positive correlation between salary and profits, meaning that if one increases, the other does as well. Based on the coefficient, every increase of one in profits corresponds to an average increase of 0.576 in salary. For every increase of 1 in Salary there will be a 1.74 increase in profit on average.



1. Test the hypothesis that SALARY is not related to PROFITS (2 marks) **Specify the hypothese and your decision process clearly.**

Salary is significantly related to Profit. This is because p = 0.000, and thus p < 0.05, indicating that p is significant and therefore changes in salary is significantly related to changes in profits.



1. Regress SALARY on PROFITS and SALES. Interpret your estimate of the coefficients on PROFITS and SALES (3 marks) **Say what the numbers mean for the estimated relationship between salary and profits.**

y = b1x1 + b2x2 + c.

salary = b1(profits) + b2(sales) + constant.

salary = 0.364(profits) + 0.017(sales) + 728.98

Rearranging for profits:

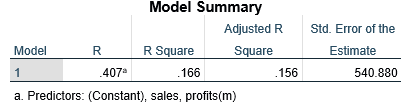
(salary - 0.017(sales) -728.98)/0.364 = 0.364(profits)/0.364

profits = 2.75(salary - 0.017(sales) - 728.98)

profits = 2.75(salary) - 0.047(sales) - 2004.695

1. How much of the squared variation in SALARY across students is explained by PROFITS and SALES? (2 marks) **State and interpret the appropriate figure from the regression output.**

The R square value of .166 indicates that about 16.6% of the variation in salary is explained by Sales and Profits.

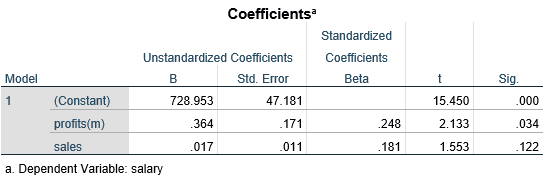


1. Why is the estimate of the coefficient on PROFITS different in a) and c)? (2 marks) **Use your knowledge of simple regression versus multiple regression here**.

In question a) we are only considering 2 variables; The independent variable Profit and the dependant variable salary. In question c) we are conducting a multiple regression as we consider another independent variable – sales. In question a) there is a single relationship between Profit and Salary, while in question c) we are looking at multiple relationships between profit, sales and salary, thus giving us different estimates on the coefficient Profit.

1. Using your results, what advice would you give to CEOs about boosting their salary? (4 marks) **This is asking you to interpret regression coefficients**

In order to boost salary, I would recommend that the CEO focuses on increasing profit. My interpretation of the coefficients of variables profits and sales suggests that if profit is increased by 1 there will be a 0.364 or 36.4% increase in salary, while an increase of 1 in sales will result in a 0.017 or 1.7% increase in salary.



Total marks: 14 marks

**Question 4:**

1. Open the file MARKETDATA-ASSESSMENT

**This file contains data on the sales and price of a good in different regions and consumer incomes in those regions.**

1. Regress LOGPRICE on INCOME and save the predicted values (1 mark) **No need to do this…you will be a bonus mark for this question!**

1. What do the predicted values in a) purport to represent? (3 marks) **You will need to look at the lecture/notes/your notes on simultaneous equations for this**
2. Use the predicted values from a) to gain an estimate of the response of *supply* to a change in LOGPRICE. (3 marks) **You will need to quote the right figure from the right regression**

1. Is the equation that you are trying to estimate in c) identified? (3 marks) **Explain your answer**

1. Use the Hausmann test for exogeneity to evaluate whether the responsiveness of *supply* to price should be estimated using ordinary least squares or two stage least squares (4 marks) **Quote the relevant figures from the regression output and use them to conduct the test and reach the correct conclusion**.

Total marks 14 marks

**Question 5:**

1. Open the file LABOUR-CAPITAL

In this question we wish to test what percentage of Labour-Capital Share of Income rate (L-C%), is due to Growth Rate (GR), to capital investment (GFCF) and to trade union density (UNDENS), the natural logarithm of Working Hours (lnWKHRS) based on data from Organisation of Economic Co-operation and Development (OECD).

1. Write down the whole countries regression model for this and estimate the model (1 mark)
2. It is believed that the kind of political party in government (right-left) determines the Labour-Capital Share of Income rate (L-C%) and interacts with the rest of the independent variables too. We now introduce a dummy variable for political party in government (POLGOV=1 for left or centre left parties and POLGOV =0 for right wing parties). Test the null hypothesis that after ‘holding all the factors constant (ceteris paribus), POLGOV has no effect on Labour-Capital Share of Income rate (L-C%). State the alternative hypothesis that POLGOV has an effect on Labour-Capital Share of Income rate (L-C%). From your regression, what percentage of the squared variation in Labour-Capital Share of Income rate (L-C%) is due to the variations in the abovementioned variables? What is the predicted effect of the variables on Labour-Capital Share of Income rate (L-C%)? What is the p value for their t statistic? Does using a 1%, 5% or 10% significance level make a difference to your decision? (4 marks)
3. Do countries at years with centre-left governments have a different intercept? (Use the Chow test) (3 marks).
4. We wish to see the effect of political party in government on the above regression, in combination with the capital variables and labour variables (POLGOV- GR, POLGOV-GFCF, POLGOV - UNDENS, POLGOV-lnWKHRS). Can we capture any effect because of political party. Do we have a different slope or different intercepts for (centre) left now? Is there any evidence of multi-collinearity? (4 marks)
5. Now we want to see wh t is the probability of a country to have a (centre) left government (POLGOV =1) depending on the abovementioned variables (Labour-Capital Share of Income rate (L-C%), is due to Growth Rate (GR), to trade union density (UNDENS), to capital investment (GFCF), the natural logarithm of Working Hours (lnWKHRS))? Use the Linear Probability Model (Ordinary Least Squares –OLS) and the LOGIT Model (Maximum Likelihood Estimations –MLE) and interpret the results. Compare the two models

(8 marks)

Total marks: 20 marks

**Question 6:**

1. Open the file STOCKPRICES.

**This file contains data on a stock market index (INDEX) and the value of a portfolio of shares (PORT).**

1. Is the variable INDEX stationary or non-stationary? If it is non-stationary what form of non-stationarity does it exhibit? (6 marks) **Describe the test/tests you are conducting and quote the right figures from the regression output to support your argument.**
2. Is the variable PORT stationary or non-stationary? If it is non-stationary what form of non-stationarity does it exhibit? (6 marks) **Describe the test/tests you are conducting and quote the right figures from the regression output to support your argument**
3. Conduct a regression to test the relationship between INDEX and PORT and justify your choice of technique. (8 marks) **Describe the test/tests you are conducting and quote the right figures from the regression output to support your argument**

Total marks 20 marks

**-END OF PAPER-**