Assignment 2

Problem 1:
Regression Output.
The Stata regression output below reports the results of the regression of the net financial wealth ($nettfa$) on the annual family income (inc) and the age of the individual (age): $nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u$.

1. Interpret the coefficients. (inc and nettf are 1000's, age is measured in years). What is the interpretation of R^2?
2. Test the hypothesis that both slope coefficients are equal to zero. (State also the null hypothesis and alternative hypothesis!)
3. Test the following null hypothesis: $\beta_1 + \beta_2 = 2$!

(Note that the estimated coefficient for β_0 is denoted by _cons in the regression output)

```
      Source |       SS       df       MS              Number of obs =    9275
-------------+------------------------------           F(  2,  9272) =  943.21
       Model |   6414618.8     2   3207309.4           Prob > F      =  0.0000
    Residual |  31528770.7  9272  3400.42825           R-squared     =  0.1691
-------------+------------------------------           Adj R-squared =  0.1689
       Total |  37943389.5  9274   4091.3726           Root MSE      =  58.313

------------------------------------------------------------------------------
      nettfa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .9533566   .0252775    37.72   0.000     .9038072    1.002906
         age |   1.030777   .0591226    17.43   0.000     .9148838     1.14667
       _cons |  -60.69654   2.596333   -23.38   0.000    -65.78592   -55.60715
------------------------------------------------------------------------------
```

```
Covariance matrix of coefficients of regress model

        e(V) |        inc         age       _cons
-------------+------------------------------------
         inc |  .00063895
         age | -.00015787   .00349548
       _cons | -.01859631  -.13739775   6.7409438
```

Problem 2:
Use the data hprice1.dta (provided on the homepage) to run the following regression:
$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + \epsilon$

1. Before you estimate the model plot the dependent variable against each independent variable. What are your conclusions, is the relationship linear?
2. Provide some descriptive statistics of your data. (number of observations, mean, min, max?)
3. Perform a linear regression, report the estimated results and whether they have a statistically significant influence on the *price*. Additionally do the following tasks:
   - Plot the residuals against the fitted values of the dependent variable (Stata command: rvfplot). Which pattern can you see here?
   - Test for multicollinearity, do we have a problem here? (Stata command: vif)
   - Test for the normality of the residuals using a Q-Q Plot, what do you conclude?
4. Do the same regression again but now use the logarithm of the price, lotsize and sqrft. Plot again the residuals versus the fitted values of the dependent variable, what do you conclude now?

(Note that you do not have to provide interpretations of the coefficients in this example)

Problem 3:
Use the dataset jtrain.dta (provided on the homepage) to run the following regression:
$scrap = \beta_0 + \beta_1 hrsemp + \beta_2 sales + \beta_3 employ + \epsilon$

1. Before you estimate the model plot the dependent variable against each independent variable. What are your conclusions, is the relationship linear?
2. Provide some descriptive statistics of your data. (number of observations, mean, min, max?)
3. Perform a linear regression, report the estimated results and whether they have a statistically significant influence on the *price*. Additionally do the following tasks:
   - Plot the residuals against the fitted values of the dependent variable (Stata command: rvfplot). Which pattern can you see here?
   - Test for multicollinearity, do we have a problem here? (Stata command: vif)
   - Test for the normality of the residuals using a Q-Q Plot, what do you conclude?
4. Now use the log of *scrap*, *sales* and *employ* (do not use the logarithm of hrsemp) and plot again the log of the dependent variable versus the log of the independent variables. What do you conclude now?
5. Perform a linear regression with the logarithmic variables and report the results. Plot again the residuals versus the fitted values of the dependent variable, what do you conclude now? Additionally do again a Q-Q Plot and compare it to the Q-Q Plot of the estimated model without using logs.

(Note that you again do not have to provide interpretations of the coefficients in this example)