3. Use the data WAGE2.DTA for this exercise. This data provides individual information on wages, education and a number of other demographic characteristics. The primary question of interest is: "What is the effect of education on wages?".

    a. Estimate and report the coefficients from following regression using OLS:

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

    For these estimate to describe the causal effect of education on wages it must be true that $E(u|educ) = 0$. Is this assumption likely to be valid? Provide an intuitive explanation for your answer.

    b. It is suggested that the number of siblings an individual grows up with (variable *sibs*) can be used as an instrument for that individual's education in the above regression. To be a good instrument it must be **valid** and **relevant**. Discuss whether siblings meet these criteria. Test and comment on relevance by regressing education on wages as such:

$$educ = \alpha_0 + \alpha_1 sibs + e$$

    c. Use Stat's IV command (*ivreg*) to estimate the "causal" effect of education on wages under the assumption that *sibs* is a good instrument for *educ*. Comment on the estimated effect of education on wages ($\hat{\beta}_1$) versus that found in part a).

    d. Re-estimate the OLS regression in part a) replacing *educ* with the values predicted from the regression in part b) as such:

$$\widehat{educ} = \hat{\alpha}_0 + \hat{\alpha}_1 sibs$$
$$\log(wage) = \beta_0 + \beta_1 \widehat{educ} + u$$

Compare the resulting estimated effect of education on wages ($\hat{\beta}_1$) with that found in part c). Does the standard error of this estimate change? Explain this result.

Hint: the following code will generate the predicted values you are interested in (saved as variable *p_educ*):

reg educ sibs

predict p_educ