

```

-----
name: <unnamed>
log: C:\Documents and Settings\Michael Rosenfeld\My Documents\newer web
pages\soc_meth_pr
> oj3\fall_2010_s381_logs\class7.log
log type: text
opened on: 12 Oct 2010, 13:59:36

```

* First of all, I created a new variable, `hourwage_adj`, from `hourwage`. First I set the missing value codes to missing, then I divided by 100 to put the units from cents to dollars. You will note that MOST of the values are missing, which means only a minority of workers have earnings expressed as an hourly wage.

```

. gen hourwage_adj= hourwage

. replace hourwage_adj=. if hourwage>9997
(125702 real changes made, 125702 to missing)

. replace hourwage_adj= hourwage_adj/100
(8005 real changes made)

. twoway (scatter hourwage_adj yrsed) if age>29 & age<40 & sex==1

```

* Above is the plot just of the wages and education. Below I add the fitted line (and the graph is reproduced below).

```

. twoway (scatter hourwage_adj yrsed) (lfit hourwage_adj yrsed) if age>29 &
age<40 & sex==1

```

```

. regress hourwage_adj yrsed if age >29 & age<40 & sex==1

```

Source	SS	df	MS	Number of obs =	955
Model	4486.04715	1	4486.04715	F(1, 953) =	77.41
Residual	55226.3995	953	57.9500519	Prob > F =	0.0000
				R-squared =	0.0751
				Adj R-squared =	0.0742
Total	59712.4466	954	62.5916631	Root MSE =	7.6125

hourwage_adj	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yrsed	.8634938	.0981418	8.80	0.000	.6708947 1.056093
_cons	2.865037	1.277871	2.24	0.025	.3572705 5.372803

* So what is the relationship between the graph and the regression? Well, the regression line is the line on the graph. The slope is 0.86, meaning each additional year of education raises your hourly wage by 86 cents. The T statistic (which is just the coefficient divided by std error) is 8.8, which is highly significant, P value very small, meaning we are sure this slope (in the wider society) would be nonzero. The constant represents the predicted value of earnings when `yrsed==0`. If you look at the graph below, you will see that everyone with `yrsed==0` has earnings greater than the predicted value of 2.86, that is they are above the line.

Lastly, the R-square tells us how much of the variance in hourly wages is explained by years of education. The answer is 7.5% This means that there may be other things that explain hourly wages that we are not taking account of here...

