Econ 113
Problem Set 2
Answer Key

May 12, 2014

Problem 1

Using WageData.dta from the course website, please run the following regression:

$$log(wage) = \beta_0 + \beta_{Educ} Educ + \beta_{Educ^2} Educ^2 + u$$

Here, *wage* is the monthly wage and *Educ* is years of education.

a. Suppose I claim that the returns to education are not constant in education. What is the probability that I'm wrong?

Code:
gen educ2 = educ*educ
reg lwage educ educ2

Answer:

Source	SS	df	MS
Model	16.2782368	2	8.13911841
Residual	149.378057	932	.160276886
Total	165.656294	934	.177362199

Number of obs =	935
F(2, 932) =	50.78
Prob > F =	0.0000
R-squared =	0.0983
Adj R-squared =	0.0963
Root MSE =	.40035

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1327884	.0781342	1.70	0.090	-.0205509	.2861277
educ2	-.0025924	.0027686	-0.94	0.349	-.0080258	.0028409
_cons	5.473306	.5398847	10.14	0.000	4.413775	6.532836

For the returns to education to be constant, a one year increase of education should always be associated with the same increase in log wage, regardless of the initial number of years of education (i.e. the relationship between education and log wage must be linear). In this case, the coefficient on the squared term, β_{Educ^2}, would not be statistically different from 0. Our null and alternative hypotheses would be:

$$\text{H}_0\text{: } \beta_{Educ^2} = 0$$
$$\text{H}_A\text{: } \beta_{Educ^2} \neq 0$$

The probability that the claim "the returns to education are not constant in education" is wrong (i.e. that the alternative hypothesis is wrong) is simply the p-value on $\hat{\beta}_{Educ^2}$. As can be seen in the Stata output this is .349 or 34.9%.

b. Please solve for the value of education at which the returns to additional education are zero. Is this a maximum or minimum? Why?

Answer: To answer this, we need to find the point at which $\frac{\delta wage}{\delta educ} = 0$. To start, take the derivative of the regression equation with respect to education:

$$log(wage) = \beta_0 + \beta_{Educ}Educ + \beta_{Educ^2}Educ^2 + u$$
$$\frac{\delta wage}{\delta educ}\frac{1}{wage} = \beta_{Educ} + 2\beta_{Educ^2}Educ = 0$$
$$\rightarrow Educ = -\frac{\beta_{Educ}}{2\beta_{Educ^2}}$$

Though we don't know the true values of β_{Educ} and β_{Educ^2} we can substitute in $\hat{\beta}_{Educ}$ and $\hat{\beta}_{Educ^2}$ from the Stata regression output:

$$Educ = -\frac{\beta_{Educ}}{2\beta_{Educ^2}}$$
$$Educ = -\frac{.1327884}{2(-.0025924)} = 25.611$$

However, in order to know if this value of education gives us a maximum or minimum log-wage, we need to look at the second derivative of the regression equation:

$$\frac{\delta^2 wage}{\delta educ^2} = 2\beta_{Educ^2} = 2(-.0025924) < 0$$

Since the second derivative is negative we know the function is concave and the value we found is where log-wage is maximized. Therefore, we can say that, according to the regression estimates, the log-wage is maximized at 25.6 years of education.

Problem 2

Next, please estimate the effect of categorical factors on wages using the following specification:

$$wage = \beta_0 + \beta_{Married}Married + \beta_{Urban}Urban + \beta_{South}South + u$$

where Urban = 1 if the respondent lives in an Metropolitan Statistical Area and 0 otherwise, and South = 1 if the respondent lives in the South and 0 otherwise, and Married = 1 if the respondent is married and 0 otherwise.

a. Please construct and interpret a 90% confidence interval for β_{Urban}

Code: reg wage married urban south, level(90)

Answer:

Source	SS	df	MS
Model	12249788.9	3	4083262.96
Residual	140466379	931	150876.884
Total	152716168	934	163507.675

Number of obs =	935
F(3, 931) =	27.06
Prob > F =	0.0000
R-squared =	0.0802
Adj R-squared =	0.0772
Root MSE =	388.43

wage	Coef.	Std. Err.	t	P>|t|	[90% Conf.	Interval]
married	192.7671	41.14338	4.69	0.000	125.0249	260.5094
urban	169.4475	28.41228	5.96	0.000	122.6669	216.228
south	-121.0465	26.9615	-4.49	0.000	-165.4384	-76.65458
_cons	705.4899	45.79462	15.41	0.000	630.0894	780.8903

As shown in the Stata output, we are 90% confident that, holding all else constant, living in an urban area is associated with earning on average between \$122.67 and \$216.23 more than non-urban areas.

b. Please construct and interpret a 99% confidence interval for β_0

Answer: To answer this we can utilize the info from the Stata output:

$$\hat{\beta}_0 \pm SE_{\beta_0} * T_{crit,99\%}$$

$$705.4899 \pm (49.7946) * (2.576)$$

$$\rightarrow (577.21, 833.76)$$

We are 99% confident that on average an unmarried person living in the rural north earns between \$577.21 and \$833.76 a month.

Problem 3

Finally, please estimate the effect of continuous and categorical factors on wages using the following specification:

$$log(wage) = \beta_0 + \beta_{Educ}Educ + \beta_{Urban}Urban + \beta_{EU}Educ * Urban + u$$

a. Please construct and carefully interpret a 90% confidence interval for β_{Educ}.

Code:
gen EU = educ*urban
reg lwage educ urban EU, level(90)

Answer:

Source	SS	df	MS
Model	22.0116562	3	7.33721872
Residual	143.644638	931	.154290696
Total	165.656294	934	.177362199

Number of obs =	935
F(3, 931) =	47.55
Prob > F =	0.0000
R-squared =	0.1329
Adj R-squared =	0.1301
Root MSE =	.3928

lwage	Coef.	Std. Err.	t	P>|t|	[90% Conf.	Interval]
educ	.0419299	.0111968	3.74	0.000	.0234945	.0603654
urban	-.1119765	.1773231	-0.63	0.528	-.4039375	.1799845
EU	.021211	.0131455	1.61	0.107	-.0004329	.042855
_cons	6.088102	.149938	40.60	0.000	5.841231	6.334974

As can be seen from the Stata output, we are 90% confident that the return to an additional year of education for someone in a non-urban area is, on average, between 2.35% and 6.04%.

b. Many large firms locate in and around cities, and therefore the returns to education may depend on whether one lives in a urban or rural location. Please conduct a two-sided hypothesis test to evaluate this question, using the 99% level of confidence. Please interpret your results.

Answer: If the return to education depends where someone lives, the coefficient on the interaction term, β_{EU}, would be statistically different from zero. Therefore we test the following hypothesis:

$$\text{H}_0\text{: } \beta_{EU} = 0$$
$$\text{H}_A\text{: } \beta_{EU} \neq 0$$

From the output in part a. we can see that the p-value for $\hat{\beta}_{EU}$ is .107. Since .107 is greater than .01, we fail to reject the null hypothesis that the slope of the interaction is different from zero. In other words, we cannot reject the assertion that the returns to education are the same across urban and rural locations.