Econmetrics II: Exercise 1

Econometrics II

Exercise 1

(a) **Load the data into Stata and familiarize yourself with the structure and the variables using different descriptive statistics.**

This dataset has 17 columns and 935 rows. We can look at the descriptions of variables in the following table.

Table 1: Description of variables

Variable name	**Variable label**
wage	monthly earnings
hours	average weekly hours
iq	IQ score
kww	knowledge of world work score
educ	years of education
exper	years of work experience
tenure	years with current employer
age	age in years
married	=1 if married
black	=1 if black
south	=1 if live in south
urban	=1 if live in SMSA
sibs	number of siblings
brthord	birth order
meduc	mother's education
feduc	father's education
lwage	log(wage)

(b) **Have a closer look at the wage variable (monthly earnings in $). How does the mean wage compare to the median? Plot the wage distribution. What is the expected wage of a working man if he is married?**

Table 2: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
wage	935	957.950	404.360	115	669	1,160	3,078
hours	935	43.929	7.224	20	40	48	80
iq	935	101.280	15.053	50	92	112	145
kww	935	35.744	7.639	12	31	41	56
educ	935	13.468	2.197	9	12	16	18
exper	935	11.564	4.375	1	8	15	23
tenure	935	7.234	5.075	0	3	11	22
age	935	33.080	3.108	28	30	36	38
married	935	0.893	0.309	0	1	1	1
black	935	0.128	0.335	0	0	0	1
south	935	0.341	0.474	0	0	1	1
urban	935	0.718	0.450	0	0	1	1
sibs	935	2.941	2.306	0	1	4	14
brthord	852	2.277	1.596	1.000	1.000	3.000	10.000
meduc	857	10.683	2.850	0.000	8.000	12.000	18.000
feduc	741	10.217	3.301	0.000	8.000	12.000	18.000
lwage	935	6.779	0.421	4.745	6.506	7.056	8.032

Mean of **wage** = \$957.94545 and median \$905.

The **expected wage** of a **married man** is \$977.0479. The difference between the average wage of married and unmarried person is \$178.6079. Unmarried person earns way less.

(c) **Consider the following equation for men under the assumption that all explanatory variables are exogenous**

$$wage = \beta_0 + \beta_1 married + \beta_2 high\_iq + \beta_3 medium\_edu + \beta_4 high\_educ + z\gamma + u \quad (1)$$

$E(u|married, iq, educ, z) = 0$.

How do you interpret β_1?

- A married person would get β_1 more dollars of wage over unmarried one, **on average**, **other factors held intact**.

How would you interpret the constant term?

- Usually, the intercept tells us what the value of the dependent variable (here wage) would be when all independent variables are zero. However, in this typical regression setup, the intercept captures the average wage of unmarried person, while all other factors are set to zero.

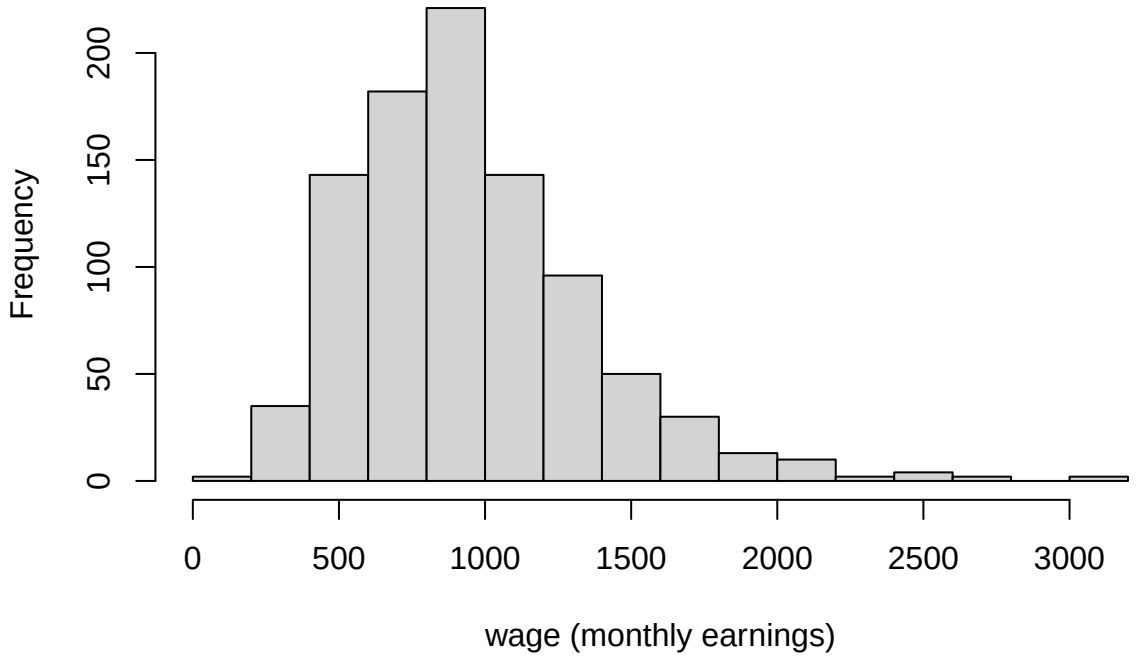

Figure 1: Distribution of monthly earnings

(d) Run the regression and interpret the regression coefficients. Include all education dummies to the regression model. Why does Stata drop one of the dummies? How does the interpretation of the constant change if you change the included education dummies?

When we include all the education dummies, Stata/R drops `low_educ` because of **singularity** with the other educ dummies; one is a linear combination of the other, in fact perfectly correlated: `corr(high_educ, medium_educ, low_educ)` = 1, -0.25533, -0.83109

$$wage = \beta_0 + \beta_1 married + \beta_2 high\_iq + \beta_4 high\_educ + z\gamma + u \tag{2}$$

(e) Why would you like to interact the married-dummy with the high_iq-dummy? Run the regression including the interaction term and interpret the coefficient.

(f) Consider now the a standard log(wage)-equation:

Table 3: Regression output: level-level

	Dependent variable:		
	wage		
	(1)	(2)	(3)
married1	159.700***	160.700***	128.500**
	(43.246)	(43.334)	(60.588)
high_iqhigh	96.844***	103.970***	45.819
	(31.028)	(30.891)	(82.446)
high_educ	196.410***	168.760***	169.020***
	(37.475)	(34.955)	(34.967)
medium_educ	106.340**		
	(52.664)		
low_educ			
exper	9.706**	8.172*	8.146*
	(4.253)	(4.194)	(4.195)
age	15.228***	15.759***	15.820***
	(5.286)	(5.291)	(5.293)
tenure	4.401	4.077	4.122
	(2.862)	(2.864)	(2.865)
feduc	18.365***	19.589***	19.581***
	(4.570)	(4.539)	(4.540)
married1:high_iqhigh			65.409
			(85.978)
Constant	−139.040	−133.050	−106.660
	(161.190)	(161.500)	(165.230)
Observations	741	741	741
R^2	0.179	0.174	0.175
Adjusted R^2	0.170	0.166	0.166
Residual Std. Error	369.820 (df = 732)	370.590 (df = 733)	370.700 (df = 732)
F Statistic	19.929*** (df = 8; 732)	22.101*** (df = 7; 733)	19.399*** (df = 8; 732)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Regression output: log-level, log-log and level-log

	Dependent variable:		
	ln(wage)		wage
	(1)	(2)	(3)
married1	0.184***	0.183***	172.010***
	(0.044)	(0.044)	(43.192)
educ	0.060***		
	(0.008)		
log(educ)		0.849***	795.990***
		(0.111)	(108.800)
exper	0.017	0.016	20.400
	(0.015)	(0.015)	(14.369)
I(exper^2)	−0.0001	−0.00004	−0.365
	(0.001)	(0.001)	(0.645)
age	0.011*	0.011*	13.032**
	(0.006)	(0.006)	(5.707)
tenure	0.009***	0.009***	4.318
	(0.003)	(0.003)	(2.851)
feduc	0.017***	0.017***	17.802***
	(0.005)	(0.005)	(4.575)
Constant	5.061***	3.669***	−2,065.100***
	(0.204)	(0.295)	(288.580)
Observations	741	741	741
R^2	0.189	0.192	0.181
Adjusted R^2	0.182	0.184	0.173
Residual Std. Error (df = 733)	0.378	0.377	369.140
F Statistic (df = 7; 733)	24.478***	24.866***	23.100***
Note:			*p<0.1; **p<0.05; ***p<0.01