Name ____________ **Surname** ____________ **Student ID.** ____________

DUE DATE : Thursday 24, November 2016.

Assignment 5: (86 points)

I pledge to the Honor Code and to obey all rules for taking and performing homework assignments as specified by the course instructor.

Student Signature: ____________

1. (25 points) Empirical Exercise :

Use the data set **CPS12** described in Empirical Exercise AEE4.1 to answer the following questions.

a. Run a regression of average hourly earnings (*AHE*) on age (*Age*). What is the estimated intercept? What is the estimated slope?

b. Run a regression of *AHE* on *Age*, gender (*Female*), and education (*Bachelor*). What is the estimated effect of *Age* on earnings? Construct a 95% confidence interval for the coefficient on *Age* in the regression.

c. Are the results from the regression in (b) substantively different from the results in (a) regarding the effects of *Age* and *AHE*? Does the regression in (a) seem to suffer from omitted variable bias?

d. Bob is a 26-year-old male worker with a high school diploma. Predict Bob's earnings using the estimated regression in (b). Alexis is a30-year-old female worker with a college degree. Predict Alexis's earnings using the regression.

e. Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and $\bar{R}^2$. Why are the R^2 and $\bar{R}^2$ so similar in regression (b)?

f. Are gender and education determinants of earnings? Test the null hypothesis that *Female* can be deleted from the regression. Test the null hypothesis that *Bachelor* can be deleted from the regression. Test the null hypothesis that both *Female* and *Bachelor* can be deleted from the regression.

2. (25 points) Empirical Exercise: Heteroskedasticity Problem.

Consider the following model to explain sleeping behavior:

$$sleep = \beta_1 + \beta_2 totwrk + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + \beta_6 male + u$$

2.1 Use data in **SLEEP75.xls** to estimate the parameters of the above model.

2.2 Plot the diagram to show the relationship between ${\hat{u}_i}^2$ and the estimated Y_i from the regression line $\hat{Y}_i$. What does this figure suggest about heteroskedasticity?

2.3 Compute the Breusch-Pagan test for heteroskedasticity by using LM-test version and report your result. What do you conclude?

2.4 Compute the White-test for heteroskedasticity. Use the F-statistic version and report your result. How strong is the evidence for heteroskedasticity?

2.5 Compute the heteroskedasticity-robust standard errors , then compare the usual OLS standard errors with the heteroskedasticity-robust standard errors.

3 (18 points) Consider the log monthly earnings equation as follows:

$$\begin{aligned}\log(\text{wage})_i &= \beta_1 + \beta_2\text{educ}_i + \beta_3\text{exper}_i + \beta_4\text{tenure}_i + \beta_5\text{married}_i \\ &\quad \beta_6\text{black}_i + \beta_7\text{south}_i + \beta_8\text{urban}_i + u_i\end{aligned} \tag{Eq.1}$$

where educ, exper, and tenure are all relevant productivity characteristics. Married, black, south, and urban are qualitative variables.

$\log(\text{wage})_i$ =natural log of wage

educ_i =years of education

exper_i =years of work experience

tenure_i =years with current employer

Married_i =1 if married,

$\text{black}_i = 1$ if black,

south_i =1 if living in the south,

urban_i =1 if living in urban.

The estimation result model Eq.1 is reported as below:
Table 3.1 the regression result of model Eq.1

Source	SS	df	MS
Model	41.8377619	7	5.97682312
Residual	123.818521	927	.133569063
Total	165.656283	934	.177362188

Number of obs	= 935
F(7, 927)	= 44.75
Prob > F	= 0.0000
R-squared	= 0.2526
Adj R-squared	= 0.2469
Root MSE	= .36547

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0654307	.0062504	10.47	0.000	.0531642	.0776973
exper	.014043	.0031852	4.41	0.000	.007792	.020294
tenure	.0117473	.002453	4.79	0.000	.0069333	.0165613
married	.1994171	.0390502	5.11	0.000	.1227801	.276054
black	-.1883499	.0376666	-5.00	0.000	-.2622717	-.1144281
south	-.0909036	.0262485	-3.46	0.001	-.142417	-.0393903
urban	.1839121	.0269583	6.82	0.000	.1310056	.2368185
_cons	5.395497	.113225	47.65	0.000	5.17329	5.617704

3.1 (2 points) Write out the regression equation for log monthly earning based on model Eq.1.

3.2 (4 points) Which of the coefficients are individually statistically significant at the 5 percent level of significance? State the critical value for hypothesis testing to receive full points.

3.3 (5 points) Holding other factors fixed, what is the approximate difference in monthly earnings between black and nonblacks? Is this difference statistically significant?

Next, we extend the original model to allow the interaction term between the married variable and the black variable by adding the "$\text{married}_i\text{black}_i$" to the equation. The new model is

$$\begin{aligned}\log(\text{wage})_i &= \gamma_1 + \gamma_2\text{educ}_i + \gamma_3\text{exper}_i + \gamma_4\text{tenure}_i + \gamma_5\text{married}_i \\ &\quad \gamma_6\text{black}_i + \gamma_7\text{south}_i + \gamma_8\text{urban}_i + \gamma_9\text{married}_i\text{black}_i + u_i\end{aligned} \tag{Eq.2}$$

The estimation result is reported as below:

Table 3.2 the regression result of model Eq.2

Source	SS	df	MS
Model	41.8849359	8	5.23561699
Residual	123.771347	926	.133662362
Total	165.656283	934	.177362188

Number of obs =	935
F(8, 926) =	39.17
Prob > F =	0.0000
R-squared =	0.2528
Adj R-squared =	0.2464
Root MSE =	.3656

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0654751	.006253	10.47	0.000	.0532034	.0777469
exper	.0141462	.003191	4.43	0.000	.0078837	.0204087
tenure	.0116628	.0024579	4.74	0.000	.006839	.0164866
married	.1889147	.0428777	4.41	0.000	.1047659	.2730635
black	-.24082	.0960229	-2.51	0.012	-.4292677	-.0523723
south	-.0919894	.0263212	-3.49	0.000	-.1436455	-.0403333
urban	.1843501	.0269778	6.83	0.000	.1314053	.2372948
marriedblack	.0613537	.1032747	0.59	0.553	-.1413259	.2640333
_cons	5.403793	.1141222	47.35	0.000	5.179825	5.627762

Note: marriedblack = married*black

3.4 (4 points) Write out the regression equation for log monthly earning based on model Eq.2. Does there appear to be a significant interaction effect among the new terms? Assess this with respect to both the black dummy variable and their interaction term and explained the results. What is the conditional expectation of $\log(\text{wage})_i$ for married blacks?

3.5 (3 points) What is the estimated wage differential between married blacks and married nonblacks?

4 (18 points) To assess the feasibility of a guaranteed annual wage (negative income tax), the Rand Corporation conducted a study to assess the response of labor supply (average hours of work) to increasing hourly wages.The data for this study were drawn from a national sample of 6,000 households with a male head earning less than $ 15,000 annually. The data were divided into 39 demographic groups for analysis. Because data for four demographic groups were missing for some variables, the data given in this example refer to only 35 demographic groups. Estimate the model Eq.3 reports in the Table 4.1

$$\begin{aligned} \text{Hours}_i &= \beta_1 + \beta_2 \text{rates}_i + \beta_3 \text{ERSP}_i + \beta_4 \text{ERNO}_i + \beta_5 \text{NEIN}_i \\ &\quad \beta_6 \text{asset}_i + \beta_7 \text{age}_i + \beta_8 \text{DEP}_i + \beta_9 \text{school}_i + u_i \end{aligned} \tag{Eq.3}$$

where

Hours_i = average hours worked during the year

rate_i = average hourly wage (dollars)

ERSP_i = average yearly earnings of spouse (dollars)

ERNO_i = average yearly earnings of other family members (dollars)

NEIN_i = average yearly nonearned income

assets_i = average family asset holdings (bank account, etc.) (dollars)

age_i = average age of respondent

DEP_i = average number of dependents

school_i = average highest grade of school completed

The estimation result model Eq.3 is reported as below:

Table 4.1 the regression result of model Eq.3

Source	SS	df	MS
Model	115385.114	8	14423.1393
Residual	24381.6287	26	937.754952
Total	139766.743	34	4110.78655

Number of obs	=	35
F(8, 26)	=	15.38
Prob > F	=	0.0000
R-squared	=	0.8256
Adj R-squared	=	0.7719
Root MSE	=	30.623

hrs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
rate	-93.75259	47.14499	-1.99	0.057	-190.6605	3.155329
ersp	.0002254	.0382548	0.01	0.995	-.0784084	.0788592
erno	-.2149663	.0979392	-2.19	0.037	-.4162832	-.0136495
nein	.1572073	.5164059	0.30	0.763	-.9042802	1.218695
asset	.0155724	.0254048	0.61	0.545	-.0366479	.0677928
age	-.3486302	3.72233	-0.09	0.926	-7.999989	7.302729
dep	20.72802	16.88047	1.23	0.230	-13.97029	55.42633
school	37.32565	22.66518	1.65	0.112	-9.263304	83.9146
_cons	1904.577	251.9332	7.56	0.000	1386.721	2422.433

4.1 (2 points) Write out the regression equation for model Eq.3.

Table 4.2 the correlation matrix

	rate	ersp	erno	nein	asset	age	dep	school
rate	1.0000							
ersp	0.5717	1.0000						
erno	0.0590	-0.0410	1.0000					
nein	0.7018	0.2344	0.3591	1.0000				
asset	0.7789	0.2741	0.2922	0.9875	1.0000			
age	0.0442	-0.0153	0.7755	0.5024	0.4171	1.0000		
dep	-0.6014	-0.6929	0.0502	-0.5208	-0.5136	-0.0484	1.0000	
school	0.8813	0.5491	-0.2986	0.5392	0.6309	-0.3311	-0.6026	1.0000

4.2 (8 points) Is there evidence of multicollinearity in the data? How do you know? Explain your answers in detail and state the critical value for hypothesis testing to receive fully points. If there is the multicollinearity problem, what remedial action, if any, would you take?

Table 4.3 Variance Inflation Factors and Tolerance (VIF

Variable	VIF	1/VIF
asset	192.57	0.005193
nein	180.51	0.005540
school	25.40	0.039367
rate	17.08	0.058540
age	9.72	0.102869
dep	4.53	0.220992
ersp	3.50	0.285841
erno	3.14	0.318680
Mean VIF	54.55	

4.3 (3 points) Explain the outcome of Variance Inflation Factors and Tolerance (VIF).

4.4 State with reason whether the following statement are true or false.

a. (2.5 points) Despite perfect multicollinearity, OLS estimators are BLUE.

b. (2.5 points) In case of high multicollinearity, it is not possible to assess the individual significance of one or more partial regression coefficients.