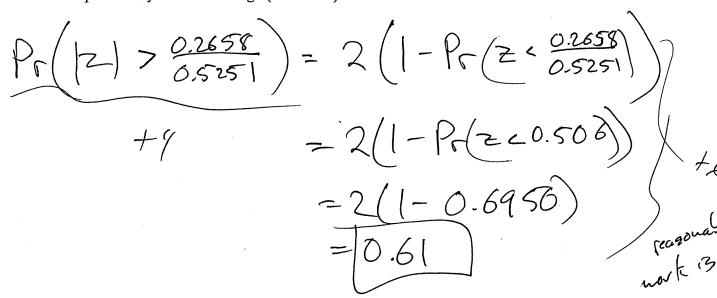
Economics 113 Professor Spearot			
Introduction to Econometrics Fall 2011 – Final			
Name	ID		
	Final – 120 Points		
notes. You may use calculators, but they own scratch paper.	rite your name on every page. The exam is closed book and closed would must not be graphing calculators. No cell phones. Do not use your		
You must show your work to receive full credit			
I have neither given nor received unauth misconduct by others.	norized aid on this examination, nor have I concealed any similar		
Signature Ke			
Purklam 1 (40 Paints)			
Problem 1 (40 Points)			
Consider the following simple specificat	tion that tests for regional differences in hours worked:		
	$hours = \beta_0 + \beta_1 urban + u$		
hours is average hours worked per week respondent lives in a metropolitan area,	and <i>urban</i> is a dummy variable that takes on a value of 1 if the and 0 otherwise. The results from estimating this equation are below:		
(Intercept) 43.7386	Std. Error t value Pr(> t ) 0.4448 XXXXXXXXXXXXXXX 0.5251 XXXXXXXXXXXXXX		
Multiple R-squared: 0. F-statistic: 0.2563 or	0002747, Adjusted R-squared: -0.0007969 1 and 933 DF, SSR=48731.95		
a.) Please construct and interpret a 95%	confidence interval for the intercept. (10 Points)		
Gent = 1.96 A			
43.7 - 0.44.1.9	16 KBO < \$3.7 + 0.44.1.96		
/ 4.2.81	1 < B < 44.56 / > +4		
With 95% confidure	, a person that lives in a moral (non-uton		
location entre bet	neen 42.84 and 44.56 hours		
perweek.	+4.		

b.) I claim that urban residents work a number of hours that is significantly different than rural residents. What is the probability that I'm wrong? (10 Points)



c.) Suppose that instead of the regression in 'a', I run the following regression:

$$hours = \beta_0 + \beta_1 urban + \beta_2 educ + u$$

where *educ* is the years of education of the respondent. The results from estimating this equation are below:

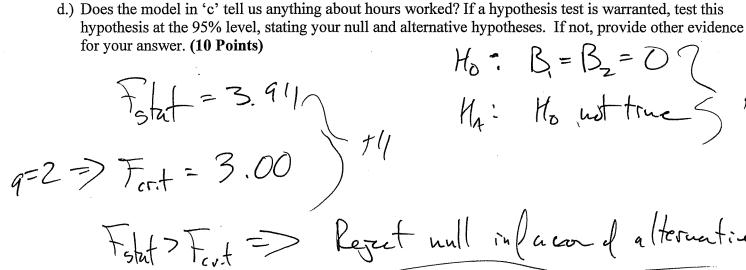
	Estimate	Std. Error	t value Pr(> t )
(Intercept)	39.8146	1.4889	XXXXXXXXXXXXXX
urban	0.1613	0.5246	XXXXXXXXXXXXXXX
educ	0.2969	0.1076	XXXXXXXXXXXXXX

Multiple R-squared: 0.008383, Adjusted R-squared: 0.006255

F-statistic: 3.94 on 2 and 932 DF, SSR=48336.7

In comparing the regression in 'a' and the regression in 'c', what is the correlation between *educ* and *urban*? Why? (10 Points)

The correlation is positive. In 'a' educe is an omitted variable. Dien the positive valationship between educ and hours, and sien Burhan goes down, there must be a positive correlation between educ and urbay.



e.) Suppose that I modify the regression in 'c' to include age and age 2, which are the age and age squared of the respondent.

$$hours = \beta_0 + \beta_1 urban + \beta_2 educ + \beta_3 age + \beta_4 age^2 + u$$

The results from this regression are below:

Multiple R-squared: 0.009078, Adjusted R-squared: 0.004816 F-statistic: 2.13 on 4 and 930 DF, SSR=48302.81

At what age is average hours worked minimized? Show your work!! (10 Points)

$$\frac{\partial hours}{\partial use} = \hat{B}_{3} + 2\hat{B}_{4} ase = 0 + 5$$

$$\frac{1}{3}e = \frac{\hat{B}_{3}}{2\hat{B}_{4}}$$

$$= \frac{0.200}{2.0.0031}$$

$$\frac{1}{3}e = \frac{25.64}{1}$$

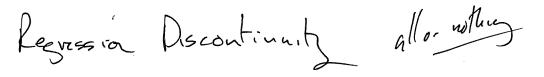
f.) Is the model in 'e' preferred to the model in 'c'? If a hypothesis test is warranted, test this hypothesis at the 95% level, stating your null and alternative hypotheses. If not, provide other evidence for your answer. (10 Points)

## Problem 2 (40 Points)

a.) For this problem, we wish to study the impact of health insurance on the smoking behavior of pregnant mothers. While difficult to assess, we will leverage a family's eligibility for prenatal care via Medicaid to determine the effects of health insurance on behavior. To do so, we run the following regression:

$$smoke = \beta_0 + \beta_1 faminc + \beta_2 medicaid + \beta_3 faminc \cdot medicaid + u$$

Here, *smoke* takes on a value of 1 if a mother smoked during pregnancy, and zero otherwise. Further, *faminc* is yearly family income (in thousands) and *medicaid* is a dummy variable taking a value of 1 if *faminc* is below 22 (which is \$22,000) and zero otherwise. What kind of regression technique is this? (10 **Points**)



b.) The results from estimating the regression in 'a' are below:

Please use a t-test to test whether Medicaid eligibility (at the eligibility threshold) affects smoking behavior. Please state your null and alternative hypotheses, and test the null against the alternative at the 99% level. (10 Points)

c.) Does the relationship between family income and maternal smoking behavior depend on whether the family is eligible for Medicaid? Test this hypothesis at the 98% level using a two-sided test. State your null and alternative, and show your work! (10 Points)

d.) Suppose that instead of the above model, we estimate the following model:

$$smoke = \beta_0 + \beta_1 motheduc + \beta_2 medicaid + u$$

where *motheduc* is the mother's education level in years. The results are below:

Multiple R-squared: 0.05677, Adjusted R-squared: 0.05519 F-statistic: 35.75 on 2 and 1188 DF, SSR=131.3309

Please interpret the coefficient on *medicaid*, and test whether this coefficient is significantly different from zero. Please state your null and alternative hypotheses, and test the null against the alternative at the 90% level. (10 Points)

e.) Which regression is preferred, the regression in '2b' or the regression in '2d'? If a hypothesis test is warranted, test this hypothesis at the 95% level, stating your null and alternative hypotheses. If not, provide other evidence for your answer. (10 Points)

Non-Nested = Agustend R<sup>2</sup>
Ad R<sup>2</sup> = 0.0252 + 4
Ad; R<sup>2</sup> = 0.0552 + 1

=> Model in 1011 13 present +2

+3 la F-test

f.) Using the previous regression equation in 'd', we wish to predict the probability of smoking for a mother with 20 years of education that is eligible for Medicaid. Please derive a regression equation that allows us to generate this prediction with standard error, and write the R commands that would estimate this particular equation. Show your work!! (10 Points)