

There are ten questions on the exam. Each question is worth ten points.
Conduct all hypothesis tests at the five percent level of significance.

1. The American Civil Liberties Union (ACLU) made the following statistical finding: “Racial disparities in New Jersey marijuana arrests are at an all-time high. The racial disparity in marijuana possession arrests reached an all-time high in 2013. That year, Black New Jerseyans were three times more likely to be arrested for marijuana possession than whites, despite similar usage rates.” Explain the importance that the phrase “similar usage rates” has to the finding of “[r]acial disparities.”

2. A student estimated the following savings function using ordinary least squares (saving = Total value of savings accounts held by household):

Source	SS	df	MS		n=4,418	
Model	6.25E+11	9	6.94E+10	Prob > F		
Residual	1.68E+13	4,408	3.81E+09	R-squared	0.0359	
				Adj R-square	0.0339	
Total	1.74E+13	4,417	3.95E+09	Root MSE	61746	
saving	Coef.	Std. Err.	t	P>t	[95% Interval]	
income	0.0153901	0.0032247	4.77	0	0.009068	0.0217122
educ	1253.95	353.1145	3.55	0		
married	2157.613	1979.709	1.09	0.276	-1723.611	6038.837
kids	1716.015	889.2228	1.93	0.054	-27.30828	3459.338
age	134.5031	59.63375	2.26	0.024	17.59104	251.4153
white	789.7357	4690.805	0.17	0.866	-8406.598	9986.07
black	-1520.538	5288.081	-0.29	0.774	-11887.83	8846.758
hispanic	-624.5823	5601.883	-0.11	0.911	-11607.09	10357.92
networth	0.0017863	0.0003523	5.07	0	0.0010957	0.002477
_cons	-21559.49	7861.382	-2.74	0.006	-36971.74	-6147.231

Use the statistical results to test the overall significance of the model at the 5% level of significance.

3. Use the results from the regression to construct a 95% confidence interval for educ (the number of years of schooling, which range in value from 1 to 17).
4. Explain the meaning of the confidence interval estimate you calculated in response to question 3. Assume I know nothing about econometrics or statistics. Explain what the estimated interval suggests about the coefficient for educ.
5. The student noticed that the coefficients for the race variables had large p-values. He reran the regression, after dropping the three variables. Test the hypothesis that the three variables are jointly equal to zero at the five percent level of significance.

Source	SS	df	MS	Number of obs =	4,418
				F(6, 4411) =	
Model	6.22E+11	6	1.04E+13	Prob > F =	
Residual	1.68E+13	4,411	3.81E+11	R-squared =	0.0357
				Adj R-squared =	0.0344
Total	1.74E+13	4,417	3.95E+09	Root MSE =	61730
saving	Coef.	Std. Err.	t	P>t	[95% Conf Interval]
income	0.0154246	0.0032236	4.78	0	0.0091048 0.0217444
educ	1298.147	338.4203	3.84	0	
married	2353.71	1956.309	1.2	0.229	-1481.638 6189.058
kids	1656.493	884.9214	1.87	0.061	-78.39669 3391.383
age	143.1119	58.31384	2.45	0.014	28.78748 257.4363
networth	0.0017893	0.0003521	5.08	0	0.001099 0.0024796
_cons	-22313.2	6050.333	-3.69	0	-34174.89 -10451.51

6. The student accepted the null hypothesis that the coefficients were not statistically different than zero. He was surprised that race was not statistically significant. He wrote: “Normally I believe that race does play an important role in determining household savings based on the idea that white people have a higher...income...With income you have more money to spare so you would save more...”

Wooldridge points out that the beauty of econometrics is it allows us to hold other things constant: “The notion of ceteris paribus—which means ‘other relevant factors being equal’ plays an important role in [econometric] analysis....For example, in analyzing consumer demand, we are interested in knowing the effect of changing the price of a good on its quantity demanded, while holding all other factors—such as income, prices of other goods, and individual tastes—fixed. If other factors are not held fixed, then we cannot know the causal effect of a price change on quantity demanded.”

Knowing what you know about econometrics, do you think the student analysis is sensible (white households have higher incomes and therefore they should have higher savings)? Explain why or why not.

7. Draw a diagram that illustrates the relationship between education (educ) and savings, for married and non-married households using the regression results from problem 5.

8. I thought that the savings function might have an inverted U-shape with respect to income. The inverted U-shape could reflect that while savings increases at first, as income increases, at some point it starts to decline as the household moves money out of low-interest paying saving accounts and into other financial instruments, like bonds or real-estate investment trusts (REITs). I therefore added to the regression $\text{income2} = \text{income} * \text{income} = \text{income}^2$. Test at the 5% level of significance my hypothesis that income2 belongs in the econometric specification, in order to reflect a non-linear (U-shaped) savings pattern. Note -2.64e-10 is equivalent to moving the decimal point 10 spots to the left. Therefore -2.64e-10 = 0.000000000264.

Source	SS	df	MS	Number of obs	=	4,418
				F(7, 4410)	=	26.47
Model	7.0291e+11	7	1.0042e+11	Prob > F	=	0.0000
Residual	1.6728e+13	4,410	3.7932e+09	R-squared	=	0.0403
				Adj R-squared	=	0.0388
Total	1.7431e+13	4,417	3.9463e+09	Root MSE	=	61589

saving	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	.0303075	.004555	6.65	0.000	.0213774	.0392376
income2	-2.64e-10	5.72e-11	-4.61	0.000	-3.76e-10	-1.52e-10
educ	1149.937	339.1691	3.39	0.001	484.9952	1814.879
married	1669.082	1957.457	0.85	0.394	-2168.517	5506.682
kids	1562.987	883.1257	1.77	0.077	-168.3827	3294.357
age	140.1844	58.18365	2.41	0.016	26.11526	254.2536
networth	.0012695	.0003689	3.44	0.001	.0005462	.0019928
_cons	-20646.75	6047.261	-3.41	0.001	-32502.42	-8791.083

9. 2,458 of the 4,418 saving observations are equal to zero.
I decided to re-estimate the model using Tobit, instead of OLS, and obtained the following results.

Tobit regression		Number of obs =		4,418	
		Uncensored =		1,960	
Limits: lower = 0		Left-censored =		2,458	
upper = +inf		Right-censored =		0	
		LR chi2(9) =		139.38	
		Prob > chi2 =		0.0000	
Log likelihood = -27044.036		Pseudo R2 =		0.0026	

saving	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	.0109859	.0061069	1.80	0.072	-.0009866	.0229584
educ	5131.392	704.4603	7.28	0.000	3750.296	6512.488
married	9204.378	3728.333	2.47	0.014	1894.973	16513.78
kids	3180.489	1639.952	1.94	0.053	-34.6414	6395.619
age	-87.7224	113.2327	-0.77	0.439	-309.7154	134.2706
white	10110.63	8678.986	1.16	0.244	-6904.541	27125.8
black	-4574.214	9898.52	-0.46	0.644	-23980.28	14831.86
hispanic	-10098.86	10629.62	-0.95	0.342	-30938.25	10740.52
networth	-.0046425	.000745	-6.23	0.000	-.0061031	-.003182
_cons	-117143.6	15124.62	-7.75	0.000	-146795.4	-87491.72
var(e.saving)	9.78e+09	3.24e+08			9.16e+09	1.04e+10

- a. Why did it make sense for me to use Tobit when estimating the savings function?
-
-
-
- b. Are you surprised that the slope term for educ is higher with Tobit than the for the OLS results provided in problem 2? Explain.
-
-
-

10. Your data set included a variable hdebt. Hdebt is a binary variable that takes on a value of 0 if the household has no debt, and 1 otherwise. Assume that Probit provides the following estimate: $Pr(\widehat{hdebt} = 1 | wageinc, married) = \Phi(2.01 - .0274 * wageinc - 0.31 * married)$

Hdebt = 1 if household has debt, 0 otherwise
Married = 1 if married, 0 otherwise
Wageinc = wage income, measured in thousands of dollars

- a. For a married household, with a wageinc of 50, the predicted probability of having debt is (show your work):
- b. For a non-married household, with a wageinc of 50, the predicted probability of having debt is:
- c. The difference in debt probabilities between these two households, one married and the other not, all else equal, is: