EXAM 2 K Foster, Stats and Intro to Econometrics, Eco B2000, CCNY, Fall 2012	ID#:

The questions are worth 120 points. You have 120 minutes to do the exam, one point per minute.

All answers should be submitted electronically and delivered via Blackboard and email.

You may refer to your books, notes, calculator, computer, or astrology table. The exam is "open book."
However, **you must not refer to anyone else, either in person or electronically**!
You must do all work on your own. Cheating is harshly penalized.
Please silence all electronic noisemakers such as mobile phones.
Good luck. Stay cool.

1. (15 points) You might find it useful to sketch the distributions.
   a. For a Normal Distribution with mean 1 and standard deviation 9.6, what is area in both tails farther from the mean than 23.1? A. 0.1251 B. 0.0214 C. 0.4585 D. 0.9893
   b. For a Normal Distribution with mean 5 and standard deviation 7.6, what is area in both tails farther from the mean than 14.1? A. 0.2743 B. 0.1587 C. 0.2301 D. 0.4603
   c. For a Normal Distribution with mean -2 and standard deviation 3.8, what is area in both tails farther from the mean than 2.9? A. 0.7007 B. 0.1936 C. 0.3872 D. 0.2578
   d. For a Normal Distribution with mean -7 and standard deviation 5.1, what is area in both tails farther from the mean than -1.9? A. 0.3173 B. 0.0849 C. 0.6346 D. 0.9151
   e. For a Normal Distribution with mean 13 and standard deviation 3.5, what is area in both tails farther from the mean than 7.8? A. 0.2672 B. 0.1336 C. 0.1587 D. 0.7734
   f. For a Normal Distribution with mean -12 and standard deviation 9.6, what values leave probability 0.003 in both tails? A. (-2.9677, 2.9677) B. (-38.3787, 14.3787) C. (-36.1166, 12.1166) D. (-40.4903, 16.4903)
   g. For a Normal Distribution with mean -2 and standard deviation 9.1, what values leave probability 0.092 in both tails? A. (-1.6849, 1.6849) B. (-17.3330, 13.3330) C. (-1.9047, 1.4652) D. (-16.3950, 14.3950)
   h. For a Normal Distribution with mean 0 and standard deviation 4.0, what values leave probability 0.039 in both tails? A. (-5.6746, 5.6746) B. (-7.0496, 7.0496) C. (-2.0642, 2.0642) D. (-8.2567, 8.2567)
   i. A regression coefficient is estimated to be equal to -12.684 with standard error 9.4; there are 16 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero? A. 0.1960 B. 0.4810 C. 1.8228 D. 0.9323
   j. A regression coefficient is estimated to be equal to 10.030 with standard error 4.0; there are 5 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero? A. 0.2030 B. 0.3354 C. 0.0300 D. 0.0540
   k. A regression coefficient is estimated to be equal to 0.559 with standard error 0.2; there are 3 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero? A. 0.6797 B. 0.9320 C. 0.0680 D. 0.7121
   l. A regression coefficient is estimated to be equal to -3.564 with standard error 1.9; there are 22 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero? A. 0.2100 B. 0.0740 C. 0.8950 D. 0.9393

2. (15 points) A Japanese study looked at the effect of "kawaii" (cute) on test performance. (I am inferring numbers from the graphs shown so these might not exactly match the study!) There were 24 subjects; half were shown pictures of cute baby animals and half were shown pictures of full-grown animals (not cute). The ability to complete tasks, for those shown cute pictures, changed by 4, with a

standard deviation of 3.5. The ability to complete tasks, for the control group, changed by 0.5 with a standard deviation of 2.9. What is the difference in means? What is the standard error of the difference? What is the normalized value for the difference? What are the degrees of freedom? What is the p-value? Is this difference statistically significant? Comment on the study. Should you immediately Google images of cute baby animals to help the rest of your exam performance?

Nittono H, Fukushima M, Yano A, Moriya H (2012) The Power of Kawaii: Viewing Cute Images Promotes a Careful Behavior and Narrows Attentional Focus. PLoS ONE 7(9): e46362.

3. (35 points) Consider the BRFSS dataset (*I know, all the variable labels are ALL CAPS like some elderly aol user typed them, sorry*), the Behavioral Risk Factor Surveillance Study. There are many observations on a wide variety of risky behaviors: smoking, drinking, poor eating, flu shots, whether household has a 3-day supply of food and water… For now concentrate on BMI, which is Body Mass Index, a measure of whether a person is obese. Construct at least one good model to explain BMI. What are descriptive statistics for BMI? Does this suggest some way you ought to limit the sample? What explanatory variables could be in the model? Which are available in the BRFSS data? Construct at least one good regression model; discuss the estimates (including statistical significance but also relevance and whether the estimates accord with theory). How could you improve the model? Consider nonlinear age terms, gender-age interactions, race-age interactions, state dummies and more. Note that INCOME2 is not a continuous variable but you would need to create dummy variables for the relevant income levels.

4. (15 points) The BRFSS data includes information on household disaster preparedness. Two measures are whether the household has a 3-day supply of food and water. We might believe these to be similar so want to examine the marginal probabilities.

Count

		Water– 3 day supply		
		= "Yes"	= "No"	
Food – 3 day supply	= "Yes"	5528	3190	
	= "No"	484	1130	

What is the probability that a household has sufficient food and water for 3 days? Given that a household has sufficient food, what is the probability that it has sufficient water as well? If a household does not have sufficient food, what is the probability that it has sufficient water? Are the latter two proportions statistically significantly different?

5. (40 points) Continuing with the BRFSS, examine the 0/1 dependent variable of whether the person ever smoked seriously (which they define as at least 100 cigarettes); the variable is SMOKE100. We are interested in de-tangling the effects of both income and education. (Recall note from previous question about INCOM2 variable.) Estimate both probit and logit models; explain different predictions from each model. Explore various specifications possibly including interactions, dummies, etc. Carefully explain the results that you find.