EXAM 1 K Foster, Statistics and Introduction 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 put into the blue books or submitted electronically.
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.
If you do work on the computer, please submit all those files via Blackboard and email.
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. If a variable has a Standard Normal Distribution, what is the probability of observing a value less than 0.3?
    b. If a variable has a Standard Normal Distribution, what is the probability of observing a value farther from the mean (both tails) than 1.8?
    c. If a variable has a Normal Distribution with mean –4 and standard deviation 7.4, what is the probability of observing a value less than –7.7?
    d. If a variable has a Normal Distribution with mean 6 and standard deviation 5, what is the probability of observing a value greater than 7?
    e. For a Normal Distribution with mean 3 and standard deviation 9.8, what values leave probability 0.02 in both tails?

2. (15 points) A survey from eFinancialCareers found that, despite predictions from NY State that the Wall St bonus pool would drop by about 35%, a full 48% of the 911 respondents believed that their own bonuses would rise.
    a. Test the null hypothesis that the fraction of respondents expecting a bigger bonus is different from 35%. What is the p-value?
    b. Create a 95% confidence interval for the fraction expecting a bigger bonus. What is the 90% confidence interval? The 99% interval?
    c. Discuss. What other survey question might help explain this difference?

3. (15 points) Dan Ariely and co-authors report a study that asks participants to solve complicated addition tasks but gives them an opportunity to cheat: they self-report how many problems they correctly solve. Every participant got a pair of fashion sunglasses but some were told that the sunglasses were counterfeit. Forty-two people were told they got counterfeit sunglasses and 30 of them cheated; 43 people were told that they got authentic sunglasses and 13 of them cheated.
    a. Test the null hypothesis that there is no difference in the rates of cheating for people wearing counterfeit or authentic sunglasses. What is the p-value?
    b. Create a 95% confidence interval for the difference in cheating rates. What is the 90% confidence interval? The 99% interval?
    c. Discuss. Are there other factors that could be at work?

F Gino, M I Norton, D Ariely (2010). "The Counterfeit Self: the Deceptive Costs of Faking It," *Psychological Science* 21:712.

4. (15 points) An audit study emailed professors to ask for an appointment but the names of the 'students' were randomly varied to be typically male or female; white, African-American, Hispanic, Chinese, or Indian. White men were 26% more likely to get an appointment than minority women. Suppose you wanted to do a replication study for CUNY faculty. The original study emailed 6500 professors, you would like to study fewer.
    a. If the true difference in response to white male vs other names is 26%, what is the minimum sample size that could distinguish a statistically significant difference (at 95% level)?
    b. If the true difference were half as large, what is the minimum sample size to distinguish a difference?
    c. What if the true difference were just 6%? What is the minimum sample size necessary?
    d. Discuss. Can you suggest additional tests?

K L Mikman, M Akinola, D Chugh, 2012. "Temporal Distance and Discrimination: An Audit Study in Academia," Psychological Science 23:7.

5. (15 points) In recent news a study of adolescent girls compared those who had received a vaccination against HPV (a sexually transmitted virus that is linked to certain cancers) with those who had not received the vaccine. Some parents had been reluctant to get their children vaccinated because they believed this would encourage sexual activity. The study compared 493 who got the vaccine agains 905 who did not. Of the girls who got the vaccine, 61 got any of testing, diagnosis or counseling for pregnancy/sexually-transmitted disease; of those who did not get the vaccine, 76 got testing, diagnosis, or counseling.
    a. Test the null hypothesis that there is no difference between outcomes for those who got the vaccine versus those who did not. What is the p-value for this difference?
    b. Create a 95% confidence interval for the difference between the groups. What is the 90% confidence interval? The 99% interval?
    c. Discuss why this difference might be observed and how someone might critique the study.

R A Bednarczyk, R Davis, K Ault, W Orenstein, S B Omer (2012). "Sexual Activity-Related Outcomes After Human Papillomavirus Vaccination of 11- to 12-Year-Olds," Pediatrics.

6. (20 points) Use the Fed SCF 2010 data (available from Blackboard). This is the Survey of Consumer Finances, which is not representative (without using the weights, which you need not do for now) – it intentionally oversamples rich people to find out about their finances. Concentrate for now on the variable "SAVING" (about the 100th variable in the list) which is the amount that people have in their savings accounts.
    a. Test the null hypothesis that there is no difference between people who are older or younger than 65. What is the p-value for this test?

7. (25 points) Use the ATUS data (available from Blackboard) on the time that people spend in different activities. Construct a linear regression explaining the time that people spend on enjoyable activities (t_enjoy which includes most of the T12 items). Restrict the data to include only those people spending a non-zero amount of time on such activities.
    a. What are likely to be some of the most important determinants of time spent on enjoyable activities? Which of these are in the ATUS data? Should the person's wage be included (do you think income or substitution effect would dominate)? What are some important determinants, that you could imagine a survey measuring, that are not in the ATUS data? You might find descriptive statistics for the included variables.
    b. Carefully specify and estimate a linear regression. What are the statistically significant coefficients? Which explanatory variables are most important? Are there surprises? Discuss your results. (You might want to estimate more models or create additional variables.)