

K Foster, Statistics and Introduction to Econometrics, Eco B2000, CCNY, Fall 2013

> 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. No need to put your name, just last 4 digits of ID to identify yourself, so grading is blind. 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.

> > Exam has 3 pages.

- 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 -0.5?
 - b. If a variable has a Standard Normal Distribution, what is the probability of observing a value farther from the mean (both tails) than 1.6?
 - c. If a variable has a Normal Distribution with mean 2 and standard deviation o.8, what is the probability of observing a value less than 0.2?
 - d. If a variable has a Normal Distribution with mean 10 and standard deviation 8.7, what is the probability of observing a value farther from the mean (to both sides) than -5?
 - e. For a Normal Distribution with mean -8 and standard deviation 4.2, what values leave probability 0.026 in both tails (combined)?
- 2. (15 points) I tracked down this reference from a sign on the bus, from Tobacco Free NY. A survey of 1681 adolescents (age 11-14) in California asked if they had tried smoking and how often they went to convenience, liquor, or small grocery stores. The study finds that 452 kids rarely went to these stores and 81 had tried cigarettes; 458 kids visited these stores often (more than twice a week) and 133 had tried cigarettes. The authors assert that visiting these stores exposed the kids to more tobacco advertising.
 - a. What is the difference in means?
 - b. What is the standard error of the difference in means?
 - c. Is this difference statistically significant? What is the p-value? Explain.
 - d. The kids were also asked if their grades were likely to be at the level of B or below; 52 of the rare-frequency kids had below-average grades, while 63 of the high-frequency visitors had below-average grades. Is this difference statistically significant?
 - e. When asked about how often they had seen tobacco advertising, low-frequency visitors reported a mean of 3.1 (with standard error of 0.8) on a scale of 1-4 where 4 means "often,"; high-frequency visitors reported a mean of 3.4 (with standard error of 0.8). Is this difference statistically significant?
 - f. Discuss the study; what else might you add? Hendrick, L, N C Schleicher, E C Feighery, and S P Fortmann, (2010). "Longitudinal Study of Exposure to Retail Cigarette Advertising and Smoking Initiation," *Pediatrics*.

- 3. (15 points) You might have missed this in the news about Alice Munro winning the Nobel, but there was a study done, showing that reading literature such as Munro and Chekov tended to make people score higher on psychological tests of Affective Theory of Mind. Consider the difference between two groups of people: either they read from a selection of literary fiction or they read non-fiction articles about non-human subjects (e.g. potatoes). They were all given a test to determine how well they could identify emotion from a picture of a person's eyes. (I'm making up some of these numbers.) The Fiction group tests at 25.6 with standard deviation of 4.38; the Non-fiction group tests at 23.5 with standard deviation of 5.17. There were 41 people in the first group and 45 in the second group.
 - a. What is the difference in means?
 - b. What is the standard error of the difference in means?
 - c. Is this difference statistically significant? What is the p-value? Explain.
 - d. In another test, people read either literary fiction (that had won awards) or pop fiction (ie good sales but no awards). The lit fiction group scored 26.1 with standard deviation of 5.43 while the pop fiction group scored 23.7 with standard deviation of 5.08. Is this difference statistically significant?
 - e. Discuss the study, both in strengths and limitations.

Kidd, D C, and E Castano, (2013). "Reading Literary Fiction Improves Theory of Mind," Science.

- 4. (25 points) After the Nobel Prize awards to Fama, Hansen, and Shiller, we look at predictability of stock returns, using data on stocks in the S&P500. There are some days where many of these company's shares have negative returns; other days where many have positive. In 2012, more than 70% of the companies had positive returns on about 25% of the days; on another 25% of the days fewer than 30% had "up" returns. On the days following "70% up" days, the average return was .06 percent, with standard deviation of 1.72; on days following "30% up," the average return was .10 percent, with standard deviation of 1.66. There were 65 days of 70% or more up; there were 59 days of 30% or fewer up.
 - a. (1 pt) What is the difference in means?
 - b. (2 pts) What is the standard error of the difference in means?
 - c. (2 pts) Is this difference statistically significant? What is the p-value? Explain.
 - d. (20 pts) Using the data given on Blackboard, specify more hypotheses about stock behavior and test these.

SPSS data has first column as the fraction that were up on a given day, the the other columns are the returns for each company. R workspace has "ret_m" the matrix of returns, where each row is a date (given by m_dates) and each column a company (given by co_txs), then the variable frac_up is the fraction that were up on that date. The workspace data.frame sp500data has price, volume, etc in a big file.

5. (30 points) With the NSA spying revelations, we return to questions of whether there is wage discrimination against people with ancestry from the Middle East or North Africa (MENA). I've created program in SPSS syntax and R that you can run, which will define MENA_ANC if the person's ancestry is from MENA (except Israel) or MENA_BPL if the person's birthplace is MENA. You should consider whether there are differences in wages and incomes between people from the MENA or others; of course one decision to make is who is a relevant comparison group. Calculate averages between groups, considering also things like education; which are statistically significant? Explain in detail.

SPSS syntax is in MENA_coding.sps; R program to load data and create the variables is load_IPUMSdata.R

- 6. (20 points) You are comparing two groups: the first has X=0 and Y=1 and Y=3; the second has X=10 and Y=9 and Y=7. [So there are four data points: (0,1), (0,3), (10,9), (10,7).]
 - a. What is the difference in means between the groups?
 - b. What is the standard error for the difference in means?
 - c. Is the difference in means statistically significant?
 - d. What is the slope of a regression line fitted to the four points?
 - e. What is the standard error of the slope?
 - f. Is the slope statistically significant?
 - g. If the Y-observations in the second group were bigger, they might test as significant for the difference in means. If the second group were (10,9+A) and (10,7+A), what value(s) of A would make the difference significant?
 - h. Now if the values of Y were changed to (10,9+B) and (10,7+B), what value(s) of B would make the slope significant?
 - i. What if, instead, the X-values were changed by (10+C,9) and (10+C,7) what value(s) of C would make the slope significant?
 - j. Would changing X values change the estimate for the difference in means? Explain.