ID:

# Exam 1

**Econ B2000, MA Econometrics**
**Kevin R Foster, CCNY**
**Fall 2014**

The questions are worth 120 points. You have 120 minutes to do the exam, one point per minute.
All answers should submitted electronically. Please submit all relevant computer files.
No need to put your name, just last 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.
Please silence all electronic noisemakers such as mobile phones.
Good luck. Stay cool.

1. (25 points) Consider the following simple distribution questions. Your answer can be as detailed as necessary. You might sketch each case.
    a. For a Normal Distribution with mean 2 and standard deviation 5.6, what is area to the right of 10.4?
    b. For a Normal Distribution with mean -1 and standard deviation 4.9, what is area to the right of -7.86?
    c. For a Normal Distribution with mean 10 and standard deviation 5.2, what is area to the left of 0.12?
    d. For a Normal Distribution with mean 7 and standard deviation 4, what is area to the left of 9.8?
    e. For a Normal Distribution with mean 11 and standard deviation 8.4, what is area in both tails farther from the mean than 19.4?
    f. For a Normal Distribution with mean -8 and standard deviation 0.4, what is area in both tails farther from the mean than -7.52?
    g. For a Normal Distribution with mean 11 and standard deviation 6.3, what is area in both tails farther from the mean than 0.29?
    h. For a Normal Distribution with mean 10 and standard deviation 1.5, what is area in both tails farther from the mean than 12.7?
    i. For a Normal Distribution with mean -1 and standard deviation 1.2, what is area in both tails farther from the mean than 1.52?
    j. For a Normal Distribution with mean 9 and standard deviation 4.9, what values leave probability 0.08 in both tails?
    k. For a Normal Distribution with mean 3 and standard deviation 9.6, what values leave probability 0.311 in both tails?
    l. A regression coefficient is estimated to be equal to 8.25 with standard error 7.5; there are 35 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?
    m. A regression coefficient is estimated to be equal to 10.08 with standard error 5.6; there are 7 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?
    n. A regression coefficient is estimated to be equal to -12.04 with standard error 8.6; there are 14 degrees of freedom. What is the p-value (from the t-statistic) against the null hypothesis of zero?

2. (20 points) A recent report asserted that people who worked more hours also tended to be fatter (among those in certain occupations). (The paper doesn't give precise numbers so I'll make them up – don't bother with Google.) The paper did much more econometric analysis of course. Nevertheless, suppose that, of the 7219 women working non-strenuous occupations, 23% are working more than 40 hours/week. Of those women in non-strenuous occupations working more than 40 hours/week, 27.3% were obese; of those women in non-strenuous occupations working less than 40 hours/week, 24.6% were obese. There were also 714 women in strenuous occupations with 21% working more than 40 hours/week. Of the women in strenuous occupations working more than 40 hours/week, 28.1% were obese while 37.4% were obese among those working fewer hours. Does it seem likely that overtime makes certain groups more likely to be obese? J Abramowitz, "Working Hours, Body Mass Index, and Health Status: A Time Use Analysis"

3. (20 points) I used the CEX data to look at the fraction of spending going to health insurance. I get the following table, grouped by education of the reference person:

| %Insurance | No HS | HS diploma | Some college, no degree | Assoc degree | Bach degree | Adv degree |
|---|---|---|---|---|---|---|
| less than 10% | 467 | 1385 | 1191 | 615 | 1181 | 521 |
| 11% - 20% | 82 | 231 | 157 | 71 | 122 | 58 |
| 21% - 30% | 21 | 65 | 27 | 10 | 32 | 7 |
| more than 30% | 8 | 18 | 14 | 1 | 3 | 2 |

   a. Conditional on the reference person having a college degree (Associate's, Bachelor's or Advanced), what fraction devote more than 20% of spending to health insurance?
   b. Conditional on the reference person having less than a college degree, what fraction spend more than 20% on health insurance?
   c. Is this difference statistically significant?
   d. What is the overall share (in this sample) of people with any college degree? What share of people spending more than 20% is made up of people with any college degree?
   e. Are those break points (+/- 20%; any degree) reasonable? Can you suggest better? Explain.
   f. What problems might there be, with the classification and analysis here? Can you do better with the CEX data?

4. (10 points) One of the first notes in class emphasized "know your data." You've done a few homework assignments using Consumer Expenditure Survey data, tell me about that data. How do they calculate expenditure on food away from home?

5. (20 points) I will consider a simple question of the relation of employment to production – relevant both for questions of "jobless recovery" and worker productivity. In the R dataset, "`macro_data1.Rdata`", I give monthly data for the US on payroll (total non-farm), the unemployment rate, and an index of industrial production for the period from February 1948 to August 2014. There is also a dummy variable for when the US was in a recession (as defined by NBER). The dataset has both the level of each of these (denoted `lvl_`) and log difference (denoted `ld_`), where $ld\_z(t) = log\big(lvl\_z(t)\big) - log\big(lvl\_z(t-1)\big)$. You can use the command, `load("macro_data1.RData")`, to get the data in. I estimate the following regression for the period from 2000-date:

$$\frac{dPayroll}{Payroll} = .000739 + .0512\frac{dProduction}{Production} - .00270Recession$$

   The intercept coefficient has standard error of .00011, the slope coefficient on percent change in production has standard error or .0161, and the Recession dummy has standard error of .0003. The R-squared is 0.4943.

a. What is the t-statistic for the slope coefficient? What is its p-value? (Carefully specify the null hypothesis.) What is a 95% confidence interval for the slope coefficient?
b. Suppose that next month (not a recession month), the percent change in production is 0.004 – what would the regression predict is the percent change in payroll?
c. How would you critique this regression? What might be improved?
d. Can you find some other interesting results from the data given? Explain.

6. (25 points) Using the CPS dataset, form a regression explaining the amount of wage and salary earned by workers. Consider the question, does it pay to get a college degree? What do the data tell us about this?