PRINCETON UNIVERSITY
Woodrow Wilson School

To: Students Taking the QE1
From: Robert Willig, Faculty Chair
Date: March 23, 2004
Re: Hints on the QE1

Three former students in our program -- Greg Felker, Varun Guari, and Meg Murray -- were once asked by the MPA office to set out guidance for prospective takers of the QE1. Their memo (which has been modified very slightly since they wrote it) is reprinted below and contains excellent advice.

I would, however, like to add a note of interpretation. The memo refers to statistical, economic, and political "components" of the QE1. (Since the memo was written before the introduction of WWS 502, there is no discussion of a psychology component.) Strictly speaking, the QE1 does not have "components." Rather, it is read separately for the knowledge shown and quality of the statistics, economics, politics, and psychology incorporated in the written work.

The ideal of the QE1 is a piece of applied work that seamlessly incorporates the analytical approaches emphasized and taught in the first year core curriculum. That said, the QE1 is an exam, and, in the interest of making clear your mastery of those approaches, you are well advised as a matter of exam strategy to make sure it will be evident to the readers, even at the price of the seams showing.

Two other memos are also attached. Jim Laity, who taught statistics and did a number of QE1 review sessions for us in the past, wrote up a memo on How to Do Statistical Analysis on the QE1. Last year's students found this memo helpful and suggested that you might as well. And finally, another memo which was generated by questions in last year's statistics review session demonstrates examples of calculating z scores and running a very simple z test on comparing a sample mean with a population mean.

I urge you to prepare thoughtfully for the QE1. Attend the "practice/review" sessions. And do try to enjoy what your examiners try very hard to make an interesting experience.

Best wishes to you all.
MEMORANDUM

TO:    First Years Taking the QE1
FROM:  Greg Felker, Varun Gauri, Meg Murray
RE:    Hints for the QE1

Below are some comments, hints, and suggestions to help you prepare and master the QE1. These are not the definitive last word on taking the QE1, but merely our humble suggestions drawn from our own QE1 experience.

1.  "Stick to the knitting": The memo's structure and clarity of organization matters a lot. There are almost always questions, posed by your fictional boss in the IPE materials, which encapsulate the basic analytic issues the faculty want you to cover. If the format of your memo differs substantially from the way these questions are presented, make sure that the answers to them emerge clearly in your text.

2.  "Focus on the bottom line and the analytics": The point of the exercise is to see you apply the economic, political, and statistical analytics to the fictional problems. Thus the focus is on "how" you back up your answer. Summarizing and building up to an answer with background info is good, but stating the bottom line early and then showing how you picked it and discarded the opposing views should get more weight. Invoke as much as possible the analytic language and concepts taught in the core courses to support your arguments.

3.  "Don't get mugged by reality." The background materials include all the information, often including the pros and cons of major policy arguments that you will need. While outside knowledge might round out your answer, make sure you evaluate and use the arguments contained in the background materials. Similarly, the premium is on arguing a particular recommendation well, making good use of analytic concepts. A wholly original answer to the problem, no matter how creative, is not the central goal of the exercise.

3.  "Look for the analytic 'nuggets'." The fictional situation is designed to motivate the application of economic, statistical, and political analytics. The faculty embed certain questions or problems in the briefing materials and include opposing arguments about them. Again, these are usually highlighted in the questions posed by your fictional boss.

4.  "Separate but equal." - The QE1 is an integrated exercise in the sense that your final recommendation draws jointly on political, economic, and statistical analysis. However, there are indeed economic, statistical, and political components which are graded separately, and you need to make sure you have a good balance in your memo.

FOR THE STATISTICAL COMPONENT:
- Make sure you include t-tests and F-tests, if applicable.
- Use the model to predict an outcome and judge its validity.
- Consider non-technical aspects of the analysis, like sample bias, omitted variable bias, or the appropriateness of the model to answer the question.
- Remember the 11th commandment - "Correlation does not prove causation!"
FOR THE ECONOMICS COMPONENT:
- Graph the economics "nugget(s)" in the appendix.
- Analyze both the micro- and macro-economic aspects of the problem.
- The problems often involve concepts such as elasticity, dead weight loss, consumer surplus, efficiency, public goods, and externalities. You might want to review some of these concepts, and try to use such "buzz words" in the memo.

FOR THE POLITICAL COMPONENT:
- Analyze the politics on at least three levels:
  
a) The politics of the memo - where do you, as a fictional character, stand vis a vis other relevant actors, including your boss, or whoever the memo's audience is? Where does your boss stand; what is her interest in this issue?
  
b) The political pros and cons of each policy recommendation. Most people do this well, and factor them into their final recommendation.
  
c) The strategic politics - how can your recommendation be packaged and sold to the relevant players and institutions? E.g. how will your recommendation be sold to the public, or put through Congress, or sold to the other Federal bureaucracies, etc?

GOOD LUCK!
HOW TO DO STATISTICAL ANALYSIS ON THE QE1

Jim Laity
April 1996

For each statistical table or regression analysis presented, you need to answer three types of questions:

1. What do the results purport to show?
2. Has the author correctly interpreted them?
3. Are there confounding factors which limit the applicability of these results to the problem under consideration?

For regressions, the answer to the first question can always be stated as follows:

A one [x-unit] increase in [x var being considered] is associated on average with an estimated [coefficient] [y-unit] [increase/decrease] in [y var], controlling for [all other x vars].

For the regression in the blue memo from the Thomas Malthus Foundation (in the Third World Debt exercise) this would become:

A one percentage point increase in foreign aid as a percentage of GNP averaged over the period 1969-76 is associated on average with an estimated 109 1977 US $ decrease in 1977 per capita income. (Note there are no other x variables controlled for in this case.)

Note several features of this generic interpretation:

-- A change in the x var is associated with a change in the y var; this does not say anything about the direction of causation. Could be x to y, y to x, or z (some other omitted var) to both x and y (see below).

-- The relationship between x and y holds only on average, for any individual case there will be some deviation from the predicted association. This is called the random component of the process being modeled and is represented by the residual for that case.

-- The numerical coefficient is only estimated; if we took a different set of data points governed by the same underlying process, we would get at least a slightly different estimate. This is because the random component of each data point affects the estimate. To determine how accurate the estimate is, we need to look at the standard error (see below).

-- The estimated relationship is based on controlling for (or "holding constant") all other x variables. For example, suppose we find that on average women make less than men, but it is also true that on average women have less job experience than men, which might account for the difference. If we include among the x variables both gender and years of experience (with wage as the y variable), then the estimated effect of gender will have
controlled for the difference in experience between the two groups. That is, it is as if we had compared the wages of groups of men and women with identical experience, even though this is not actually the case in our sample.

To answer the second question for regressions, you should consider the following:

**Variables:** What is dependent (y) variable? What units is it measured in? What are independent (x) variables? Units? Be sure you know the differences between:

- **numeric** variables
- **dummy** variables
- **categorical** variables (group of dummies with one missing category; effect of each variable should be interpreted relative to the omitted category)
- **interaction terms**
- **constant** (prediction when all independent variables are zero, *if this makes sense* in the context of the model, *and* if such data points, with all independent variables zero or close to zero, were included in the range of data used to estimate the regression)

**Sign:** What is the sign of the effect of each independent variable? Does it support or refute whatever the author is trying to claim?

**Size:** What exactly does the coefficient mean in practical terms? Is the estimated effect large or small from an economic or policy point of view? NOTE ON UNITS:

- **X and Y both in natural units:** "a one unit increase in x is associated with a b unit change in y, on average, holding all other variables constant."

- **X in natural units, Y in logs:** "a one unit increase in x is associated with a b*100 percent change in y, on average etc." Reminder, this is an approximation, acceptable for changes less than 15% (.015). For changes greater than 15%, use the formula: \((e^b - 1)\times100\%\) to calculate the percent change.

- **X and Y both in logs:** "a one percent increase in x is associated with a b percent change (not a b*100 percent change) in y, on average etc."

- **Dummy/Categorical variables:** a "one unit" change in a dummy variable means the change associated with being in, as opposed to not being in, the corresponding class.

**Significance:** Which effects are statistically significant (ie: not likely to be due simply to random variation)? Use t test:

\[ t_{n-k} = (b - \mu_0) / SE_b \] (where k is the number of regressors, including the intercept)
Be sure you know the difference between a one-tailed and a two-tailed test and when to use each. A one-tailed test is easier to "pass" (i.e.: a less strict test of statistical significance); it should be used only when you can rule out a priori that the effect goes in a particular direction, or you are interested in proving an effect in one direction only. For example, if you believe you can rule out in advance that there is wage discrimination in favor of African Americans, you could use a one-tailed test of the coefficient on race in a wage equation to decide if there was evidence of negative wage discrimination. Or, if you are only interested in demonstrating discrimination against African Americans, and would not have drawn any conclusion if your results had appeared to show reverse discrimination in their favor, a one-tailed test of negative discrimination would also be appropriate. In contrast, if you are not sure whether the effect of a particular drug is to increase or decrease the incidence of cancer, and you would like to know what the evidence shows either way, a two-tailed test is appropriate. When in doubt, use the more conservative two-tailed test.

Reminder: Stata always tests \( H_0: \beta = 0 \) using a two-tailed test. You must calculate your own t statistic/p-value if you wish to use a one-tailed test, or a null hypothesis other than \( \beta = 0 \). You must decide this from the context of the problem. The p-value for a one-tailed test is always one half that for a two-tailed test. The table in your book gives one-tailed p-values; to get two-tailed p-values you must double them. To get one-tailed p-values from Stata output, take half of the (two-tailed) p-value shown in the listing.

Note: You don't need to bother interpreting sign and size for variables that are not even close to statistically significant. As a general rule of thumb, if t is close to or greater than 2, than the result is statistically significant. Another way of stating this is that you can be 95% confident that the true underlying coefficient governing the model is within approximately two standard errors of the estimated one, so if the range \( b \pm 2SE_b \) does not include zero, it is unlikely that the estimated effect is due simply to random variation (i.e.: that the true effect is 0). This is equivalent to saying that the estimated effect is statistically significant. Be sure you are clear on the distinction between statistically significant, and economically significant, or significant from a policy point of view. The former refers to a low probability that the estimated effect is due only to chance; the latter to the practical consequences of the effect for policy decisions.

\( R^2 \): percent of the total variation in y explained by the regression.

**F-test for single regression:**

\[
F_{k-1,n-k} = \frac{(R^2/(k-1))}{((1-R^2)/(n-k))}
\]

This tells you if the total amount of variation explained by the model is significant (i.e.: if it is, you can reject \( H_0: \beta_1 = \beta_2 = ... = \beta_n = 0 \)).

Reminder: \( TSS = ESS + SSR \) and \( R^2 = ESS/TSS = 1 - SSR/TSS \), so another version of the F-statistic is:

\[
F_{k-1,n-k} = \frac{(ESS/(k-1))}{(SSR/(n-k))}
\]

**F-test for diff between two nested regressions:**
F_{ku-kr,n-ku} = \frac{\left( R^2_u - R^2_r \right) / (k_u - k_r) }{ \left( 1 - R^2_u \right) / (n - k_u) }

or

F_{ku-kr,n-ku} = \frac{\left( ESS_u - ESS_r \right) / (k_u - k_r) }{ SSR_u / (n - k_u) }

Tests hypothesis that the additional coefficients in the unrestricted (fuller) specification jointly equal zero (i.e., no new explanatory power is added to the model by the additional variables in the fuller model).

**Note:** The coefficient estimate for each x variable depends on all the x variables included in the model (unless they are all perfectly uncorrelated, which is very unlikely). Thus, you **cannot** combine coefficient estimates taken from different models as if they were comparable, since they are based on controlling for different things; the analysis in the Third World Debt exercise erroneously tries to do this.

You should also know the formulas for testing null hypotheses about single means or proportions:

\[ t_{n-1} = \frac{\bar{X} - \mu_0}{SE} \quad \text{where} \quad SE = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} \]

\[ z = \frac{P - \pi_0}{SE} \quad \text{where} \quad SE = \sqrt{\frac{P(1-P)}{n}} \quad \text{assuming} \ n \ \text{large (greater than 30)} \]

Note that especially for proportions, the null value you are testing against may not be zero. For example, to see whether polling data shows that a particular candidate will win an election, you would want to test \( H_0: \pi \leq .5 \). The distinction between one and two tailed tests also applies to hypothesis testing of means and proportions.

You should also know the formula for checking the statistical significance of the **difference between two means**. The basic form of this formula is:

\[ t_{df} = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{SE_{diff}} \]

Here, it is usually the case that \( \mu_2 - \mu_1 = 0 \) since you are testing the null hypothesis that there is no difference between the underlying means.

You should be able to distinguish the following three situations:

---

**Comparison between whole samples (not nec the same size) where it is reasonable to assume the variance within each sample is the same (pooled estimation);** in this case the formula for \( SE_{diff} \) is:

\[ SE_{diff} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \text{where} \quad s_p^2 = \frac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2} \]

and the df for this t statistic is \( n_1 + n_2 - 2 \).
comparison between whole samples (not nec the same size) where it is not reasonable to assume the variance within each sample is the same (non-pooled estimation)

$$SE_{diff} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$ where $s_i^2 = \frac{\sum(X_i - \bar{X}_i)^2}{n_i - 1}$

and the df for this t statistic is the minimum of $n_1 - 1$ and $n_2 - 1$

comparison between pairs of individuals who have been matched (as much as possible) on variables not of interest to the analysis (matched-sample); here the two samples must be of the same size since there is a one-to-one correspondence between them, and you perform your t-test on the null hypothesis that the within-pair difference is zero, using the statistic:

$$t_{n-1} = \frac{D - D_0}{SE_D}$$ where $n$ is number of pairs

and the SE of the differences is computed just like any other single sample SE.

You also need to know the formula for checking the difference between two proportions:

$$z = \frac{(P_2 - P_1) - (\pi_2 - \pi_1)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

Again, $\pi_2 - \pi_1$ will usually be zero, since you are usually testing the null that there is no difference between the proportions. Note that you use a z test; this requires the n's to be "large"; generally greater than 30 is large enough.

To answer the third type of question, pay particular attention to:

Direction of causality: Does x "cause" y or does y "cause" x? Or is there another variable omitted from the analysis which "causes" both x and y. This is no longer a question of math, but of intuition. Which interpretation seems more reasonable? Is the author getting it backwards? In the blue memo from the Malthus Foundation it is claimed that foreign aid "causes" low GNP. Might it be the other way around?

Omitted variables: If relevant variables are omitted which are correlated with the included x variables, estimates of the coefficients for these included variables will be biased. To determine the sign of the bias so you can judge whether it strengthens or weakens the apparent effect:

sign of bias = sign of omitted coef * sign of omitted/included corr

Thus for example, if the omitted variable has a positive effect on the dependent variable and the included and omitted variables are positively correlated with each other, this will produce a positive bias in the estimated coefficient for the included variable, since some of the effect of the omitted variable is being attributed to it; that is, the estimated coefficient is too high. This means the effect is overstated if it is positive, and understated if it is negative (a "higher" negative number has a smaller absolute value). If the omitted variable has a positive effect but is
negatively correlated with the included variable, this will produce a negative bias in the estimated coefficient for the included variable, since some of the effect of the omitted variable is countering it; that is, the estimated coefficient is too low. This means the effect is understated if it is positive, and overstated if it is negative.

To give a concrete example, if being a woman and years of job experience are negatively correlated, and years of job experience tends to increase wages, then if only gender is included in a wage regression, the bias on its coefficient will be negative (negative * positive); that is estimated coefficient will be too low. If the direct effect of being a woman on wages is negative, the estimated effect is overstated (too large a negative number) since some of the negative effect of less job experience is being attributed to gender. This is why it is important to use multiple regression and include as many relevant variables as possible to control for their effects.
Mechanics of Z Scores

1) For 5459 pregnant women using Aarhus University Hospital in Denmark in a two-year period who reported information on length of gestation until birth, the mean was 281.9 days, with standard deviation 11.4 days. A baby is classified as premature if the gestation time is 258 days or fewer. If gestation times are normally distributed, what proportion of babies would be born prematurely?

2) According to *Current Population Reports*, self-employed individuals in the United States in 1990 worked an average of 44.6 hours per week, with a standard deviation of 14.5. Assume this variable is approximately normally distributed; find the proportion who averaged more than 40 hours per week.

3) The mean score for all High School seniors taking a college entrance exam equals 500. A study is conducted to see whether a different mean applies to those students born in a foreign country. For a random sample of 100 such students, the mean and standard deviation on this exam equals 508 and 100.

   a) Set up hypothesis for a significance test.
   b) Compute the test statistic
   c) Report the p-value, and interpret
   d) Based on (c), can you conclude that the population mean for students born in a foreign country equals 500? Why or why not?
   e) Make a decision about $H_0$, using $\alpha = .05$
   f) Construct a 95% confidence interval for $\mu$. Show the correspondence between whether 500 falls in the interval and whether $H_0: \mu = 500$ is rejected in favor of $H_a: \mu \neq 500$. 
Answers

1) 258 days have a z-score of $z = (258-281.9)/11.4 = -2.10$; proportion born prematurely are those with z-scores to the left of -2.10, which is .018

2) 40 has a $z$-score of $(40-44.6)/14.5 = -.32$; the left tail probability below this is .3745, so the probability above 40 is $1 - .3745 = .6255$

3) 
   a. $H_0: \mu = 500$, $H_a: \mu \neq 500$.
   b. $\hat{\sigma}_f = \sqrt{\frac{100}{100}} = 10.0$, $z = (508 - 500)/10.0 = .80$
   c. $P = 2 (.212) = .424$; the data do not contradict $H_0$.
   d. No, it does not make sense to “accept $H_0$” and conclude that the parameter does equals the $H_0$ value. It is simply plausible that $\mu = 500$, but this is only one of many plausible values, as the confidence interval in (f) shows
   e. Since $P = .42 > .05$, we cannot reject $H_0$.
   f. $508 \pm 1.96 (10) = 508 \pm 20$, or (488, 528). 500 is in the 95% confidence interval, and we do not reject $H_0$ at the .05 level.