INTERNATIONAL UNIVERSITY OF JAPAN
Public Management and Policy Analysis Program
Graduate School of International Relations

Guidelines for Quantitative Data Analysis

Wednesday, April 25, 2012

Hun Myoung Park

1. General

Organize ideas systematically. Highlight key research questions. Explain why your research is needed and valuables and how your research differs from other studies. You may mention your questions in Introduction and/or literature review chapters.

DO NOT omit necessary citations; otherwise, you are stealing others' ideas and research. This must be considered academic dishonest and misconduct including plagiarism, fabrication (falsification), and collaboration (cheating).

2. Organization

2.1 Data Description

Provide detailed explanation about data (population, sampling process, level of representativeness, units of measurement, descriptive statistics, etc.) and explore variables to be used. Overall properties and strength (and weakness) of your data should be presented. You have to make sure you manipulate your data correctly in order to avoid funny conclusion.

2.2 Rescaling

If you are using too big numbers or too small numbers, please adjust scales appropriately in order to avoid unintended overflow or underflow in computation. For instance, if original data contain big numbers such as "24,879,574,258,578," divide them by 1,000,000 to change unit of measurement to 1 million. By contrast, multiply 1,000,000 if your numbers are like ".0000005547" and then change the unit of measurement.

2.3 Method

Be specific. Explain a specific model that you want to use. And why. Do not simply say, "I want to use a panel data model," which sounds almost the same as "I am analyzing data." Do not list all methods you are considering. Choose the best model (your model). Do not forget to include relevant citations to acknowledge scholars who developed the model you want to use unless you developed the model by yourself.

2.4 Analysis and result

Report the results in a professional manner. See Tables and Figures section below. You should report goodness of fit measures such as R2 and chi-squared for the likelihood

ratio test in order to show that your model fits the data well. If your model does not fit the data, DO NOT try to interpret coefficients; your model is wrong or effects of IV are zero.

2.5 Interpretation

Interpret the result in substantive way. DO NOT simply report signs and magnitudes of parameter estimates (coefficients). Your boss or other audience in general does not know exactly your model; they are not econometricians. You should provide intuitive interpretation. Instead of saying, "If you increases one percent of an IV (investment), GDP (DV) will increase (or decrease) 3 percent," (As a methodologists, I do not know exactly what this interpretation means) you may try this way: "If you invest 1 million US dollars more from current status, the GDP is expected to increase by 2.5 billion US dollars, assuming you do not change any others factors other than investment." You are responsible for reporting data analysis clearly and present the results in substantive manner; you should be able to appeal to the audience, who oftentimes tends to have poor knowledge to understand your model . In addition, when you take a logarithmic transformation of an IV, transform its impact back to the original scale when interpreting it.

2.6 Discussion

In addition to substantive interpretation, discuss your findings and try to make them sense before making conclusions. If your findings are quite opposite to those of existing literature, please double-check all analysis processes (from getting data through running models) and then suggest possible reasons for getting odd results.

3. Expression

Do not repeat common knowledge that other in the field already know. For example, "panel data models examine if there are fixed and/or random effects," or "GNP is computed as..." Your thesis is not a dictionary or textbook!!!

Avoid these expressions. "Regarding (…), people ..."→ Explain (…) in the sentence. and "According to World Bank, it was found that (...)" → (…) and add citation.
Minimize use of transitional phrase (e.g., on the other hands, moreover, as a result, for example). Make your paragraphs and statements flow logically.
Use % in parenthesis only; otherwise spell out (percent) in contents.

4. Common Errors

- Significant level → significance level
- at 5% level → at the .05 level
- at the level of significance α=5% → at the .05 significance level
- at the conventional level → omit this expression or use specific significance level
- P-value is significant → P-value is small (keep in mind that the p-value itself is neither significant nor insignificant)
- The null hypothesis is b1 = 0 → The null hypothesis is $\beta 1 = 0$. You already know the value of b1; what you don't know is $\beta 1$. That is, you do not need to test if b1=0.
- The coefficient of $\beta 1$ → Parameter estimates of $\beta 1$ or the coefficient of DV1.
- Accept (or confirm) the null hypothesis → do not reject the null hypothesis
- I do not believe that H0 is true → reject the H0 at the .01 level

5. References

All research and studies you cited in your thesis SHOULD APPEAR in the reference section. Otherwise, you are committing plagiarism!!! PLEASE double-check if you omit anything before finalizing your thesis.

Some References for Data Analysis

Baltagi, Badi H. 2001. *Econometric Analysis of Panel Data*. Wiley, John & Sons.

Greene, William H. 2003. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences*. Sage Publications.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

http://www.indiana.edu/~statmath/stat/all/panel/panel.pdf (Panel Data Models)

http://www.indiana.edu/~statmath/stat/all/cdvm/cdvm.pdf (Binary Logit/Probit Models)