November 30, 2003. 7 November 2006. Eric Rasmusen, Erasmuse@indiana.edu.

Common Mistakes

This is a list of the most common mistakes students made in one particular semester of papers they wrote for my industrial organization course.

1. Don't use computer codes as names or abbreviations for variables.

2. Don't put boxes around tables and figures. Do not use more horizontal and vertical lines than are useful.

3. Include a title page, an abstract, a date, and your email address.

4. Say what software you use.

5. Do not report F-statistics unless you have some reason to report them.

6. Do not report numbers to more decimal places than is interesting (i.e., say 2.13, not 2.134292022).

7. Do not use keys to graphs or charts— label the curves or regions directly, so the reader does not have to move his eyes between a key and the chart itself.

8. If you are using time series data, include time-period dummies or a time trend, since serial correlation will be a big problem. Fixed effects are also usually necessary for time-series cross-section data. Use robust standard errors.

9. Tables should explain what is going on in them. If you exclude outliers, for example, say how you define "outlier".

10. Label each regression column in a table with a number.

11. Give sources for your data.

Regression Presentation

1. Discuss in your text the size of the important coefficients, not just their significance. If the x-value changes by 10%, how much does the y-value change? You do not need to do this for all x-variables, but do it for the ones whose effects you are really interested in (as opposed to control variables that are just holding everything else constant).

2. Do not write 1.23423 when rounding to 1.23 will do just as well. Fewer digits yield greater clarity.

3. Use correlation matrices to show the simple correlations between important variables.

4. Give summary statistics. Think about which are most useful. Think about presenting the mean, median, mode, minimum, maximum, standard deviations, and number of observations. Do not present all of these–think.

5. Use words for variable names, not computer codes. "Density" is a much better name than the unpronounceable and mysterious "POPSQMI."

6. Present the coefficients, standard errors or t-statistics (not both), R^2, and number of observations. Do not present other statistics (e.g. Aikake, an F-test for all coefficients equalling zero) unless you have a reason to. Maybe use stars for significance– * for 10% level, ** for 5%, and *** for 1%.

7. If the left-hand variable (y-variable, dependent variable, endogenous variable) takes only a few values (e.g., 0 and 1) then use a special technique such as logit or tobit. If a right-hand variable (x-variable, independent variable, exogenous variable) takes only a few values, that does not create a need to use anything besides OLS.

8. If you use a technique such as logit for which the coefficient values have little meaning, do not report them in your tables. Instead, report the "marginal effects" which show how a small change in the x-value affects the y-value, evaluated at the average or median values of all the x-values. You do not need to do this for OLS or 2SLS; you do need to for logit, probit, or tobit.

FIGURES

1. Keep the data-to-ink ratio high.

2. Erase data-ink for useless data.

3. Erase non-data ink.

4. Refer to every figure in the text.

5. Don't use boxes and grids.

6. Don't use cute pictures that obscure what you are saying.

7. Consider labelling individual data points of special interest.

8. Write text horizontally, not vertically.

9. Make the figure self-contained. Don't require the reader to refer to the text or a previous table. Include the source and the units of measurement.

10. Number and title every figure.

TABLES

1. Keep the data-to-ink ratio high.

2. Leave out dividing lines and boxes unless you have a good reason for them.

3. Leave off repetitive, useless numbers.

4. Don't use just capital letters.

5. Circle or otherwise mark important numbers, in particular, ones you mention in the text or talk.

6. Make the table self-contained. Don't require the reader to refer to the text or a previous table. Include the source and the units of measurement.

7. Number and title every table. Use meaningful titles (not "Regression Results," "Regressions Explaining Profit Rates". Refer to each in the text.