REGRESSION ANALYSIS II: Linear Models

Tuesday, June 23 - Friday, July 17, 2020 — On-line (synchronous and asynchronous)

Tim McDaniel
Buena Vista University — Storm Lake, Iowa USA — mcdaniel@bvu.edu

Joint appointment: The Siebens School of Business and the School of Science & Mathematics

Introduction

This course is intended for social scientists who are comfortable with algebra and basic introductory-level statistics, and now want to learn applied ordinary least squares (OLS) multiple regression analysis for their own research and to understand the work of others.

As social scientists, it is important that we know how to use multiple regression. But it is also important for us to know why and how multiple regression works (and fails) under varying conditions. Given this, we will discuss much of the mathematical and statistical theory behind multiple regression and also some potential drawbacks and circumstantial limitations. So while the presentations will not be purely theoretical, neither will this course be “cookbook” in nature.

The primary goal of the course is to develop an applied and intuitive (as opposed to purely theoretical or mathematical) understanding of the topics. Whenever possible, presentations will be in “Words,” “Picture,” and “Math” languages.

Course Topics

We will begin with a quick overview of basic statistics and hypothesis testing (including the use and misuse of p-values). Most (but probably not all) of this material should be a review for you.

After that we will cover various topics in bivariate and then multiple regression, including:

- Model specification and interpretation.
- Diagnostic tests and plots.
- Analysis of residuals.
- Dichotomous (dummy) variables.
- Outliers, influence, and leverage.
- Logit models and analysis, and Probit models.
- Advanced diagnostic plots and graphical techniques.
- Regression models from a matrix perspective.
- Various F-tests (e.g., group significance test; Chow test; relative importance of variables and groups of variables; comparison of overall model performance).
- Categorical independent variables (e.g., new tests for “Intervalness” and “Collapsing”).
- Heteroskedasticity: Definition, consequences, detection, and correction.
- Autocorrelation: Definition, consequences, detection, and correction.
- Generalized Least Squares (GLS) and Weighted Least Squares (WLS).

- $R^2$ and Standard Error of Regression.
- Transformations to induce linearity.
- Multicollinearity.
- Multiplicative interaction terms.
- Categorical (e.g., Likert) variables.
Technology We Will Use

We will use the University of Michigan Canvas platform to disseminate Lecture videos, Lecture Note Packets, Optional Learning Exercises, data files, and other materials in this course.

Additional information and details will be provided to you before the course begins.

Lecture Note Packets

This course will utilize approximately 925 pages of Lecture Notes, which will be furnished to you as pdf files, organized in twenty Packets. These will serve as the sole required “textbook” for this course and also as an information resource for you after the course ends. These Lecture Note Packets will significantly reduce the amount of notes you have to write during each lecture, which means you can concentrate much more on learning and understanding the material itself.

A detailed outline of the Lecture Note Packets is included in this syllabus. Each of the twenty Lecture Note Packets also has a Table of Contents with page number information.

I urge you to use this year’s edition of the Lecture Note Packets, as opposed to using an earlier edition; I have made several substantive additions, deletions, and other changes this year.

Although these Lecture Note Packets are detailed, comprehensive, and self-contained, it is still advisable for you to study the relevant Packets before and after each class, to ask questions during class, and to meet with either me or one of the three Teaching Assistants (TAs) outside of class so you can maximize your learning and other benefits from this course. These Packets contain several algebraic derivations and proofs. We will not take time to work through most of them, but instead they are provided for your information and consideration outside of class.

Textbooks and Other Readings

There is no required textbook (other than the aforementioned twenty Lecture Note Packets, which constitute a de facto textbook) for this course. Later in this syllabus I provide you with information about several optional supplemental readings from various textbooks and journals, organized across each of the twenty Lecture Note Packets. See the beginning of the “Some Suggested Readings” section for a discussion of the four textbooks (each employing a different style) included in these optional readings.

I do not think that any one of these four textbooks is significantly better than the others; instead, the one(s) that you might find best will depend upon a number of personal factors, including that with which you are comfortable and familiar. Therefore, instead of just picking one textbook, I have designed the course so that you can experiment and pick-and-choose which one(s) you prefer. Of course, you may decide to read and study (and maybe purchase) none, one, two, three, or even all four of these textbooks; again, that is entirely up to you. See the alphabetized list of “Some Suggested Readings” near the end of this syllabus for the textbook readings corresponding to each of the Lecture Note Packets. Finally, other textbooks are also appropriate for this course; see me if you have any questions about this.
At the end of this syllabus is a bibliography for the textbooks and all other optional readings. Note that I have included quite a few of the Sage “Little Green Book” monographs. You may notice that some of the readings are from political science books and journals; however, this is not a cause for either rejoicing or concern since they deal with methodological topics that are easily and broadly generalizable across other disciplines. There also reading and examples from areas such as business, criminal justice, economics, law, sociology, and social science in general.

Optional Learning Exercises and Software

You will receive seven Optional Learning Exercises, each at an appropriate time. These are not submitted for grading and do not count toward your course letter grade; they truly are “optional.” I will also distribute an extensive and comprehensive “Tutorial and Answer Key” for each one. Each of these will provide you with another excellent learning opportunity in this course.

The first Optional Learning Exercise covers some basic regression terminology, notation, and concepts. Then for the remaining six Optional Learning Exercises you use data to generate computer output; that part will be relatively trivial (and not necessary, since the output is also provided in the accompanying “Tutorial and Answer Key” document. Then your substantive learning task will be to interpret, analyze, and explain that output.

The Teaching Assistants (and I) are fully available to provide you with advice, clarification, discussion, assistance, etc., as you use the Optional Learning Exercises as a way to learn.

This is a course on Regression Analysis, NOT on statistical software. The emphasis, including in the Optional Learning Exercises, is on understanding and interpreting (not generating) output.

• Since it is so quick to learn and easy to use, SPSS is used to generate all in-class output.
• Relevant output is always included in the “Tutorial and Answer Key” documents...
• ...But you can instead use any software you choose to generate your own output!
• The TAs are primarily a resource for assistance in understanding and explaining substantive material and interpreting output, but also potentially for basic help generating output using SPSS, R, and Stata.

The data files for all of the Optional Learning Exercises, and also for each of the in-class examples, will be furnished to you via the University of Michigan Canvas platform. In addition to SPSS format, I will also give you access to each data file in Stata format.

Matrix Algebra

I have designed the course so that matrix algebra will not be used during the first three weeks. Instead, you will learn as much of the regression material as possible within a simple scalar algebra environment. Packet 15, which you can study during the weekend between the third and fourth weeks, consists of an introduction to matrices and matrix algebra. Then parts of some of the presentations covering the more advanced topics in the fourth week will involve matrices.

Many of you should consider attending the “Mathematics for Social Scientists II” lectures.
Lectures: Asynchronous Videos

Lecture video files (in mp4 format) will be furnished asynchronously through the University of Michigan Canvas platform. Each day’s lecture video files will be approximately (give-or-take...) a total of two hours or so in length, including a short “Introduction” video and also a short “Wrap-up” video file.

Supplemental document (pdf) files will be furnished through the University of Michigan Canvas platform and discussed in that day’s Wrap-Up video file.

Here is the Lecture Packet (and video files) coverage schedule for each of the lectures:

Tuesday, June 23 — Introduction and the Syllabus – Packet 1, pages 1-33
Thursday, June 25 — Packet 3, pages 22-End – Packet 4; pages 1-22
Friday, June 26 — Packet 4, pages 23-End – Packet 5 pages 1-End

Monday, June 29 — Packet 6, pages 1-End
Tuesday, June 30 — Packet 7, pages 1-End – Packet 8, pages 1-16
Wednesday, July 1 — Packet 8, pages 17-End – Packet 9, pages 1-13A
Thursday, July 2 — Packet 9, pages 13A-End – Packet 10, pages 1-2J
Friday, July 3 — No Class (“Fourth of July” holiday)

Monday, July 6 — Packet 10, pages 3-End – Packet 11, pages 1-11 & 57-End
Tuesday, July 7 — Packet 11, pages 12-56.5
Wednesday, July 8 — Packet 12, pages 1-End – Packet 13, pages 1-2 & Review
Thursday, July 9 — Packet 13, pages 2-12I
Friday, July 10 — Packet 13, pages 13-End – Packet 14, pages 1-21

Monday, July 13 — Packet 14, pages 21-End – Packet 15, pages 1-End
Tuesday, July 14 — Packet 16, pages 1-End
Wednesday, July 15 — Packet 17, pages 1-End – Packet 18, pages 1-6
Thursday, July 16 — Packet 18, pages 7-End
Friday, July 17 — Packet 19, pages 1-End – Packet 20, pages 1-End

The video files will be posted prior to each listed lecture date.

Remember, each Lecture Note Packet is available (as a pdf file) through the University of Michigan Canvas platform. I strongly recommend that you have access to, and follow along with, the appropriate Packet as you watch and learn from each Lecture video.
Questions and Discussions: Synchronous Zoom Sessions

We will be using synchronous Zoom sessions, reached through the University of Michigan Canvas platform, for questions, discussion, etc., about the material in the lecture videos and also about anything else about Regression II topics, or the course in general, that you wish to discuss.

If some questions arise from these Zoom sessions that might be best addressed in a lecture format (that is, by a “Bonus” mini-lecture, or whatever) then I will record one, and make it available to everyone, as soon as possible after that Zoom session.

If you want to have a more private conversation with me – for example, perhaps about your own work – then of course we can arrange to do so (probably via Zoom).

The first Zoom sessions are on Tuesday, June 23. The last Zoom sessions are on Friday, July 17. Here is the Monday-Friday Zoom session schedule:

- Tim McDaniel (Instructor): 9:00 - 11:00 a.m.*
- Drew Winters (Teaching Assistant): 12:00 - 2:00 p.m.
- Nathan Tarr (Teaching Assistant): 3:00 - 5:00 p.m.
- Kim Turner (Teaching Assistant): 6:00 - 8:00 p.m.**

* Friday, June 26: Tim McDaniel will not have a Zoom session.
** Thursday, June 25, and Thursday, July 2: Kim Turner’s Zoom session will be 5:00 - 6:00 p.m.

All times are Ann Arbor (USA Eastern Daylight time zone [i.e., EDT; GMT/UTC –4h]) times.

Learning, Teaching Assistants, and the Lecture Note Packets

The primary purpose of the asynchronous lectures, Lecture Note Packets, and the Optional Learning Exercises is to help you learn this material. You are not in this alone. Studying and learning with other participants is probably a very good idea for many of you, as is taking advantage of Zoom session opportunities involving myself and the Teaching Assistants.

We have participants from all over the USA and all over the world; that is a lot of time zones! The Teaching Assistants and I make every effort to be accessible to you (e.g., eight hours per day, spread out across an eleven-hour window, Monday through Friday!). I encourage you to attend our Zoom sessions or make a Zoom appointment if those hours do not fit your schedule.

Note: The “many time zones” situation is also why the Lectures are provided asynchronously.

You need to read, search, and study the Lecture Note Packets as you study the Optional Learning Exercises (if you choose to do so; they really are optional...) and before you ask any questions. After that, if you still have a question then that is fine; ask away! Now the resulting discussion will be much better informed, more efficient, and more productive. Also: The outline in this syllabus, and the Table of Contents (which includes page numbers) in the front of each of the twenty Lecture Note Packets, might be useful to you...so use them.
Grades, Grades, Grades...

Given the new course content and delivery mechanism being used this summer, I am going to try something a bit different when it comes to grades. I think it the best way to do it this way, under these circumstances (which is why I am doing it!). Remember, you can also use the seven Optional Learning Exercises (and the Tutorial and Answer Keys) to help you learn the material!

One Exam will be given near the end of the course. This Exam will determine your bottom-line course letter grade. You will have plenty of time (at least 24 hours) to complete this Exam, even though it should only take you two or three or so hours to actually do the entire thing.

All answers can and will be typed into a Word document, which you will then send to me when you are finished with the Exam. I will design the questions so that you will not draw any pictures, or prove/derive anything using mathematics, on this Exam. So do not worry about typing fancy stuff like subscripts, Σ’s, Greek letters, plots, graphs, tables, etc.; instead, simply make sure I know what you are trying to say (e.g., “X1” is okay; “X₁” is not necessary).

You can use your personal notes, all handouts (including the Optional Learning Exercises), and all twenty of the Packets on the Exam. But I want your answers to be your answers! So use these artifacts only as a resource. Answer each question in your own words, not in mine. Feel free to use, but do NOT simply directly copy from, the handouts or the Packets.

You will answer a subset of questions and problems on the Exam. You get to choose which ones you (and, perhaps more importantly, which ones you do not!) answer. Each of your answers typically will be two or three paragraphs or so in length. Remember that none of the questions will be purely mathematical in nature, nor will you draw pictures as part of your answer.

The basis for answering each of the questions/problems can be found in the twenty Packets and also will have been discussed in the video Lectures.

Of course, be complete... but there is no need to write a novel. I am not looking for 100%-expert answers here; instead, I want to see that you understand this material well enough to do a good job of interpreting and explaining it. That is what the Exam is all about.

I will be the only person who grades your exam (i.e., not the Teaching Assistants). As usual in this class, course letter grades will also be determined by me, based upon the University of Michigan Graduate grading norms and conventions and, of course, my judgment.

Additional “practical matter” details about this Exam will be provided later in the course.

Concluding Remarks

Learning the material in this course will require a substantial amount of effort...but that is why you are in the Summer Program! The return-on-investment payoff will be worth it. Let me know if the Teaching Assistants or I can be of any assistance to you in this endeavor.

It is an honor to be your instructor for this course.
Course (and Lecture Note Packets) Outline

Packet I. Basic Statistics Review

1 A. Summations and Sigma Notation
B. Basic Statistics
   1. Mean
   2. Variance and Standard Deviation
   3. Probability
   4. Random Variables
      a. Continuous versus Discrete
      b. Nominal, Ordinal, and Interval
   5. Standardized Variables
   6. Expected Value
   7. Covariance, Correlation, and Causality
   8. Independence
   9. Normal Distribution: Notation, Skewness & Kurtosis, Areas
   10. Central Limit Theorem
   11. Student’s t Distribution
   12. Hypothesis Testing
   13. P-Values (“Prob-Values”): Interpret, Explain, and Use
   14. To “p” or Not to “p”? That is the Question...
   15. Confidence Intervals

Packet II. Supplement to Basic Statistics Review

2 A. A Closer Look at Population and Sample Variances
B. Hypothesis Testing: Summary, Flowchart, Protocol, and p-Values
C. Some Abuses and Misuses of Probability and Statistics
D. Symbol Glossary

Packet III. Bivariate Regression

3 A. Notation
B. Fitting a Line
C. Ordinary Least Squares Assumptions
D. Deriving the Sample Intercept and the Slope Coefficient Using Algebra
E. Centered Variables
F. The Estimated Slope Coefficient (“b”)
   1. Variance and Confidence Interval
   2. Confidence Interval and Hypothesis Testing
G. The Gauss-Markov Theorem
H. Appendix: Deriving the Formulas for “a” and “b” Using Calculus

Optional Learning

Exercise #1

Distributed
Packet 4
I. Residuals
   1. Definition and Estimation
   2. Population “Error” as a Disturbance, or Stochastic Element
J. Explained, Unexplained, and Total Deviations and Sums of Squares
K. Goodness of Fit
   1. Coefficient of Determination (R-Squared)
      a. Correlations (Again...)
      b. Why the R-Squared Can Be Inappropriate and Misleading
      c. Perils of Maximizing R-Squared: A Monte Carlo Simulation
   2. Standard Error of Regression (SER)
L. Standardized Variables and Beta Weights
M. Reporting OLS Regression Results
N. Regression Forced Through the Origin
   1. Definition, Illustrations, and Examples
   2. The Importance of Theoretical and Substantive Justifications
O. Comparison of Centering, Standardizing, and Forcing Through the Origin
P. Another Note on the Meaning and Interpretation of “a,” “b,” and Ŷ
Q. An Analogy: Means, Slopes, Standardization, Samples and Populations

Packet 5
R. Functional Transformations of Independent Variables
   1. The Need for Transformations
   2. The Regression is Still Linear
   3. The Natural Log and the Square Root Transformations
   4. The X-Squared Transformation (a First Look)
   5. Presenting Findings with Transformed Independent Variables

Packet 6
S. Interpolation, Predictive Intervals, and Extrapolation
T. Some Simple Diagnostic Plots
   1. Y vs. X (and an Introduction to Outliers)
   2. Y vs. Ŷ
   3. Residual vs. X
   4. Residual vs. Case Number
   5. Residual vs. Lagged Residual
   6. Residual vs. Ŷ
Optional Learning
   3. Residual vs. X
   4. Residual vs. Case Number
Exercises
   5. Residual vs. Lagged Residual
   6. Residual vs. Ŷ
   U. Simpson’s Paradox, Aggregation Bias, and the Ecological Fallacy
   #2 and
   #3
   V. Generation and Interpretation of Computer Output Using Real Data
   #2 and
   Distributed
   1. Setting Up the Substantive Example
   2. Generating and Analyzing the Output
   W. The Usefulness of Simple Scatter Plots: An Illustration
   X. R-Squared, “b,” and SER: A Monte Carlo Simulation
<table>
<thead>
<tr>
<th>Packet</th>
<th>IV. Multiple Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>A. Limits of Bivariate Regression</td>
</tr>
<tr>
<td></td>
<td>B. Trivariate Regression</td>
</tr>
<tr>
<td></td>
<td>1. Visualization</td>
</tr>
<tr>
<td></td>
<td>2. The Residual Term: Definition, the SER, and Degrees of Freedom</td>
</tr>
<tr>
<td></td>
<td>3. The (Two) Estimated Slope Coefficients</td>
</tr>
<tr>
<td></td>
<td>a. Partial Effects Equations, Models, and Venn Diagrams</td>
</tr>
<tr>
<td></td>
<td>b. Computing the Two Slope Coefficients</td>
</tr>
<tr>
<td></td>
<td>c. Variance and Confidence Intervals</td>
</tr>
<tr>
<td></td>
<td>d. Meaning and Interpretation</td>
</tr>
<tr>
<td></td>
<td>4. Holding One Variable Constant: What’s That All About?</td>
</tr>
<tr>
<td></td>
<td>5. The Impact on OLS Assumptions</td>
</tr>
<tr>
<td></td>
<td>C. Multiple Regression: The General OLS Model</td>
</tr>
<tr>
<td></td>
<td>1. The Slope Coefficient</td>
</tr>
<tr>
<td></td>
<td>a. Partial Effects Equations</td>
</tr>
<tr>
<td></td>
<td>b. Computing the Slope Coefficients</td>
</tr>
<tr>
<td></td>
<td>c. OLS Assumptions and the Gauss-Markov Theorem</td>
</tr>
<tr>
<td></td>
<td>d. Meaning and Interpretation</td>
</tr>
<tr>
<td></td>
<td>e. Variance and Confidence Intervals</td>
</tr>
<tr>
<td></td>
<td>2. The Residual Term</td>
</tr>
<tr>
<td></td>
<td>a. Definition</td>
</tr>
<tr>
<td></td>
<td>b. The SER and Degrees of Freedom</td>
</tr>
<tr>
<td></td>
<td>3. Degrees of Freedom: From One to Many Variables</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Packet</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4. Summary and Review of Partial Effects and the Interpretation of “b”</td>
</tr>
<tr>
<td></td>
<td>5. Summary and Review of t Stats, p-Values, and Hypothesis Tests</td>
</tr>
<tr>
<td></td>
<td>6. Units of the SER and Comparing SER’s Across Equations</td>
</tr>
<tr>
<td></td>
<td>7. R-Squared and Adjusted R-Squared</td>
</tr>
<tr>
<td></td>
<td>a. Review of the R-Squared Statistic and Inappropriate Uses</td>
</tr>
<tr>
<td></td>
<td>b. R-Squared and Functional Transformations</td>
</tr>
<tr>
<td></td>
<td>c. Multiple Independent Variables and the Effect on R-Squared</td>
</tr>
<tr>
<td></td>
<td>d. Understanding and Interpreting the Adjusted R-Squared Value</td>
</tr>
<tr>
<td></td>
<td>8. Multiplicative Interactions</td>
</tr>
<tr>
<td></td>
<td>a. Introduction, Description, Analogy to Functional Transformations</td>
</tr>
<tr>
<td></td>
<td>b. Models with an Interaction of</td>
</tr>
<tr>
<td></td>
<td>i. Two Dummy Variables</td>
</tr>
<tr>
<td></td>
<td>ii. One Dummy Variable and One Continuous Variable</td>
</tr>
<tr>
<td></td>
<td>iii. Two Continuous Variables</td>
</tr>
<tr>
<td></td>
<td>c. Interpreting and Understanding: Algebra, Components, and Theory</td>
</tr>
<tr>
<td></td>
<td>d. The Components of an Interaction Model: A Comparative Diagram</td>
</tr>
<tr>
<td></td>
<td>e. Conditional Impact: “Can I Exclude a Stand-Alone Term?”</td>
</tr>
<tr>
<td></td>
<td>f. Presenting Results, Graphs, and Significance Tests</td>
</tr>
<tr>
<td></td>
<td>g. Additional Readings and Concluding Remarks</td>
</tr>
<tr>
<td></td>
<td>h. Bivariate Regression and Interaction Term Models: A Comparison</td>
</tr>
</tbody>
</table>
9. Multicollinearity and Multicollinearity Diagnostics
   a. Perfect Multicollinearity: An Example
   b. Auxiliary R-Squared, Tolerance, and Variance Inflation Factor
   c. Explaining Multicollinearity Using Venn Diagrams
   d. Consequences, Including Possible “Backdoor Bias”
   e. When to Suspect Multicollinearity Problems
   f. Do Not Use Bivariate Correlations to Diagnose Multicollinearity
   g. Possible Remedies
   h. A Statistics, Estimation, and Information (Not a Theory!) Problem
   i. Multicollinearity and Models with Interaction Terms
   j. Narrowing Down the Source of the Multicollinearity

10. Dummy and Categorical Independent Variables
    a. Definition and Interpretation
    b. The Importance of “Intervalness” for Independent Variables
    c. Replacing a Categorical Independent Variable with Dummies
        i. Implementation and Interpretation
        ii. Excluding One Dummy Variable from the Model
    d. Graphing Models with Dummy Variables (and Interactions)
    e. Comparison: “Regression Forced Through Origin” to “Having X in an Interaction, but Not as a Stand-Alone, Term”
    f. Interpretation of Category Dummies and “Jumps”

11. Functional Transformations
    a. Things to Consider Regarding Transforming Y
    b. Log Transformations and Constant Elasticity Models
    c. Conditional Impacts and Slopes in Models with Interaction Terms
    d. More on Models with an X-Squared (e.g., Threshold Models)
    e. The Bend Rule

12. Model Specification
    a. Review of Types of Specification Error
    b. Omitting Relevant Variables: Derivation and Consequences
    c. Including Irrelevant Variables: Derivation and Consequences
    d. Variable Selection
    e. Perils of Stepwise Regression
    f. An Alternative to Standardization for Interval-Level Discrete X’s

13. Missing Data
    a. Data Missing at Random: Dependent and Independent Variables
    b. “Solutions” and Their Potential Problems
        i. Casewise (Listwise) Deletion
        ii. Pairwise Deletion
        iii. Mean and Conditional Mean Substitution
        iv. Other Methods (e.g., Profile Models)

14. Measurement Errors: In Y and in an X

15. Partial Effects Plots and Linearity in Multiple Regression Models
   a. Type I and Type II Errors  
   b. Flowcharts: Review of the Logic of Hypothesis Testing  
   c. Essay: Hypothesis Testing and American Criminal Trials  
   Optional Learning  17. Review: Category Dummy Variables  
   18. Review: Models with Continuous, Dummy, and Interaction Terms  
   Exercises #4  19. Review: Diagnostic Partial Plots  
   #5  20. Omitted Variable Plots  
   and  21. Another Look at Outliers: What They Are and Why They Matter  
   23. Multiple Regression: A Computer Example Using ANES Data  
      a. Setting Up the Example (American National Election Study)  
      b. Generation and Interpretation of Computer Output  
         i. Control Variables, Multicollinearity, and Diagnostic Plots  
         ii. An Illustration of Partial Effects  
         iii. Categorical X’s: Intervalness, Jumps, and Collapsing  
         iv. Creating an Interaction Term  
   24. Subgroup Differences: “Dummy*Continuous” Interaction Terms  

Packet 12  V. Analysis of Variance and the F Test  
   A. The F Distribution  
   B. Total, Regression, and Error Sum of Squares...and the ANOVA Table  
   C. The F Statistic and the F Test  
      1. A Test Involving All of the Regression Coefficients  
      2. When the F Test and Student’s t Test Are Identical  
   D. Analysis and Demonstrations Using a Regression Simulation  
      1. Interpreting the F Test: A “Randomly Generated Values” Illustration  
      2. A Closer Look at R-Squared, Adjusted R-Squared, and SER  

Packet 13  VI. Categorical Independent Variables (and F Tests): A Closer Look  
   A. Problems using Categorical and Non-Interval Independent Variables  
   B. Review: Replacing a Categorical X with Dummy Variables in OLS  
   C. Various Tests Involving the F Distribution  
      Optional Learning  1. A Review of the Whole-Model F Test in a Regression Environment  
      2. Nested F Tests: Testing Groups (Subsets) of Regression Coefficients  
         a. Purpose, Hypotheses, and Procedures  
         b. Group Significance of Multicollinear Variables (e.g., Interactions)  
      Distributed  c. Equivalent to Testing the SER’s  
      3. The Chow Test  
         a. Purpose, Hypotheses, and Procedures  
         b. Applications: Structural Shifts and Whole-Model Performance  
         c. Generalized (Multiple Subgroup) Chow Test and Aggregation Bias  
      4. Overview of Comparing “Safe” versus “Risky” Models  

(Packet 13 is continued on the next page)
(Packet 13, continued from the previous page)

5. Comparison of SER’s with No Prior Knowledge of Relative Model Performance (i.e., When There is No “Safe” or “Risky” Model)
7. A General Discussion of F Tests, Using ANOVA Terminology

D. Collapsing Categories of a Categorical Variable
   1. Setup and Notation
   2. Statistical Hypothesis Test for Collapsing Categories

E. Treating a Categorical Variable as Interval-Level
   1. Setup and Notation, and “Contextually Interval”
   2. Statistical Hypothesis Test for Intervalness

F. Monte Carlo Simulation Results: Which Dummy to Exclude

G. Generalizations of the F Test
   1. For Collapsing Categories and for Intervalness
   2. For “Safe” and “Tested” Models in General

H. Optional: Discussion and Analysis of a Substantive Example (Venezuela)
   1. Description of the Study
   2. Statistical Tests Performed
   3. Potential Effects of Multicollinearity
   4. Different Specifications, Interpretations, and Conclusions
   5. Interpretation of Manipulated Coefficients

Packet VII. Dichotomous Dependent Variables and the Logit Model

A. Examples

B. The Binary Choice Model
   1. Description and Illustration
   2. Problems Using OLS with a Dichotomous Dependent Variable
      a. Over- or Under-Estimating Y
      b. Var(e), Heteroskedasticity, and the Goldberger Procedure
      c. Probable Non-Linearity
   3. Using an S-Shaped Curve Instead of a Line
      a. Description and Theory
      b. Fit and Bias

C. The Logit Model
   1. The Logistic Function and Log-Odds
   2. Interpreting Logit Coefficients
   3. Interpreting and Presenting the Results of Logit Analysis: An Example
   4. How Logit Works: Big Picture and Comparison to Linear OLS
   5. Estimating Coefficients: Maximum Likelihood Estimation
   6. Significance of Coefficients: The Likelihood Ratio Test and the Z Test
   7. Using and Interpreting Logit Output: An Education Example
### Packet VIII. Matrices and Matrix Algebra

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>A. General Terms and Definitions</td>
<td>For you C. Inverse Matrices</td>
</tr>
<tr>
<td></td>
<td>B. Addition, Subtraction, and Multiplication: How to Do It and Examples</td>
<td>to review D. Determinants and Condition Numbers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>largely E. Rank and Singularity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>on your F. Properties of Matrix Algebra</td>
</tr>
<tr>
<td>(also with own)</td>
<td>G. Kronecker (Tensor) Product</td>
<td>a short H. Eigenvalues and Eigenvectors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lecture) I. Positive Definite Matrices</td>
</tr>
</tbody>
</table>

### Packet IX. Outliers and Graphical Techniques (“Weird Points and Pretty Pictures”)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>A. Outliers: Detection, Classification, and Possible Effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Influence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Definition and Description</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Detection Using Statistics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>i. DFBETA and DFBETAS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ii. Cook’s Distance (D) and DFFITS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Leverage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Definition and Description</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Hat Values</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Studentized Residuals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Definition, Calculation, and the Bonferroni Adjustment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. A Note on Notation and Terminology (It’s a Mess...)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. Case Dummies and Joint Cases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B. Diagnostic Plots and Graphical Techniques</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Proportional Leverage Plots: Definition and Different Types</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Jointly Influential Observations: Definition and Examples</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Conditional Effects Plots</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a. Definition, Description, Presentation, and Interpretation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Models with Interaction Terms (Including Confidence Bounds)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. The Importance of “Conditional”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Jittered Data Plots (for Categorical Independent Variables)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Scatterplot Matrix</td>
<td></td>
</tr>
</tbody>
</table>

(Packet 17 is continued on the next page)
(Packet 17, continued from the previous page)

I. Ordinal and Dummy Variables: An Example Using Matrices
J. Example: A Singular X’X Matrix
K. Transformations and Interactions in Matrix Form: OLS is Still Linear
L. Appendix: A Proof that the Matrix Derivation of “B” Minimizes the SSE

Packet XI. Heteroskedasticity (and Generalized/Weighted Least Squares)

A. Definition and Omega Matrices
B. Variance-Covariance Matrices
C. Detection of Heteroskedasticity Using Diagnostic Plots
D. Types of Heteroskedasticity
E. When to Expect Heteroskedasticity
F. Consequences of Using OLS with Heteroskedastic Errors
G. Detection of Heteroskedasticity Using Statistical Tests
   1. Goldfeld-Quandt Test
   2. Glejser Test
   3. Likelihood Ratio Test
H. Correcting for Heteroskedasticity Using Generalized Least Squares
   1. Introduction, and Deriving the Slope Coefficients using GLS
   2. Benefits of Using GLS When You Have Heteroskedastic Errors
   3. GLS as Weighted Least Squares (WLS)
   4. The Underlying Logic of WLS: An Example Using State-Level Data
   5. Using GLS Given a Model for Var(ε): The General Case

Packet XII. Autocorrelation

A. Definition of First-Order Autocorrelation
B. When to Expect Autocorrelation
C. Algebraic Interpretation of First-Order Autocorrelation
D. Examples of Positive and Negative First-Order Autocorrelation
E. Four Common Types of Time Series Models
F. Consequences of Using OLS with Autocorrelation
G. Detection of Autocorrelation
   1. Diagnostic Plots
   2. Using a Bivariate Model
   3. The Durbin-Watson Statistic (First-Order Autocorrelation)
      a. Algebraic Derivation and Implementation
      b. Some Things to Keep in Mind
   4. The Wallis Test (Fourth-Order Autocorrelation)
H. Correcting for First-Order Autocorrelation
   1. The Prais-Winsten GLS Estimator: Algebraic Procedure and Proofs
   2. Some Other GLS Methods
Packet XII. More on Dichotomous Dependent Variables: The Probit Model

20 A. The Normal Distribution Revisited
   1. The Standard Normal Distribution
   2. The Normal (Gaussian) CDF
B. The Probit Function
C. Finding Probabilities Using the Normal Distribution
D. Interpreting and Presenting Probit Coefficients
E. Maximum Likelihood Estimation of Probit Models
F. Deciding Between Using Logit versus Probit Models

XIV. Chart: Some Potential Problems When Using OLS Regression Models

XV. Final Remarks
A. More than One Equation: Simultaneous Equations Models
B. Additional Topics for Further Study
C. The Importance of Parsimony and Presentation
D. The Dangers of Over-Reliance on Statistical Procedures

---

Some (Optional) Suggested Readings

Again: The twenty Lecture Note Packets that I wrote and we use serve as the de facto required textbook for this course. I advise you to use the outline in this syllabus (along with the Table of Contents at the start of each Packet) as a guide for these Lecture Note Packets.

The additional readings in this section of the syllabus (details of which can be found in the Bibliography section that immediately follows) are totally optional for you in this course.

Several of these readings are from the following four traditional textbooks, each of which I like a lot. While there are many similarities between them, I think that each takes a somewhat different (though at times only subtly different) pedagogical approach.

- Gujarati takes more of a “Math Language” (but not too advanced or difficult) approach.
- Hamilton takes more of a “Picture Language” (lots of graphics, etc.) approach.
- Kennedy takes more of an “English Language” (narrative explanations) approach.
- Wooldridge takes more of a “Combined” (and accessible and comprehensive) approach.

Also, Fox’s (2017) textbook provides broad coverage, thorough explanations, and creative uses of graphics that are among the best I have ever seen. This book might [?] be a bit too advanced for some participants in this course, but it could be a wonderful post-course reference. So I have included it in the bibliography, but not in the list of suggested optional readings, in this syllabus.

You will notice several “little green books” from the QASS Sage Series and also numerous articles from a number of journals across several academic disciplines. Both the more recent ones and the older classics are very useful as learning vehicles...which is why I included them!

The suggested optional readings for each of the Lecture Note Packets are on the next four pages.
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
</table>
| Packets 1 & 2  | Baumgartner, Breunig, et al. (especially pages 606-613)  
Gujarati and Porter: Intro; Sections 5.5-5.8, 6A.2; Appendix A  
Hamilton: Pages 289-296  
Kennedy: Sections 2.5-2.8  
Mohr  
Wasserstein, Schirm, and Lazar  
Wooldridge: Appendices A.1-A.4, B, C.2-C.6 |
| Packet 3       | Gujarati and Porter: Chapter 1; Sections 2.1, 2.2, 2.6, 2.7, 3.1-3.4, 3A.1-3A.7, 4.1-4.3, 5.1-5.3  
Hamilton: Pages 29-34, 42-49, 296-297  
Kennedy: Chapter 3; Sections 1.1, 1.4, 2.1-2.3, 2.11  
Lewis-Beck and Lewis-Beck: Pages 1-14, 23-38  
Schroeder, Sjoquist, and Stephan: Pages 11-23, 81-82  
Wooldridge: Chapter 1; Sections 2.1-2.5 |
| Packet 4       | Achen (1982): Pages 73-77  
Achen (1991)  
Gujarati and Porter: Sects. 2.4, 2.5, 3.5, 3.8, 5.4, 5.11, 6.1-6.3, 6A.1  
Hamilton: Pages 37-41, 49-51, 124-125  
Kennedy: Sections 1.2, 2.4, 2.10; Pages 109-110  
King (1986)  
King (1991)  
Lewis-Beck and Skalaban  
Schroeder, Sjoquist, and Stephan: Pages 23-29, 31-32  
Wooldridge: Section 2.6 |
| Packet 5       | Gujarati and Porter: Sections 2.3, 6A.3; Pages 164-166  
Hamilton: Pages 53-58, 148 |
Gujarati and Porter: Sections 3.6, 3.7, 5.10  
Lewis-Beck and Lewis-Beck: Pages 19-20, 39-49, 86-88 |
| Packet 7       | Achen (1982): Pages 7-51  
Asher: Pages 237-248  
Berry: Pages 1-24, 81-83  
Berry and Feldman: Pages 9-15  
Fox (1991): Pages 3-9  
Gujarati and Porter: Sections 7.1-7.4, 7.6, 8.1-8.3, 8.8  
Hamilton: Pages 65-72, 109-113  
Kennedy: Section 1.3  

(Packet 7 is continued on the next page)
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet 7</td>
<td>Lewis-Beck and Lewis-Beck: Pages 55-64</td>
</tr>
<tr>
<td><em>(Continued from previous page)</em></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 29-31</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Chapters 3, 5; Sections 4.1-4.3</td>
</tr>
<tr>
<td>Packet 8</td>
<td>Berry: Pages 24-27</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 15-16</td>
</tr>
<tr>
<td></td>
<td>Brambor, Clark, and Golder</td>
</tr>
<tr>
<td></td>
<td>Braunmoeller</td>
</tr>
<tr>
<td></td>
<td>Davenport, Armstrong, and Lichbach</td>
</tr>
<tr>
<td></td>
<td>Friedrich</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 7.5, 7.8, 9.6</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 77-80, 84-85</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Section 5.5</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 69-71</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 32-51, 54-56, 58-59</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Section 6.3</td>
</tr>
<tr>
<td>Packet 9</td>
<td>Asher: Pages 248-250</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 45-47</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 37-50, 64-70</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 10-21</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 10; Sections 8.5, 9.1-9.4</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 82, 85-92, 133-136</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 1-21, 29-48</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Chapter 12; Sections 15.1-15.4</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 64-69, 75-83</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 56-58, 71-72</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 7.1-7.3, 7.6; Pages 240-245</td>
</tr>
<tr>
<td>Packet 10</td>
<td>Achen (1982): Pages 51-73; 77-79</td>
</tr>
<tr>
<td></td>
<td>Asher: Pages 250-255</td>
</tr>
<tr>
<td></td>
<td>Baumgartner, Breunig, et al. (especially pages 613-615)</td>
</tr>
<tr>
<td></td>
<td>Berry: Pages 27-41, 60-66</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 18-37, 71-72</td>
</tr>
<tr>
<td></td>
<td>Clarke</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 53-61</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 6.4, 6.5, 6.8, 6.9, 7.7, 7.9, 7.10, 13.1-13.3, 13.5, 13A.1-13A.3; Pages 499-500</td>
</tr>
<tr>
<td></td>
<td>Halvorsen and Palmquist</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 72-77, 82-84, 148-153, 163-167, 173-174</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 5.1-5.4, 6.1-6.3; Pages 49 (the first point), 111</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 49-52, 72-73</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 53, 59-61, 65-71</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 6.1, 6.2, 9.4; Pages 303-306, 324</td>
</tr>
<tr>
<td>Lecture Packet</td>
<td>Suggested Optional Readings (In Alphabetical Order)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------------------------</td>
</tr>
</tbody>
</table>
Achen (2005)  
Berry: Pages 41-45, 49-60  
Eisenberg and Wells  
Hamilton: Pages 158-163  
Lewis-Beck and Lewis-Beck: Pages 88-93  
Schroeder, Sjoquist, and Stephan: Pages 63-65  
Wooldridge: Sections 4.6, 6.4 |
| Packet 12      | Berry and Feldman: Pages 17-18  
Gujarati and Porter: Sections 5.9, 8.4  
Hamilton: Pages 92-101  
Mock and Weisberg  
Schroeder, Sjoquist, and Stephan: Pages 51-52  
Wooldridge: Pages 152-153 |
Gujarati and Porter: Sections 8.7, 9.5, 13.7  
Hamilton: Pages 80-82, 101  
Hardy: Pages 21-29, 48-53, 78-82, 84-85  
Kennedy: Sections 4.3, 4.4  
McDaniel  
Wooldridge: Section 4.4; Pages 143-152, 245-248 |
| Packet 14      | Ai and Norton  
Berry: Pages 47-49  
Gujarati and Porter: Sections 15.1-15.8, 15A.1  
Hamilton: Chapter 7  
Hardy: Pages 75-78  
Kennedy: Sections 2.9, 4.5, 16.1  
Markus: Pages 555-556  
Menard: Chapters 1-4; Appendix  
Pampel: Chapters 1-3; Appendix  
Schroeder, Sjoquist, and Stephan: Pages 79-80  
Wooldridge: Sections 7.5, 17.1 |
| Packet 15      | Gujarati and Porter: Appendices B.1-B.3  
Hamilton: Pages 333-337, 344-345  
Wooldridge: Appendices D.1-D.5 |
| Packet 16      | Bollen and Jackman  
Chatterjee and Wiseman  
Fox (1991): Pages 21-40  
(Packet 16 is continued on the next page) |
<table>
<thead>
<tr>
<th>Lecture Packet</th>
<th>Suggested Optional Readings (In Alphabetical Order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet 16</td>
<td>Gujarati and Porter: Pages 496-497</td>
</tr>
<tr>
<td>(Continued from previous page)</td>
<td>Hamilton: Ch. 1; Pages 113-117, 125-133, 141 (note #4), 154-158</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 21.1, 21.2; Page 347</td>
</tr>
<tr>
<td></td>
<td>Narang and Nelson (especially pages 364-370)</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Pages 326-330</td>
</tr>
<tr>
<td>Packet 17</td>
<td>Achen (1982): Appendix</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 80-82, 83-85</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Appendices C.1-C.10, C.12, CA.4</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 338-344</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 4.1, 4.2</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Appendices E.1-E.3</td>
</tr>
<tr>
<td>Packet 18</td>
<td>Berry: Pages 67, 72-78</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 73-88</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 11; Appendices C.11, CA.3</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 183-189</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 53-61</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 8.1-8.3</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 75-77</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 8.1-8.4</td>
</tr>
<tr>
<td>Packet 19</td>
<td>Berry: Pages 67-72, 78-81</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Sections 12.1-12.9, 12A.1, 12A.2</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 118-124</td>
</tr>
<tr>
<td></td>
<td>Hardy: Pages 82-84</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Section 8.4</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 72-75</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Sections 10.1-10.3, 11.2, 11.5, 12.1-12.3, 12.6</td>
</tr>
<tr>
<td>Packet 20</td>
<td>Asher: Pages 255-256, 260-261</td>
</tr>
<tr>
<td></td>
<td>Berry and Feldman: Pages 88-89</td>
</tr>
<tr>
<td></td>
<td>Fox (1991): Pages 75-80</td>
</tr>
<tr>
<td></td>
<td>Goertzel (2002)</td>
</tr>
<tr>
<td></td>
<td>Granato</td>
</tr>
<tr>
<td></td>
<td>Gujarati and Porter: Chapter 18; Sections 13.13, 15.9, 15.10</td>
</tr>
<tr>
<td></td>
<td>Hamilton: Pages 246 (note #2), 249-252, 281-282</td>
</tr>
<tr>
<td></td>
<td>Kennedy: Sections 22.1-22.4</td>
</tr>
<tr>
<td></td>
<td>Lewis-Beck and Lewis-Beck: Pages 93-95</td>
</tr>
<tr>
<td></td>
<td>Menard: Chapter 5</td>
</tr>
<tr>
<td></td>
<td>Pampel: Chapters 4, 5</td>
</tr>
<tr>
<td></td>
<td>Schroeder, Sjoquist, and Stephan: Pages 77-79, 80</td>
</tr>
<tr>
<td></td>
<td>Wooldridge: Chapter 19; Sections 16.1-16.4, 17.1</td>
</tr>
</tbody>
</table>
Bibliography


Hamilton, L. *Regression with Graphics: A Second Course in Applied Statistics.* Duxbury, 1992. (Note: This is the same as the edition published by Brooks Cole in 1999.)


