PUB 316 Methodological Tools for Public Policy Spring 2008

Tuesday and Thursday 1:15pm -2:35pm in Social Science 134

Course Description:
The course will give students familiarity with basic statistical techniques as applied in the social sciences. It will consider the role of simple statistics to describe events, groups, and other things of interest to social scientists. It will also introduce students to Ordinary Least Squares (OLS) versions of multiple regression analysis. The class will be practical in its discussion of various statistical techniques, including weekly assignments to develop experience with each method.

This class meets the General Education Requirement for Mathematics and Statistics

Learning Objectives for Mathematics and Statistics
Courses in Mathematics and Statistics enable students to demonstrate:

1) knowledge of concepts, terms, and symbols used to analyze data
2) an ability to formulate problems in abstract form amenable to mathematical, statistical, or logical analysis
3) an ability to perform appropriate operations to draw conclusions from data
4) an ability to interpret and communicate quantitative information

Contacting Me:
Feel free to e-mail me at: s.barclay@albany.edu
Office Hours: Tuesday and Thursday 12.30pm – 1.15pm in Humanities 016

Requirements:

9 Assignments, 2 in-class exams, and a final exam

Assignments 1-6 ……...30% Due as designated in class
Assignments 7-8……....20% Due by the end of the semester
Assignment 9 …….……..….30% Due on day designated for class in exam schedule
2 in-class exams …….. .10% On day designated in class
Final Exam…….10% On day designated for class in exam schedule

a. If you fail to complete 2 or more of the assignments, you are automatically failed for the entire course.
b. Questions on the assignments that are not addressed will be taken as failing that section.
c. The 2 in-class exams will make you interpret statistical output – the purpose of each exam is to confirm that your understanding as displayed in the first 6 assignments is an accurate reflection of your individual ability. If your exam answers are not consistent with your apparent understanding as displayed in your prior assignments, your first 6 assignments will be retroactively re-graded after each exam. If you fail to take an exam, all of the earlier assignments will be graded as a zero.
d. The final exam will make you interpret statistical output -- the purpose of this exam is to confirm that your understanding as displayed in the last 3 assignments is an accurate reflection of your individual ability. If your exam answers are not consistent with your apparent understanding as displayed in these assignments, your last 3 assignments will be retroactively re-graded after this exam.

ATTENDANCE IS NOT REQUIRED (but very strongly suggested): You are free to miss as many classes as you like for any reason (on the clear presumption that you understand the consequences of such action). If you miss class, you are *not* required to provide a note to me, but it is your responsibility to stay up to date with the material -- you should contact your classmates to kindly request them to lend you a copy of their notes. Be aware that in office hours, I will **never** re-teach material taught in an earlier class simply because you missed that class and now realize that you do not understand the material currently being covered in class. Based on prior experience, I warn you that it is virtually impossible to successfully complete all of the assignments without consistent engagement with the class discussion. Missing even one class can cause major problems in gaining an understanding as the material is cumulative and complex.

The due date of each assignment will be only announced in class. Similarly, the date of each of the in-class exams will be only announced in class.

Group Work:
I really encourage students to work together in completing the assignments and understanding the material. In fact, I strongly suggest students create study groups that work together to complete each of the assignments.

However, it is essential that each student in the group develops their own independent knowledge of the material covered in the assignments – simply copying the work of others in the study group is not permitted. Each student must understand the techniques or ideas that created the answers you place in your assignments. Periodically, I may ask individual students to demonstrate to me (in individual sessions on the computer) that the work in their submitted assignments accurately reflects their level of understanding and your work may be retroactively re-graded in such circumstances. Similarly, the 2 in-class exams and the final exam exist primarily to confirm that the work in your submitted assignments accurately reflects your individual level of understanding. Your work will be retroactively re-graded after these exams if it is not consistent with your apparent understanding as displayed in your prior assignments. In any case of evidence of cheating, you will be failed for the entire course and your case will be forwarded to Judicial Affairs with a request for additional sanction as allowed by the University rules.

Textbook: (Available at Mary Jane's Bookstore)
Acock, Alan C. 2006. *A Gentle Introduction to Stata*. Stata Press. (ISBN-10: 1-59718-009-2 ISBN-13: 978-1-59718-009-2) --- http://www.stata.com/bookstore/acock.html ($42)

OR

Hamilton, Lawrence C. 2006. *Statistics with Stata (Updated for Version 9)*. Brooks / Cole Press. (ISBN-10: 0-495-10972-X ISBN-13: 978-0-495-10972-3) -- http://www.stata.com/bookstore/sws.html ($60)

Software:
I will demonstrate in class using STATA and Microsoft Excel. For those who do not wish to purchase it, Stata 9 is available in all of the computer labs on the Uptown and Downtown campus.

STATA 10 Intercooled – One Year License
Purchased directly through STATA through Grad Plan associated with University at Albany: Cost: $95 (http://www.stata.com/order/new/edu/gradplans/gp-campus.html)

Additional Software Used in Class Assignments:
Microsoft Excel (any recent version with the statistical package enabled)
Adobe Acrobat Reader (Version 7 or later)

Software Proviso:
I understand that you will likely be new to using this software and I will offer all sorts of helpful hints on the use of Stata in class and in the assignments. However, I still expect you to use your brain first even if you think "statistics is hard" and "the software is unfamiliar." So, I suggest the first recourse is to look at the program's help files. Both Stata and Excel have extensive help files. Only after you have exhausted these options, you should feel free to email your question or, better still, to ask it in class.

The Path to Failing the Class:
In the last few years, PUB 316 has failed approximately 30% of the students in the class. Some years, over 40% of the class have failed. This does not occur because the class is taught badly – in fact, the Department of Public Administration and Policy purposely uses some of its best teachers (among those with a statistical leanings) to teach this class. The class is also not very difficult – in the past, approximately 20% of the class have received a grade of A or A-. Finally, there is no evidence that a single student in the class ever failed because they were actually intellectually incapable of completing the material assigned.

The high failure rate occurs for two primary reasons:

a. The math-rejectionist approach. This approach usually has this format: A student thinks "I am not good at math and statistics is math. Therefore I am destined to fail the class." The assertion becomes a self-fulfilling prophecy and the student fails.

b. The student does not bother to do the work in the class because
    i. They mistakenly do not believe statistics and data analysis *really* applies to considerations of public policy, or
    ii. They mistakenly do not believe that the PUB Department would fail nearly all of the students in a class. Instead, they mistakenly think the class instructor will be blamed if the whole class does poorly and the standards will be lowered, or
    iii. They are too lazy to bother doing the work in the class and mistakenly think they can pass the class and complete the PUB major/minor without doing the required work.

The Path to Passing the Class:
Despite the fear of students, the course is designed with only a very limited involvement of the math that forms the basis of the statistics we consider in the class. While we highlight math formulas, we primarily consider these formulas in the context of understanding the "pieces" of information that are being used to calculate them. All the simple and hard math work is done by either Excel or Stata from computing averages to running regression.

Notwithstanding the myth that it is not used in public policy, the ability to understand data is a key component of policy analysis and evaluation. It is a key to determining successful from non-successful policies. Moreover, use of data is a feature of the policy input of interest groups and government agencies.

The best way to pass the class is to:

a. Treat statistics (and research design) as a new language you are learning. Allow yourself to learn new ideas and new ways of thinking about the data. But remember that learning a language takes time and you sometimes do not see how the whole thing comes together in the early stages – to learn a language, you need to first learn verbs, nouns, grammatical rules, and tense to be able to eventually construct effective sentences. So, give yourself time for it all to make sense to you.
b. Simply and consistently *try* to follow the class discussion as we proceed (by both attending class regularly **and** engaging with the material covered in class). Ask questions in class and in office hours if you do not understand it.
c. Do the assignments in small pieces in ways that follow the class. As we learn each small, new idea in class, use the separate sections of the related assignment to learn how to apply that small aspect in relation to real data.
d. Realize that the assignments sometimes take a lot of time to complete. Do not simply expect to complete them quickly, especially in the early ones when you may be new to using the software. I strongly suggest that you do not leave them to the very last minute.
e. Set aside some time to experiment with how it might apply to data or ideas you are interested in.

***PUB 316* -- First Classes**

1. **Introduction to the Idea of Statistics**
    - 1.1. Simple indicators of group behavior or social action
        - 1.1.1. Capturing some group characteristic or characteristic of social action
            - 1.1.1.1. Choice of characteristic defined by social scientist – selection issue
            - 1.1.1.2. Indicator often constructed in comparison terms
        - 1.1.2. Attempt to reduce complex information for description purposes
            - 1.1.2.1. Examples:
                - 1.1.2.1.1. GPA,
                - 1.1.2.1.2. Grades,
                - 1.1.2.1.3. Pct of Women in Class,
                - 1.1.2.1.4. Ratio of Democrats in State to Republicans
        - 1.1.3. To offer greater clarity for the description of any chosen factor (How many? How much?)
            - 1.1.3.1. "There are a lot more Democrats in NY City than Republicans" vs "There are 5 members of the Democratic Party in NY City for every Republican in the city."
            - 1.1.3.2. "She is a good student in her political science classes but performs poorly in other subjects" vs "She has a GPA of 3.8 in political science but a 2.7 GPA if we only consider her other subjects."

2. **Setting up the terms of the discussion**
    - 2.1. To successfully use statistics you must understand a few simple terms
    - 2.2. The issue of over-determination, determined, underdetermined
        - 2.2.1. Speed of Light example
        - 2.2.2. N of 1
        - 2.2.3. Not enough information
    - 2.3. The familiar univariate statistics (Consider only one characteristic)
        - 2.3.1. Sum – Total
        - 2.3.2. Percent – Converted to a number between 0.00 and 1.00 (0 and 100)
        - 2.3.3. Mean -- Average
        - 2.3.4. Mode – most commonly occurring
        - 2.3.5. Median – middle number of ordered set found at position (n+1)/2
        - 2.3.6. Total number of considered groups / events / individuals etc
    - 2.4. Some Greek letters and Quick symbols
        - 2.4.1. Sum – Σ
        - 2.4.2. Percent -- %
        - 2.4.3. Mean -- μ
        - 2.4.4. Number of groups etc – n
        - 2.4.5. Infinity -- ∞
    - 2.5. Simple Math
        - 2.5.1. Sum => ΣX_i
        - 2.5.2. Mean => $(\Sigma X_i)/ n$
    - 2.6. The nature of numbers
        - 2.6.1. Integers – real, whole numbers (usually greater than 0)
            - 2.6.1.1. Example: 1, 3, 7, 8, etc
        - 2.6.2. Negative
        - 2.6.3. Positive
        - 2.6.4. Unreal
    - 2.7. The "Distribution" of Numbers
        - 2.7.1. Count – a count of events involving integers between 0 and ∞
            - 2.7.1.1. Example: Number of Supreme Court Appointments in a time period
            - 2.7.1.2. Example: The number of votes cast on a congressional committee
        - 2.7.2. Continuous – a factor that involves numbers between $-\infty$ and ∞
            - 2.7.2.1. Example: The Gross Domestic Product of a country
            - 2.7.2.2. Example: Housing Prices in a region

2.8. Nature of Data
    2.8.1.1. Ordered – there is an order but not necessary identical interval
        2.8.1.1.1. Example: 2, 4, 5, 8, 9, 12
    2.8.1.2. Interval – ordered with a consistent interval
        2.8.1.2.1. Example: 2, 4, 6, 8, 10
    2.8.1.3. Nominal – no order in data because of nature of variable considered
        2.8.1.3.1. Example: red bus, blue bus, green bus
        2.8.1.3.2. Example: Socialist, Democrat, Green, Independent, Liberal, Republican,
2.9. Format of Numbers
    2.9.1. Numbers – Any form of real or unreal number
        2.9.1.1. Example: 6 , $\prod$, $\sqrt{11}$
    2.9.2. Strings – Text or Words
2.10. Variables
    2.10.1. Any set of indicators used to represent an observed and documented characteristic of the object under consideration
        2.10.1.1. Must vary to be a "variable" (or potential to vary in some unusual cases), otherwise determined
        2.10.1.2. May be more than one possible variable to represent same construct
            2.10.1.2.1. Example: Percent Unemployed, Total Unemployed, Seasonal Jobless Rate (excluding Farm Workers) etc

3. **Univariate Statistics**
3.1. Measures of Central Tendency
    3.1.1. Mean $_{(\Sigma X_i)}$/ n
    3.1.2. Mode
    3.1.3. Median
3.2. Why are these three statistics important in describing the characteristics of a variable?
3.3. Thinking about the distribution of each of each variable.
    3.3.1. Stem and Leaf
    3.3.2. Histograms
3.4. How are they distributed?
3.5. How do data points differ from the mean?
    3.5.1. Is it skewed to one side from the mean
    3.5.2. Variance
    3.5.3. Standard Deviation
        3.5.3.1. See p. 60 of Agresti and Findlay for interesting aspect of it

4. **Probability Distributions**
4.1. Different characteristics of Probability Distributions
    4.1.1. Many probabilities – normal, chi-square, student t, logistic etc
    4.1.2. Many variations on even "familiar" ones
        4.1.2.1. based on their parameters – mean, standard deviation, etc
            4.1.2.1.1. Mean -- μ
            4.1.2.1.2. Standard deviation – σ
    4.1.3. Normal Distribution
        4.1.3.1. z-scores
        4.1.3.2. Calculating probabilities using distributions
        4.1.3.3. The "tails"
        4.1.3.4. Standard normal
    4.1.4. Standard Error of Y
    4.1.5. Central Limit Theorem
    4.1.6. Sample mean to "True" Mean

5. Estimators and Confidence Intervals

- 5.1. Point Estimators
  - 5.1.1. Terms of the Discussion
    - 5.1.1.1. Unbiased v. Biased Estimators
    - 5.1.1.2. Efficient
    - 5.1.1.3. Large and Small Standard Errors
- 5.2. Confidence Intervals
  - 5.2.1. Related to Standard Error of Mean

6. Statistical Inference and Hypothesis Testing

- 6.1. Hypothesis Testing
  - 6.1.1. Null Hypothesis H_0
  - 6.1.2. Alternative Hypothesis H_a
  - 6.1.3. Point Estimate
  - 6.1.4. P-value
  - 6.1.5. Conclusion
- 6.2. Example 6.2 from Agresti and Findlay (pp. 162-164)
- 6.3. One tailed tests and two tailed tests
- 6.4. Significance levels – α
  - 6.4.1. Myths about significance levels
  - 6.4.2. Choosing significance levels
  - 6.4.3. Type I and Type II errors (Agresti and Findlay p. 175)
  - 6.4.4. Impact of size of n
    - 6.4.4.1. larger n reduces standard error
- 6.5. T-Tests