Draft: Spring 2018: PUBP804-001
Multivariate Statistical Analysis in Public Policy

Instructor: Naoru Koizumi, Ph.D.
[Phone] 703-993-8380
[Email] nkoizumi@gmu.edu (for main communication)
[Office Hours] Mondays 5 – 7 pm (and by appointment)
[Office] Arlington Campus: Founders Hall 520

TA: TBD
Email: TBD (for main communication)
[Office Hours] XXX (and by appointment)
[Office] Arlington Campus: XXX

TABLE OF CONTENTS
· Course Description
· Course Time and Location
· Course Materials
· Course Software
· Course Topics
· Course Grading
· Lecture Schedule
· Reading List, Lecture Slides and Hand-Outs
· Homework Assignments
· Project Information
· Student with Special Needs
· Plagiarism

COURSE DESCRIPTION

PREREQUISITE: PUBP704 or equivalent
The course examines major techniques of multivariate statistical analysis in social sciences, with emphasis on the applications in policy studies. The course begins with a brief review of research design and fundamentals of statistical analysis, and then moves on to major multivariate techniques commonly used in policy research. The course covers logic and application of causal analysis using multivariate cross-tabulation and hypothesis testing, bivariate and multivariate regression, regression diagnostics, Generalized Linear Models (GLMs), factor and cluster analyses, path analysis and structural equations, and panel data analysis. The course stresses applications and interpretations over mathematical foundations, and in particular it will emphasize the applications of multivariate statistics to a real, complex database using a statistical package, STATA.

(return to contents)

COURSE TIME AND LOCATION

· Time: Mondays 7:20 – 10:00 pm
· Location: Founders Hall 206
LEARNING OUTCOME MEASURES

1. Knowledge and understanding
   - Students should be able to read and understand the journal articles that involves basic multivariate statistical analyses.
   - Students should be able to critically assess policy arguments, comments, reports and other materials that use statistical analysis.
   - Where relevant, students should be able to make policy remarks and draw policy implications/conclusions based on the findings of various policy studies that apply statistical techniques.

2. Develop quantitative and qualitative skills
   - Students should be able to conduct basic multivariate and statistical analyses to analyze various policy issues.
   - Students should be able to interpret the STATA outputs of various statistical analyses covered under this course.
   - Students should acquire basic programming skills necessary to conduct basic multivariate statistical analysis in STATA (and to be able to proceed to more in depth advanced statistics covered in PUBP705, if so wished).
   - Students should learn how to find, collect, organize and clean datasets necessary to analyze the topic of their interests.

3. Professional development and leadership
   - Students should learn working effectively in a team setting through homework assignment and course project.
   - Students should learn how to make professional presentations through a course project.
   - Students should learn how to seek assistance when needed and communicate effectively with classmates and instructors.
   - Students should learn how to deliver assignments and outputs in a timely fashion.

COURSE MATERIALS

REQUIRED TEXTS:
   - *The Multivariate Social Scientist* by Hutcheson and Sofroniou (Paperback, SAGE Publications)
   - *Statistics with Stata: Version 12* by Hamilton (Cengage)

RECOMMENDED TEXTS:
   - *Quantitative Research Methods in the Social Sciences* by P. Maxim (Oxford University Press)
   - *Applied Logistic Regression* by D. Hosmer and S. Lemeshow (Wiley InterScience Publication)
   - *Discrete Choice Analysis* by Ben-Akiva and Lerman (MIT Press)
   - *Basics of Structural Equation Modeling* by G. Maruyama (SAGE Publications)
   - *Factor Analysis as a Statistical Method* by D. N. Lawley and A.E. Maxwell (Elsevier Publishing)
   - *Reliability and Validity Assessment* by E. Carmines and R. Zeller (Sage University Paper)
COURSE SOFTWARE

- STATA 13 or 14 IC or SE
  Students are recommended to purchase a 6-month (or longer) license of Stata, Version 13 or 14, either IC or SE. If you plan to use a large dataset for your project, you may want to invest in SE rather than IC. For the product (IC vs. SE) comparison, see: http://www.stata.com/products/which-stata-is-right-for-me/
- Software Price:
  **Stata/IC software**
  - $ 75.00/six months
  - $125.00/one year
  - $198.00/perpetual
  **Stata/SE software**
  - $235.00/one year
  - $395.00/perpetual
- How to Order:
  **Online:** http://www.stata.com/order
  **Email:** orders@stata.com
  **Phone:** 800-782-8272 (Monday-Friday, 8 to 5 Central Time)
  Delivery is via electronic download. You will receive download instructions when you place your order and then your License and Activation Key within one day.
- If you do not want to purchase the software, you can always use the computers in the Arlington computer lab (3rd floor). Stata should be installed in all machines in the lab. Note that all exams and problem sets require you to use Stata.

*You do NOT need to purchase STATA before the semester begins. We will discuss your options during the first class.*

COURSE TOPICS

### Introduction:
- Elementary Statistical Concepts
- Hypothesis Testing

Research designs and various types of data analysis with Stata
- Descriptive statistics
- Measures of association / contingency tables
- Hypothesis testing (t- and chi-sq tests)
- P-values and Confidence Intervals
- ANOVA

### Regressions and the Diagnosis

- OLS
- Elasticity and other log models
- Goodness-of-fit tests
- Regression with a binary and categorical dependent variable.
- Path analysis and structural equations

### Multivariate Topics

- ANOVA
- Principle components, factor analysis and rotations
- Cluster analysis
### Violation of Regression Assumptions and the Remedies
- Reliability and validity assessment
- Gauss-Markov (GM) and Classical Linear Model (CLM) conditions
- Tests for the GM/CLM violations
- Heteroskedascity and WLS
- Endogeneity and Instrumental Variable (IV)/2SLS regression

### COURSE GRADING

| 3 Problem Sets | 30% |
| Exam (comprehensive) | 30% |
| Project | 30% |
| Class Participation | 10% |

### LECTURE SCHEDULE

<table>
<thead>
<tr>
<th>Lectures</th>
<th>Date</th>
<th>Topic</th>
<th>Homework</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan. 22</td>
<td>Course Overview / Introduction Review of elementary statistics</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Jan. 29</td>
<td>A workshop on available GMU databases by an ARL library librarian. Review of elementary statistics continued: Central Limit Theorem, Law of Large Numbers, Probability distributions</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Feb. 5</td>
<td>Introduction to Stata (Stata Session) Descriptive analysis and hypothesis testing: t-/Fisher’s exact tests, chi-sq tests, binomial tests, spearman’s rank correlation tests, Introduction to ANOVA Regression diagnosis (multicollinearity, goodness of fit, F-test, etc.) <strong>PS1 Distributed</strong></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Feb. 12</td>
<td>One-way and Two-way ANOVA Kruskal Wallis test, Tukey’s HSD test Review OLS Regression: Goodness of Fit ($R^2$), Stepwise regression, Standardized coefficients, Rescaling of variables, Dummy variables and Interaction terms, Prediction for new observations</td>
<td></td>
</tr>
</tbody>
</table>
**Article Review Session: OLS Regression**
Brief introduction to the regressions involving a categorical dependent variable: Mutinomial and ordered logit/probit regressions and McFadden’s discrete choice model

**PS1 Due**

**Article Review Session: Logit and Poisson Regressions**
Brief introduction to Poisson, negative binomial, zero-inflated/truncated regressions

**PS2 Distributed**

**Mar. 12**
Spring Break (No Class)

**Mar. 19**
Principle component factor analysis
Cluster analysis
Reliability and validity assessment

**PS2 Due**

**Mar. 26**
Heteroskedasticity testing, Weighted Least Sq. (WLS) and Robust regressions

**Article Review Session: Factor Analysis**

**PS3 Distributed**

**Mar. 26**
Endogeneity and Omitted variable analyses: 2 Stage Least Sq. (Instrumental Variable) and Heckman’s Selection Bias Analyses

**PS3 Due**

**Apr. 2**
Panel Data Analysis: Fixed and Random Effects models
Hausman’s test

**Project Outline Due**

**Apr. 9**
Brief introduction to path analysis
Brief introduction to time-series analysis (Introduction to PUBP705)
Hierarchical Linear Model (HLM)
Comprehensive Review

**Take-home final exam distributed**

**Apr. 16**
Project Presentation I

**Apr. 30**
Project Presentation II

**May 7**
Reading Day

**Take-home exam Due Project Paper Due**

**READING LIST, LECTURE SLIDES AND HANDOUTS**

Reading list and the materials for the following week will be provided at each class. Lecture slides, handouts, STATA datasets and program files will be uploaded to the blackboard website before noon on the day of each class. The first slide of lecture slides always lists the reading for the following class. Students are encouraged to download all the files into the classroom computer before each class.

**HOMEWORK ASSIGNMENTS AND EXAM**
Both homework (problem sets) and exams will be uploaded to the course blackboard site for download.

**HOMEWORK**
The homework is collected at the end of class on the due date. Late submission will be penalized. Although it is permitted to work as a group, the final product must be original.

**EXAMS**
Both exams are open book and open notes. However, no collaboration is allowed for the exams. Students are expected to solve the exams alone (without any assistance from others). Further, the final products must be original. Late submission will be penalized.

**PROJECT INFORMATION**

1. **PROJECT REQUIREMENTS**
   Students are asked to conduct either his/her own or a team (2 students per group at max) project. Each group is expected to undertake a case study involving a statistical analysis of some data set. The only substantive requirement is that your analysis includes at least one example of advanced analysis you have learned in the class. This analysis should demonstrate a sound statistical knowledge of regression. Each group is expected to find a dataset used for the project (from websites or elsewhere) and make a brief (about 15 -20 minutes) presentation at the last class.

   The final paper is expected to be around 15 pages. The paper should include: (i) motivation of the problem and key question/s, (ii) literature review, (iii) descriptive analysis of your data, (iv) regression model specification, and (v) conclusion and discussion of your findings. All source datasets and material used in the analysis should be cited explicitly.

   Each group is expected to submit an electronic form of the presentation (ppt format). Each group is also expected to submit both hard copy and electronic forms (doc or pdf format) of the final paper.

2. **PAST PROJECT EXAMPLES**
   Past examples of project presentations and papers are available from the blackboard website.

3. **NOTE ON “NEW VOICES IN PUBLIC POLICY”**
   I will consider nominating the very best papers in this course for publication in New Voices in Public Policy. New Voices is a student- and faculty-reviewed journal that shares SPP's finest student work with the rest of the world.

**ACADEMIC ACCOMMODATION FOR A DISABILITY**

If you are a student with a disability and you need academic accommodations, please email and then make an appointment to see Professor Fritschler or Rudder and contact the Disability Resource Center (DRC) at 703-993-2474. All academic accommodations must be arranged through the DRC.
SCHAR SCHOOL POLICY ON PLAGIARISM

The profession of scholarship and the intellectual life of a university as well as the field of public policy inquiry depend fundamentally on a foundation of trust. Thus any act of plagiarism strikes at the heart of the meaning of the university and the purpose of the School of Public Policy. It constitutes a serious breach of professional ethics and it is unacceptable.

Plagiarism is the use of another’s words or ideas presented as one’s own. It includes, among other things, the use of specific words, ideas, or frameworks that are the product of another’s work. Honesty and thoroughness in citing sources is essential to professional accountability and personal responsibility. Appropriate citation is necessary so that arguments, evidence, and claims can be critically examined.

Plagiarism is wrong because of the injustice it does to the person whose ideas are stolen. But it is also wrong because it constitutes lying to one’s professional colleagues. From a prudential perspective, it is shortsighted and self-defeating, and it can ruin a professional career.

The faculty of the School of Public Policy takes plagiarism seriously and has adopted a zero tolerance policy. Any plagiarized assignment will receive an automatic grade of “F.” This may lead to failure for the course, resulting in dismissal from the University. This dismissal will be noted on the student’s transcript. For foreign students who are on a university-sponsored visa (eg. F-1, J-1 or J-2), dismissal also results in the revocation of their visa.

To help enforce the SPP policy on plagiarism, all written work submitted in partial fulfillment of course or degree requirements must be available in electronic form so that it can be compared with electronic databases, as well as submitted to commercial services to which the School subscribes. Faculty may at any time submit student’s work without prior permission from the student. Individual instructors may require that written work be submitted in electronic as well as printed form. The SPP policy on plagiarism is supplementary to the George Mason University Honor Code; it is not intended to replace it or substitute for it.

(http://www.gmu.edu/facstaff/handbook/aD.html)

(return to contents)

Last modified: November 7, 2017