PPPA 6002: Research Methods & Applied Statistics

Semester: Fall 2016 – Thursdays
• 12:45 in Duques 250 – Section 10
• 6:10 in Government 101 – Section 13

Instructor: Bill Adams (adams@gwu.edu)
Office: MPA Bldg. 601-D Phone: 202-994-7494
Meetings: 4:30-6:00 Wednesdays & Thursdays

SPSS Lab:
Arthur Laciak (alaciak@gwmail.gwu.edu)

Textbooks & Software:
• M. Patten, Understanding Research Methods. 9th edition. Glendale, CA: Pyrczak, (years vary.)
• A. Bhattacherjee, Social Science Research, 2012 PDF.
• SPSS 22/23 (available in GW computer labs)
• Many other online readings and class handouts.

Class Web Site: gwu6002.net
This web site will include posts of 6002 updates and links to readings. Please subscribe; be sure to check on the weekend and again before the class.

Assignments and Grades:
• Please post weekly worksheets on Blackboard at least one hour before class; no late postings.
• Late SPSS projects are docked a letter grade.
• Class attendance is required; no more than one unexcused absence.
• Overall = final exam (40%), midterm (20%), five SPSS projects (25%), weekly worksheets (15%).
• Like life, the final exam is cumulative.
• Lowest overall grade (no rounding) for an A is 94.00; A− 90.00; B+ 87.00; B 83.00; B− 80.00; C+ 77.00; C 73.00; and C− 70.00.

• For the weekly worksheets, genuine collaboration is encouraged. However, data analyses of the computer output for SPSS projects should be conducted independently.
• If for any reason a class is missed, in whole or in part, please obtain all announcements and assignments (especially those due at the next session), along with class notes and handouts, from your class colleagues before next class.

Learning Objectives:
PPPA 6002 focuses on practical skills for conducting and evaluating empirical and quantitative research, plus a brief look at qualitative methods. The course explores the strengths and weaknesses of experimental (RCT), quasi-experimental, and nonexperimental research; it also covers the widely used SPSS statistical software and the most widely used statistics (univariate statistics, contingency table statistics, chi-square, t-tests, bivariate regression, correlation, and multiple regression).

In particular, the course targets as learning objectives these ten key research skills:

Skill 1: How to conduct and evaluate survey research
Skill 2: How to create an SPSS data set
Skill 3: How to analyze basic univariate statistics
Skill 4: How to analyze basic bivariate statistics
Skill 5: How to conduct and evaluate RCTs
Skill 6: How to conduct and assess quasi-experiments
Skill 7: How to analyze correlation & regression stats
Skill 8: How to conduct and evaluate focus groups
Skill 9: How to conduct and evaluate content analyses
Skill 10: How to conduct and evaluate meta-analyses
<table>
<thead>
<tr>
<th>Session</th>
<th>6002 Session Topics</th>
<th>Lab</th>
<th>SPSS Due</th>
<th>Quiz</th>
<th>Worksheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1: Sept 1</td>
<td>Field trends; Research questions; Literature review; Research ethics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 2: Sept 8</td>
<td>Question and questionnaire design; Survey sampling systems</td>
<td></td>
<td>[no lab]</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 3: Sept 15</td>
<td>Mail, telephone, e-mail surveys; Measurement validation</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 4: Sept 22</td>
<td>Causal inference and RCT designs</td>
<td>[no lab]</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 5: Sept 29</td>
<td>NEC group designs; Time series and correlational designs</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 6: Oct 6</td>
<td>Qualitative research; Focus groups; Meta-analysis; Intro to statistics</td>
<td></td>
<td>[no lab]</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 7: Oct 13</td>
<td>Univariate descriptive statistics and analysis of crosstabulations</td>
<td>✓ intro</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 8: Oct 20</td>
<td>Mid-term quiz</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Week 9: Oct 27</td>
<td>Confidence intervals Hypothesis testing</td>
<td>✓ new dataset</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 10: Nov 3</td>
<td>( \chi^2 ) (and measures of association)</td>
<td>✓ tables &amp; ( \chi^2 )</td>
<td>1: Dataset &amp; Univariate Stats</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 11: Nov 10</td>
<td>T-tests; Difference of means</td>
<td>✓ t-test</td>
<td>2: Crosstabs</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 12: Nov 17</td>
<td>Bivariate regression and correlation</td>
<td>✓ correl.</td>
<td>3: T-tests</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 13: Dec 1</td>
<td>Multiple regression and correlation</td>
<td>✓ m.reg.</td>
<td>4: Bivariate correlation</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Week 14: Dec 8</td>
<td>Review for final examination</td>
<td></td>
<td></td>
<td>5: Multiple reg</td>
<td>✓</td>
</tr>
<tr>
<td>Dec. 15 (3-5pm)</td>
<td>Final examination</td>
<td>[no lab]</td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
1 Introduction to 6002, plus:
trends in methods; research questions
and hypotheses; literature review;
meta-analyses; research ethics
Readings:
- Links at class web site
- Patten, topics 12-19 (Note “topics” not pages)
- Bhattacherjee, chapter 1-4, 16
- Adams, “Using the Internet” (Blackboard)

Broad 20th century research trends
“Traditionalism”
Behavioralism / basic research
Classic model of scientific research steps
Theory; Hypothesis
Independent variable (X); Dependent variable (Y)
Operationalize concepts; operational definition

Applied research / policy analysis
Model of applied research steps

Basic structure of the written research report:
  Intro; lit review; methodology; findings; discussion
Main goals of literature review
Some tips for upgrading the literature review

Meta-analysis purpose and strengths
Steps in conducting a meta-analysis

Principles of research ethics
Institutional Review Board (IRB)
Informed consent
Issues with coercion, privacy, confidentiality, risks,
deception, common courtesy, debriefing, and vulnerable populations

2 Survey Research, Measurement Validity
and Measurement Reliability
Readings:
- Links at class web site
- Bhattacherjee, chapters 6-9
- Wheelan, chapters 7, 10
- Patten, session 2: topics 24-30
- Patten, session 3: topics 7-8, 31-33, 35, 49

Principles for designing good individual questions
Ways to filter or minimize "random responses"
Best practices for overall questionnaire flow:
  Short intro; easy start; broader to more detailed;
sensitive questions later; demographics at end
Closed-ended vs. open-ended questions
Likert item (strongly agree/agree/disagree/strongly disagree)

Census vs. sample
Random vs. nonrandom samples
Nonrandom (nonprobability) sampling such as
  convenience, snowball, and purposive sampling
Simple random sampling
Systematic random sampling
Stratified random sampling
  (Proportionate vs. nonproportionate)
Probability-proportional-to-size (PPS) sampling
Sampling frame (source/list used to draw sample)

Nonresponse bias vs. response bias
Evaluating response rates; how high is high enough?
Factors for the optimum size of a completed sample

Survey participation as a quick cost-benefit decision
Best practices for improving survey response rates
Benchmark confidence intervals (95% level) for
  n=100 (±10%); 600 (±4%); 1100 (±3%)

Operationalize; operational definition
Measurement reliability & measurement validity
Subjective validity: face validity & content validity
Criterion validity: concurrent & predictive validity

Measurement levels: nominal, ordinal, interval, ratio
Combined interval and ratio = scale or interval
Categorical vs. continuous variable
Unobtrusive measures; multiple measures
4 Causal Inference and RCT Designs

Readings:
• Links at class web site and Blackboard
• Bhattacherjee, chapter 10
• Patten, topics 2, 41-44, 46

Three elements of causal inference...
1) X & Y covary; association; concomitant variation
2) X before Y; direction; time sequence; temporal order
3) Rule out Zs; no plausible alternative; nonspuriousness

Antecedent variables and Intervening variables

Campbell & Stanley's design diagraming system
Single group posttest only
Single group pretest-posttest (aka before-and-after)
Static group design (nonequivalent comparison design)

Threats to internal validity (partial list):
- History
- Maturation
- Practice
- Instrumentation
- Regression to the mean
- Selection
- Intragroup history
- Attrition/Mortality (and how to deal with attrition)

Randomized, controlled trial (RCT) = true experiment

Elements of a true experiment (RCT):
(1) random assignment of subjects from pool to groups and (2) random assignment of X to groups

Reason for the power of experimental designs:
Comparability of the groups (i.e., only real difference between the groups is X, so X is the best explanation for differences in the groups)

Classic experimental design
(aka pretest-posttest control-group design)

Posttest only experiment
(aka posttest-only control-group design)

Factorial designs (simple or complex)
Dosage/sensitivity designs
Complex X
Multiple Ys

RCT’s two essential elements

External validity (generalizability)
Random assignment from pool of subjects to groups strengthens internal validity.
Random selection from the relevant population strengthens external validity.

Reactivity
Hawthorne effects
Placebo
Watch out for between-group reactivity as well as other types of reactivity (e.g., with X and staff)

Big four sets of validity issues:
Construct (measurement) validity
Internal validity
Statistical conclusion validity
External validity

5 NEC, Time Series, Correlational Designs

Readings: • Links at class web site and Blackboard
• Patten, topics 3-4
• Bhattacherjee, chapter 5

Practical reasons why RCTs may not be conducted
Quasi-experiments (vague term)
Causal-comparative designs
Nonequivalent comparison group (NEC) designs
Pretest-posttest nonequivalent comparison design
Posttest only nonequivalent comparison group design

Key threat to internal validity of NEC designs: selection

Retrospective matching design (ex post facto with nonrandom posttreatment matching)
Causal-comparative (another term for studies that try to infer causality with groups not randomly assigned)
Natural experiments (strict vs. broad usage of term)

Time series (aka longitudinal) research
Simple interrupted time series
Reiterative time series
Multiple time series
Panel data (aka "panel-back") vs. cross-sectional data
Deceptive time series charts (truncated base)

Retrospective pretests; proxy pretests
Fallacy of time series inferences from a single survey
Process and logic of correlational designs

Correlational design problems:
Selection threats motivation and self-selection, plus difficulty in statistically controlling all Zs;
specification error / omitted variable bias;
so findings may vary widely depending on the availability and choice of control variables

Ecological fallacy
Aggregate data (units of analysis are groups)

Overall assessment of causal designs:
Lab experiments: often strong on internal validity but weak on external validity
Nonexperimental field studies: often strong on external validity but weak internal validity

Field experiments: strong in both internal and external validity but often not feasible to conduct
The logic of inferring causality by coupling lab experiments with nonexperimental field studies

Checklist for conducting and evaluating research

### Qualitative Research

More exploratory than hypothesis testing
Small, purposive sample, not large random
Extended, intense observations or interviews
Unstructured or semi-structured data gathering
Essay reports with little or no quantitative data
Often explore the researchers’ subjective impact

Focus groups purposes:
Probing attitudes, reaction testing, brainstorming
Focus group: participant recruitment; focus group size; session length and agenda; moderator style;
and ideal focus group facilities

Content analysis
Inter-coder reliability testing
Content analysis steps:
Define scope; Operationalize variables to code
Refine and test coding system;
Code content; Analyze data

### Univariate Descriptive Statistics and Interpretation of Crosstabulations

Readings:
- Wheelan, chapters 1, 2, 3, 8
- Patten, topics 49-50, 53-56
- Links at class web site and Blackboard
- (optional: Bhattacherjee, chapter 14)

Good data analysis requires good data, plus awareness that: all summary statistics are reductionist, context dictates interpretation, small differences should not be exaggerated, correlation does not prove causation, start with univariate analysis before multivariate.

Nominal univariate statistics – percent and mode
Interpretation pitfalls include:
- Misleading pictograms; confusing absolute and relative percent; misinterpreting mode as midpoint; and misleading modal composites

Plurality vs. majority
Measures of central tendency:
- mean and median, plus trimmed mean
Mode (not necessarily a central tendency)
Measures of dispersion:
- standard deviation and interquartile range
Positive skew (high values pull mean above median)
Negative skew (low values pull mean below median)

Normal curve
\[ \pm 1 \text{ standard deviation} = 68.3\% \text{ of normal curve} \]
\[ \pm 2 \text{ standard deviations} = 95.4\% \text{ of normal curve} \]
\[ \pm 3 \text{ standard deviations} = 99.7\% \text{ of normal curve} \]

Value of examining frequency distribution charts
Descriptive vs. inferential statistics
Confidence intervals, sampling error, margin of error

Standard error of the mean (formula optional)
Standard error of the proportion (and formula)
Central limit theorem

Boxplots, stem-and-leaf plots
Histograms, bar charts, pie charts

Interpreting crosstabulations (aka: crosstabs or contingency tables) using counts, row percent, column percent, total percent, and marginal

Lab session: Review for examination over Part 1

### 10 Crosstabs & Chi Square

**Readings:**
- *Patten, topics 47-48, 51-52, 61*
- *Links at class web site and Blackboard*

Chi square ($\chi^2$) test of statistical significance
Interpreting significance levels: .05, .01, .005, .001

Is it acceptable to use .10 as the probability level?

Statistical significance vs. substantive significance
Type I Error vs. Type II Error

Chi square steps:
- State the null hypothesis,
- State the research hypothesis,
- State decision rule (alpha level),
- Examine result to reject or not reject the null

The null is the assumption, but it is never “proven.”

Interpretation caveats:

"Statistically significance" does not mean it is important, powerful, strong; it just means odds are good that there is some kind of relationship. Large samples often show statistical significance for weak, trivial relationships. Conversely, failure to detect a relationship, especially in a small sample, does not mean there is no relationship.

Formula for the chi square test for independence:
\[
\chi^2 (\text{obtained}) = \sum \frac{(f_o - f_e)^2}{f_e}
\]
where $f_o$ = observed cell frequencies and $f_e$ = cell frequencies that would be expected if the variables are independent

Formula to determine expected frequencies:
\[
f_e = \frac{\text{row marginal x column marginal}}{n}
\]

Measures of association (strength) for crosstabs
Gamma (most common measure of association)
**T-Tests**

*Readings: • Links at class web site and Blackboard  
• Adams, chapter 6 (Blackboard)  
• Wheelan, chapter 9  
• Patten, topics 3-4  
• Bhattacherjee, chapter 5*

T-tests: significance test for the difference in the means of two randomized groups – either two means from a sample (random selection) or from an experiment (random assignment)

T-test steps:
- State the null hypothesis
- State the research hypothesis
- State the decision rule (alpha)
- Employ the correct F-test step
- Examine result to reject or not reject the null

Usual issues in interpreting statistical significance...
- stat. significance vs. substantive significance;
- small vs. large samples; careful interpretations if fail to reject null; using the customary .05 criterion

**Bivariate Regression & Correlation**

*Readings:  
• Patten, topic 57  
• Wheelan, chapter 4, 13, conclusion  
• Links at class web site and Blackboard*

r (correlation coefficient)
Correlation coefficient (r) ranges from -1 to 1 and measures the strength of the relationship. No linear relationship when r = 0

Scattergram or scatterplot
- Positive/direct relationship
- Negative/inverse relationship
- Linear and curvilinear patterns

r² (coefficient of determination)
Bivariate regression: ŷ = a + bx
- ŷ (estimate of y); a (intercept); b (slope)
- Residuals represent the difference between actual and predicted values (Y_i − ŷ_i)

Analysis of residuals; analysis of outliers
Homoscedasticity and heteroscedasticity
Standard error of the estimate (analogous to standard deviation)
Ecological fallacy
Aggregate data (units of analysis are groups)

**Multiple Regression & Correlation**

*Readings:  
• Links at class web site and Blackboard  
• Wheelan, chapters 11 and 12*

Regression-discontinuity design (basic design/logic)
Multiple Regression
- Ŷ = a + b₁X₁ + b₂X₂ ... bₙXₙ
- Ŷ = β₁X₁ + β₂X₂ ... βₙXₙ

Multiple R = multiple correlation coefficient
Multiple R² = multiple coefficient of determination

Unstandardized partial regression coefficients
- (aka: unstandardized partial coefficients; unstandardized regression coefficients)

Standardized partial regression coefficients
- (aka: betas; beta weights; standardized partials; standardized regression coefficients)

Limitations of R² even when its value is high:
1. Always increases when add variables (except in the rare case where the additional variable has absolutely no effect)
2. Does not indicate which variables are significant
3. Does not mean that important variables have not been omitted

Dummy variables take the values 0 and 1. Used when nominal or ordinal independent variables are used in the multiple regression.

Multicollinearity
Dummy variables
Standard Trachtenberg School Policies

1. The Syllabus: This syllabus is a guide to the course. Sound educational practice requires flexibility and the instructor may therefore, at her/his discretion, change content and requirements during the semester.

2. Incompletes: A student must consult with the instructor to obtain a grade of “I” (incomplete) no later than the last day of classes in a semester. At that time, the student and instructor will both sign the CCAS contract for incompletes and submit a copy to the School Director. Please consult the TSPPPA Student Handbook or visit the website for the complete CCAS policy on incompletes.

3. Submission of Written Work Products Outside the Classroom: It is the responsibility of the student to ensure that an instructor receives each written assignment. Students can submit written work electronically with the express permission of the instructor.

4. Policy on Late Work: All work must be turned in by the assigned due date in order to receive full credit for that assignment, unless an exception is expressly made by the instructor.

5. Changing Grades after Completion of the Course: No changes can be made in grades after the conclusion of the semester, other than in cases of clerical error.

6. Academic Honesty: GW’s Code of Academic Integrity is at studentconduct.gwu.edu/code-academic-integrity. All exams and other graded work products are to conform to the Code. Its definition of “academic dishonesty” includes “cheating of any kind” and “misrepresenting one’s own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information.”

7. Average Minimum Independent Weekly Work. In addition to the average of over three hours weekly of direct instruction in class and the computer lab, this course should entail a minimum weekly average of 7 hours of independent reading, research, and learning.

8. Religious Holidays: Religiously observant students should notify the instructor the first week of classes regarding any session that will be missed; the courtesy of an absence without penalty will be extended.

9. Accommodation for Students with Disabilities: To receive accommodations on the basis of disability, please provide documentation from the GW’s Disability Support Services, Rome Hall 102 (202-994-8250). For more, see gwired.gwu.edu/dss.

10. Mental Health Services: This GW office 24/7 assistance to address students' personal, social, career, and study skills problems. Services include crisis and emergency mental health consultations, and counseling services as well as referrals: counselingcenter.gwu.edu

GW Bulletin Course Description (bulletin.gwu.edu/courses/pppa)

PPPA 6002. Research Methods and Applied Statistics

Development of skills and knowledge for conducting original research and critically evaluating empirical studies. Various research designs and data collection techniques are examined. Focus on computerizing data sets for quantitative analysis, analyzing strength of relationships, selecting appropriate statistical techniques, and testing statistical hypotheses.
### “To-Do” checklist: Preparation for each session

<table>
<thead>
<tr>
<th>To do before this class session</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
</table>
### Almost every week this semester:

**Internalize key concepts listed on the syllabus and covered in the prior session to get the most out of the course and prep for the midterm and final exams**

**Readings**
- Read chapters and supplemental readings listed on the syllabus for the prior class session.
- Read post(s) and linked readings at class website: www.gwu6002.net

**Worksheets**
- Submit weekly worksheet answers via Blackboard (1 hr or more before class)
- Bring a copy of worksheet questions and your answers to discuss in class.

**SPSS**
- Attend SPSS computer lab after class.
- Prepare a paper copy of your SPSS project to submit in class.

### As scheduled on syllabus:

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
</table>

### Calendar

**September**
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30

**October**
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30

**November**
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30

**December**
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31