ARIZONA STATE UNIVERSITY

GENERAL STUDIES COURSE PROPOSAL COVER FORM

Course information:

*Copy and paste **current** course information from Class Search/Course Catalog.*

Academic Unit School of Politics & Global Studies Department SPGS

Subject POS Number 401 Title Political Statistics Units: 3

Is this a cross-listed course? (Choose one)
If yes, please identify course(s)

Is this a shared course? (choose one) If so, list all academic units offering this course
Course description:

Requested designation: (Choose One)
*Note- a **separate** proposal is required for each designation requested*

Eligibility:

Permanent numbered courses must have completed the university's review and approval process.
For the rules governing approval of omnibus courses, contact Phyllis.Lucie@asu.edu or Lauren.Leo@asu.edu.

Submission deadlines dates are as follow:

For Fall 2015 Effective Date: October 9, 2014

For Spring 2016 Effective Date: March 19, 2015

Area(s) proposed course will serve:

A single course may be proposed for more than one core or awareness area. A course may satisfy a core area requirement and more than one awareness area requirements concurrently, but may not satisfy requirements in two core areas simultaneously, even if approved for those areas. With departmental consent, an approved General Studies course may be counted toward both the General Studies requirement and the major program of study.

Checklists for general studies designations:

Complete and attach the appropriate checklist

- Literacy and Critical Inquiry core courses (L)
- Mathematics core courses (MA)
- Computer/statistics/quantitative applications core courses (CS)
- Humanities, Arts and Design core courses (HU)
- Social-Behavioral Sciences core courses (SB)
- Natural Sciences core courses (SQ/SG)
- Cultural Diversity in the United States courses (C)
- Global Awareness courses (G)
- Historical Awareness courses (H)

A complete proposal should include:

- ☐ Signed General Studies Program Course Proposal Cover Form
- ☐ Criteria Checklist for the area
- ☐ Course Catalog description
- ☐ Course Syllabus
- ☐ Copy of Table of Contents from the textbook and list of required readings/books

Respectfully request that proposals are submitted electronically with all files compiled into one PDF. If necessary, a hard copy of the proposal will be accepted.

Contact information:

Name Meaghan Dirksen Phone 480-727-5568

Mail code 3902 E-mail: Meaghan.dirksen@asu.edu

Department Chair/Director approval: ***(Required)***

Chair/Director name (Typed): Cameron Thies Date: 2/10/15

Chair/Director (Signature):

Rev. 1/94, 4/95, 7/98, 4/00, 1/02, 10/08, 11/11/ 12/11, 7/12, 5/14

Arizona State University Criteria Checklist for

MATHEMATICAL STUDIES [CS]

Rationale and Objectives

The **Mathematical Studies** requirement is intended to ensure that students have skill in basic mathematics, can use mathematical analysis in their chosen fields, and can understand how computers can make mathematical analysis more powerful and efficient. The **Mathematical Studies** requirement is completed by satisfying both the **Mathematics [MA]** requirement and the **Computer/Statistics/Quantitative Applications [CS]** requirement explained below.

The **Mathematics [MA]** requirement, which ensures the acquisition of essential skill in basic mathematics, requires the student to complete a course in College Mathematics, College Algebra, or Pre-calculus; or demonstrate a higher level of skill by completing a mathematics course for which a course in the above three categories is a prerequisite.

The **Computer/Statistics/Quantitative Applications [CS]** requirement, which ensures skill in real world problem solving and analysis, requires the student to complete a course that uses some combination of computers, statistics, and/or mathematics.* Computer usage is encouraged but not required in statistics and quantitative applications courses. At a minimum, such courses should include multiple demonstrations of how computers can be used to perform the analyses more efficiently.

*CS does *not* stand for computer science in this context; the "S" stands for statistics. Courses in computer science must meet the criteria stated for CS courses.

Revised April 2014

Proposer: Please complete the following section and attach appropriate documentation.

ASU--[CS] CRITERIA

A COMPUTER/STATISTICS/QUANTITATIVE APPLICATIONS [CS] COURSE MUST SATISFY ONE OF THE FOLLOWING CRITERIA: 1, 2, OR 3

YES	NO		Identify Documentation Submitted
		1. Computer applications*: courses must satisfy both **a** and **b**:	
☒	☐	**a.** Course involves the use of computer programming languages or software programs for quantitative analysis, algorithmic design, modeling, simulation, animation, or statistics.	See yellow highlighted areas and text chapters
		b. Course requires students to analyze and implement procedures that are applicable to at least one of the following problem domains **(check those applicable)**:	
☒	☐	i. Spreadsheet analysis, systems analysis and design, and decision support systems.	See yellow highlighted areas and text chapters
☐	☒	ii. Graphic/artistic design using computers.	
☐	☒	iii. Music design using computer software.	
☒	☐	iv. Modeling, making extensive use of computer simulation.	See syllabus
☒	☐	v. Statistics studies stressing the use of computer software.	See yellow highlighted area and texts chapters
☒	☐	vi. Algorithmic design and computational thinking.	See course schedule and topic oulines in syllabus

*The **computer applications** requirement **cannot** be satisfied by a course, the content of which is restricted primarily to word processing or report preparation skills, the study of the social impact of computers, or methodologies to select software packages for specific applications. Courses that emphasize the use of a computer software package are acceptable only if students are required to understand, at an appropriate level, the theoretical principles embodied in the operation of the software and are required to construct, test, and implement procedures that use the software to accomplish tasks in the applicable problem domains. Courses that involve the learning of a computer programming language are acceptable only if they also include a substantial introduction to applications to one of the listed problem domains.

YES	NO		Identify Documentation Submitted
		2. Statistical applications: courses must satisfy **a, b,** and **c.**	
☐	☒	**a.** Course has a minimum mathematical prerequisite of College Mathematics, College Algebra, or Pre-calculus, or a course already approved as satisfying the MA requirement.	See explanation re: critical tracking
		b. The course must be focused principally on developing knowledge in statistical inference and include coverage of all of the following:	
☒	☐	i. Design of a statistical study.	See green highlighted areas
☒	☐	ii. Summarization and interpretation of data.	See green and dark purple highlighted areas
☒	☐	iii. Methods of sampling.	See dark blue highlighted area and Chptrs 6 & 7 from the text
☒	☐	iv. Standard probability models.	See red highlighted areas and Ch 10 in the companion text
☒	☐	v. Statistical estimation	See purple highlighted areas
☒	☐	vi. Hypothesis testing.	See blue highlighted area
☒	☐	vii. Regression or correlation analysis.	See purple highlighted areas
☒	☐	**c.** The course must include multiple demonstrations of how computers can be used to perform statistical analysis more efficiently, if use of computers to carry out the analysis is not required.	See yellow highlighted areas, chapters in companion text,

YES	NO		Identify Documentation Submitted
		3. Quantitative applications: courses must satisfy **a, b, and c:**.	
☐	☒	**a.** Course has a minimum mathematical prerequisite of College Mathematics, College Algebra, or Pre-calculus, or a course already approved as satisfying the MA requirement.	See explanation re: critical tracking
		b. The course must be focused principally on the use of mathematical models in quantitative analysis and decision making. Examples of such models are:	
☒	☐	i. Linear programming.	See purple highlighted areas and correlating text chapters.
☐	☒	ii. Goal programming.	
☐	☒	iii. Integer programming.	
☐	☒	iv. Inventory models.	
☐	☒	v. Decision theory.	
☒	☐	vi. Simulation and Monte Carlo methods.	See purple highlighted texts and their correlating text chapters
☐	☒	vii. Other (explanation must be attached).	
☒	☐	**c.** The course must include multiple demonstrations of how computers can be used to perform the above applications more efficiently, if use of computers is not required by students.	See yellow highlighted areas, chapters in companion text,

Course Prefix	Number	Title	General Studies Designation
POS	401	Political Statistics	

Explain in detail which student activities correspond to the specific designation criteria. Please use the following organizer to explain how the criteria are being met.

Criteria (from checksheet)	How course meets spirit (contextualize specific examples in next column)	Please provide detailed evidence of how course meets criteria (i.e., where in syllabus)
1	This course is critically tracked in term 6. The College of Liberal Arts and Sciences math requirement must be completed by term 3.	Students will have completed their math requirement prior to enrolling in POS 401 due to critical tracking.
2	This course covers various social science statistical methods, including: probability models, controlled comparisons, correlation, regression. Applications of the techniques are included that draw on standard social science resaerch designs and methods. The applications require students to design and use approparиate statistical methods.	See highlighted colors as well as text chapters. See the homeowrk assignments and problem sets.
3	This course uses R to analyze data. Students also use it in their problem sets and are tested on it.	See yellow highlighted areas and included text chapters. See the homework assignments and problem sets.

Course Catalog Description for:

POS 401 Political Statistics: Basic concepts in statistics as they facilitate the description, explanation, and prediction of social and political phenomena.

Arizona State University, Department of Political Science Political Science 401: Political Statistics, Spring 2015

Professor: Dr. Rodolfo Espino
Phone: 480-965-5884, email: respino1@asu.edu
Time/Location: T/TH, 3:00-4:15PM, Tempe WGHL 101
Office: 6732 Lattie Coor Hall

Office Hours: T/Th 1:15-2:45PM, or by appoint- ment

TA: Erik Bumgardner email: ebumgard@asu.edu

Course Description

This course is designed to introduce you to the mathematical techniques and logical processes that are used by social science researchers to organize, summarize, and interpret the results from their research studies. During the semester you will be required to demonstrate your ability to use the statistical procedures we will cover in the course. You will be required to demonstrate your abilities on homework assignments and three comprehensive examinations. In the long term, you should understand why the different statistical methods have been developed (i.e., what purpose they serve, the circumstances in which they are used, and what questions they are designed to answer).

Course Evaluation

Your grade in the course will be based on my evaluation of your performance in the following four areas:

- Problem Sets (40%)
- Two Midterm Exams (30%)
- Final Exam (20%)
- Attendance - Ten Pop Quizzes (10%)
- Research Study Participation - Pass/Fail for the course

Problem Sets. Approximately once a week, you will have problem sets related to the text readings and course lectures. Typically, you will have one week to complete the assignments. Failure to complete the assignments by the assigned deadline will result in no credit for that assignment.
Midterm Exams. The midterm exams will cover material we have covered up to that point in time on the syllabus. The dates of the midterm exams are TBA.
Final Exam. The Final Exam is comprehensive.
Attendance. **Active** attendance is critical to understanding the material covered in the course, doing well on exams, and completing the quizzes and problems accurately and efficiently. I will encourage attendance by administering four pop quizzes over the course of the semester.
Research Study Participation Requirement. You are required to participate in a research study for this course. Failure to participate in the study will result in failing this course. More details about this are explained the next section of the syllabus.

Research Study Participation Requirement

Students enrolled in this course are required to complete a research assignment that can include up to 2 hours of research study participation. These studies require that students set up an appointment to complete participation at a laboratory on campus (or via an on-line survey). Students will learn how studies are conducted and will receive a synopsis at the conclusion of the semester describing the study's goal, result, and relevance to the class. Students who prefer not to participate in research as subjects may opt for an alternative that entails reading one chapter about political science research and writing a three-page reaction paper. The typical chapter is about 20 pages and thus reading it and writing a three-page paper should take approximately two hours.

During the first week of the semester, students will receive an announcement and e-mail through Blackboard including details on how to complete either requirement. Note that if you are enrolled in multiple POS classes that require participation, you only need to satisfy the requirement one time.

The following website will direct you to the SPAGS experimental website where you can register for an experimental session.

http://spgslab.wordpress.com/experimental-participation-sign-up/

Completing the research requirement is REQUIRED for you to pass this course.

Course Administration

Absences

I cannot review lecture material for students on an individual basis. It is expected that students who miss class should rely upon fellow students for material covered in class. Make up exams will not be given. Exceptions will only be granted for extreme circumstances, which must be documented.

Disability Resources

In order to be fair, it is imperative that students with disability accommodation requests clear such re- quests with the office of Disability Resources for Students. I am not a qualified medical professional on such matters. This office works closely with both students and professors to be sure all accommodations are fair and adequate.

Email

I will expect that:

1. You check your email at least once every 48 hours.
2. You have an ASU email address or your primary email address is registered with ASU. If not, you will miss important class announcements.
3. Be sure your email account is not full. If it is, I am not responsible for any email messages you do not receive.

Late Assignments

Assignments must be turned in by the scheduled deadline. There will be no exceptions. Save yourself headaches but avoiding procrastination.

Incompletes

It is my practice not to grant incompletes other than for extreme circumstances approved by me. Again, an extreme circumstance in your mind may not constitute a valid excuse in my mind.

Academic Misconduct

It should go without saying but I need to say it anyway: Any form of academic dishonesty will not be tolerated and at a minimum will result in failure of the course. View "ASU Academic Integrity" policy online to see what constitutes academic misconduct according. to Arizona State University and the legal sanctions that will be imposed. Quite simply, it is against Arizona state law for any form of academic misconduct. I will strictly enforce all policies by the state of Arizona in instances of academic misconduct.

Textbooks

The books I ordered for this class are

Pollock, Phillip H. 2012. *The Essentials of Political Analysis*, 4th ed. Pollock, Phillip H. 2014. *An R Companion to Political Analysis.*

Computing Language

For the course, we will be using R - an efficient but powerful language written primarily for statistical usage. You can access R via ASUs citrix website, or you can download it locally to your own computer at www.r-project.org. Two advantages to R are 1) it is free, and 2) it is compatible with any major operating system. The former can also be considered a disadvantage to R as there is no helpful supporting manuals as are typically provided with proprietary software. However, over the last several years, the usage of R has increased dramatically - thus, leading to the production of free, helpful documentation/email listserves/blogs that I encourage you to make use of. In the end, you will find that the key to learning how to work with R is trial and error - you will make mistakes and you should not be afraid of making mistakes. To the goal of helping you learn R, I will make myself available during class time and during office hours to help you. However, I ask that before you seek out my assistance that you put forth a serious effort on your own to solve the problem yourself and provide documentation to me that you have made a good faith effort to do so.

Course Schedule

What follows is a topical outline of the course schedule.

Topic One

- Introductions
- Concepts, Variables, Measurement
- Chapter 1 of Essentials and Companion

Topic Two

- Hypothesis Testing
- Chapter 2 of Essentials and Companion

Topic Three

- Descriptive Statistics
- Chapter 3 of Essentials, Chapter 3-4 of Companion

Topic Four

- Controlled Comparisons
- Chapter 4 of Essentials, Chapter 6 of Companion

Topic Five

- Inferential Statistics
- Chapter 5 of Essentials, Chapter 6 of Companion

Topic Six

- Measures of Association
- Chapter 6 of Essentials, Chapter 7 of Companion

Topic Seven

- Correlation & Regression
- Chapter 7 of Essentials, Chapter 8 of Companion

Topic Eight

- Regression (cntd)
- Chapter 9 of Companion

Topic Nine

- Logistic Regression
- Chapter 8 of Essentials, Chapter 10 of Companion

Topic Ten

- Other Forms of Categorical Data Analysis
- Chapter 9 of Essentials, Chapter 11 of Companion

Topic Eleven

- Other Forms of Categorical Data Analysis (cntd)

AN R COMPANION TO Political Analysis

PHILIP H. POLLOCK III

"Pollock successfully organizes an introduction to R and its use for statistical analysis so that it is no longer difficult to learn. The Companion helps readers with concise instructions using a minimum of technical jargon. Pollock's understanding of social science is reflected in his careful treatment of data management and in the way he provides building blocks for performing statistical analysis. This R workbook is the first one written for social scientists, and is long overdue. I highly recommend this book for students and even researchers who want to start using R."

—Toshiyuki Yuasa, University of Houston

"Philip H. Pollock has written a timely, useful, and well-written book to accompany his popular text, *The Essentials of Political Analysis*. The use of R in the classroom is increasing each year, and the need for 'user-friendly' books to help integrate methodological training with this powerful statistical language has reached a critical stage. Professor Pollock's book fills this gap superbly. It takes the student from the elements of installing R on their own computer or laptop through the use of R to solve both simple and complex problems in social and political analysis. Students will love this book, as will their teachers."

—Courtney Brown, Emory University

"R is rapidly becoming the tool of choice for doing statistical analysis in political science. This companion text introduces students to R in the context of Pollock's excellent *The Essentials of Political Analysis*. But I was particularly impressed that *An R Companion to Political Analysis* also stands alone as a gentle guide to the mechanics of using R for political statistics. Drawing on real world data and providing clear step-by-step instructions Pollock demonstrates the basic procedures that will be required for any introductory or intermediate statistics class. This accessible introduction makes R a viable option for undergraduate classes in political analysis."

—Kurt T. Gaubatz, Old Dominion University

KEY FEATURES

- **Step-by-step instructions** and labeled screen shots offer students clear guidance and visual explanation.
- **Engaging exercises and research-quality datasets** provide students with hands-on practice and skill application.
- **Concepts are reinforced** by requiring students to write and run R scripts from the script window.
- **Local functions** help students automate recurring tasks.
- **A Closer Look boxes** teach students how to write R commands and are integrated throughout the workbook for easy reference.
- A full **solutions manual** is available for instructors.

SAGE | CQPRESS
Los Angeles | London | New Delhi
Singapore | Washington DC
www.cqpress.com

ISBN 978-1-4522-8731-7
90000
9 781452 287317

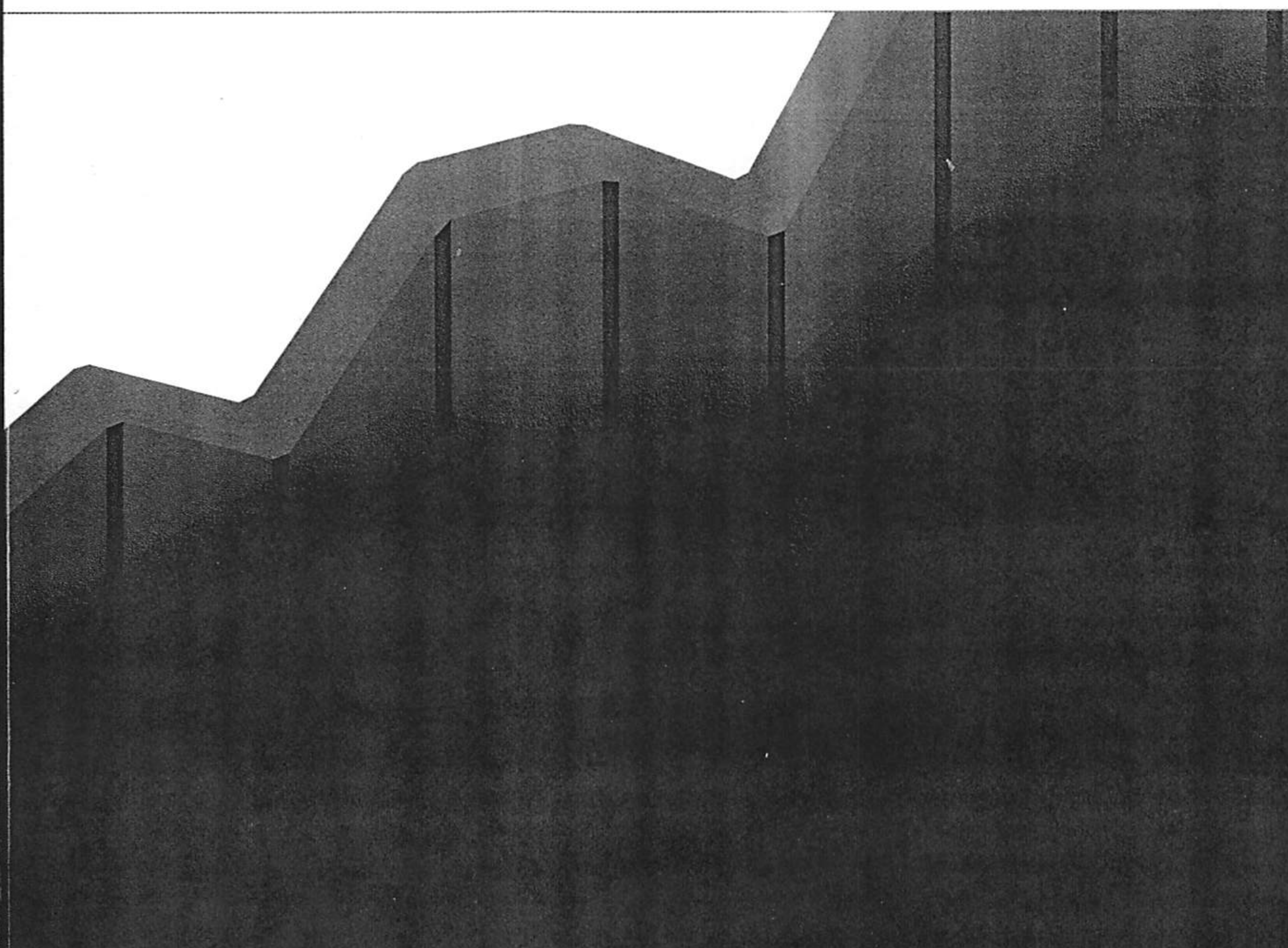
Contents

The Essentials of

Political Analysis

FOURTH EDITION

Philip H. Pollock III



The Essentials of Political Analysis
FOURTH EDITION

"Overall, this is a very good introduction to political science research methods. The *SPSS Companion* is also quite useful for students who have not used any statistical software before. The writing and examples are very clear and interesting."

—**Muhammet Bas**, *Harvard University*

"*Essentials* is very concise but covers most of the bases I do in my introductory research methods course. I like that it has lots of examples. Good, strong definitions of terms match up well with the way I teach my class. It is a well-organized text and clearly written. I switched to the *Essentials* text to use with the *Companion*—the two work well together."

—**Kathleen Ferraiolo**, *James Madison University*

"*The Essentials of Political Analysis* is well written and generally well organized. It teaches the principles of empirical political research in a manner that is neither dense nor intimidating, focusing on the application of data analysis techniques and avoiding an overabundance of equations. The exercises are very good, and my students learn a lot from doing them."

—**Jonathan Hanson**, *Syracuse University*

Philip H. Pollock III is professor of political science at the University of Central Florida. He has taught courses in research methods at the undergraduate and graduate levels for more than thirty years. His main research interests are American public opinion, voting behavior, techniques of quantitative analysis, and the scholarship of teaching and learning. His recent research has been on the effectiveness of Internet-based instruction. Pollock's research has appeared in the *American Journal of Political Science, Social Science Quarterly,* and the *British Journal of Political Science*. Recent scholarly publications include articles in *Political Research Quarterly,* the *Journal of Political Science Education,* and *PS: Political Science and Politics*.

BUY THE PACKAGE AND SAVE! Students save money when they buy *The Essentials of Political Analysis* together with Pollock's *SPSS Companion to Political Analysis* or *Stata Companion to Political Analysis*. Visit **college.cqpress.com** for more information.

A **solutions manual** to all of the book's exercises is available free to adopters. Go to **http://college.cqpress.com/instructors-resources/pollock/** to register and download the manual.

SAGE | CQPRESS

Los Angeles | London | New Delhi
Singapore | Washington DC

www.cqpress.com

ISBN 978-1-60871-686-9

9 781608 716869

Contents

Tables

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 1

Instructions: This assignment will require you to use the CPS data available on Blackboard. Include all your answers and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console. Answer the following:

1. Report the overall mean, median, quantiles, and interquartile range for Wages.
2. Show a histogram for Wages
3. Show a density plot for Wages
4. Show a histogram and density plot for Wages
5. Show a box plot for Wages

Your assignment is due on Friday, January 30, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 1, which may be found on the class Blackboard page. No late work will be accepted.

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 2

Instructions: Review Chapter 2 of the Pollock (2014) book, *An R Companion to Political Analysis*, and answer Questions 2, 3, and 4 at the end of the chapter. Please include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Wednesday, February 18, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 2, which may be found on the class Blackboard page. No late work will be accepted. 1

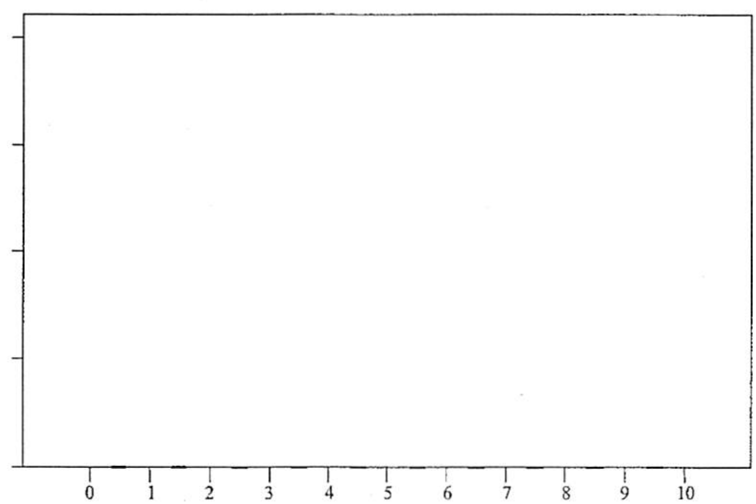
E. Create a nicely labeled version of the histogram you produced in Part A. Give the horizontal (x-axis) the following label: "Percentage of Women in Legislature." Give the chart this main title: "Percentage Women Legislators 111 Democracies." Print the histogram.

2. (Data set: gss. Variables: science.quiz.f, wtss. Remember that gss must be weighted by wtss.) The late Carl Sagan once lamented, "We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about science and technology." This is a rather pessimistic assessment of the scientific acumen of ordinary Americans. Sagan seemed to be suggesting that the average level of scientific knowledge is quite low and that most people would fail even the simplest test of scientific facts.

Dataset gss contains science.quiz.f, an ordered factor created from 10 questions testing respondents' knowledge of basic scientific facts. Values on science.quiz.f range from 0 (the respondent did not answer any of the questions correctly) to 10 (the respondent correctly answered all 10).[8]

A. Consider three possible scenarios for the distribution of science.quiz.f. All three scenarios assume that science.quiz.f has a median value of 6 on the 10-item scale. In scenario X, science.quiz.f has no skew. In scenario Y, science.quiz.f has a positive skew. In scenario Z, science.quiz.f has a negative skew. Below are three graphic shells, labeled Scenario X, Scenario Y, and Scenario Z. Sketch a bar chart, depicting what the distribution of science.quiz.f would look like if that scenario were accurate.

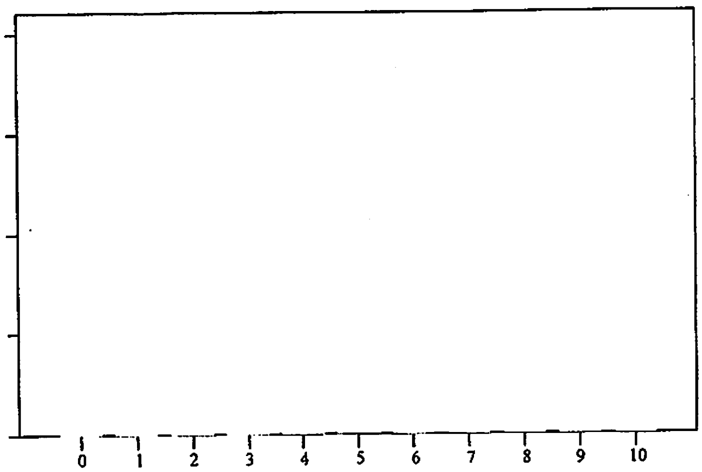
Scenario X: Median=6, No Skew



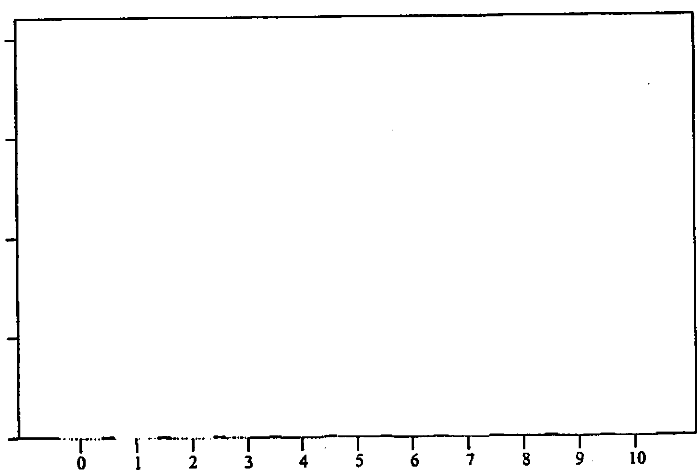
8. Science_quiz was created by summing the number of correct responses to the following questions (all are in true-false format, except for earthsun):

- The center of the Earth is very hot (General Social Survey variable, hotcore);
- It is the father's gene that decides whether the baby is a boy or a girl (boyorgrl);
- Electrons are smaller than atoms (electron);
- The universe began with a huge explosion (bigbang);
- The continents on which we live have been moving their locations for millions of years and will continue to move in the future (condrift);
- Human beings, as we know them today, developed from earlier species of animals (evolved);
- Does the Earth go around the Sun, or does the Sun go around the Earth? (earthsun);
- All radioactivity is manmade (radioact);
- Lasers work by focusing sound waves (lasers);
- Antibiotics kill viruses as well as bacteria (viruses).

Scenario Y: Median=6, Positive Skew



Scenario Z: Median=6, Negative Skew



B. Run freq to obtain a frequency distribution and bar chart of science.quiz.f, making sure to include the gss weight variable, wtss. (The freqC function is unnecessary because science.quiz.f does not have lengthy value labels.) In the following table, fill in the Valid Percent and Cumulative Percent.

science.quiz.f	*Valid Percent*	*Cum Percent*
0	?	?
1	?	?
2	?	?
3	?	?
4	?	?
5	?	?
6	?	?
7	?	?
8	?	?
9	?	?
10	?	100.00
Total	100.00	

C. Examine the frequency distribution and the bar chart. Based on your analysis, which scenario does the distribution of science.quiz.f most closely approximate? (Circle one)

Scenario X   Scenario Y   Scenario Z

Briefly explain your reasoning.

____

____

____

____

D. According to conventional academic standards, any science.quiz.f score of 5 or lower would be an F, a failing grade. A score of 6 would be a grade of D, a 7 would be a C, an 8 a B, and scores of 9 or 10 would be an A. Based on these standards, about what percentage of people got passing grades on science.quiz.f? (Circle one)

About 30 percent   About 40 percent

About 50 percent   About 60 percent

What percentage got a C or better? (Circle one)

About 30 percent   About 40 percent

About 50 percent   About 60 percent

E. Print the chart that you created for this exercise.

3. (Dataset: gss. Variables: fem.role.f, fem.role, wtss. Remember that gss must be weighted by wtss.) Two pundits are arguing about how the general public views the role of women in the home and in politics.

Pundit 1: "Our society has a sizable minority of traditionally minded individuals who think that the proper 'place' for women is taking care of the home and caring for children. This small but vocal group of traditionalists aside, the typical adult supports the idea that women belong outside the home and in the workplace."

Pundit 2: "Poppycock! It's just the opposite. The extremist 'women's liberation' crowd has distorted the overall picture. The typical view among most citizens is that women should be in the home, not in work."

A. The gss dataset contains fem.role, a numeric variable that measures respondents' attitudes toward women in society and politics. Gss also contains fem.role.f, an ordered factor version of fem.role. For both fem.role and fem.role.f, scores can range from 0 (women belong in the home) to 9 (women belong in work).

If pundit 1 is correct, fem.role will have (Circle one)

a negative skew.   no skew.   a positive skew.

If pundit 2 is correct, fem.role will have (Circle one)

a negative skew.   no skew.   a positive skew.

If pundit 1 is correct, fem.role's mean will be (Circle one)

lower than its median.   the same as its median.   higher than its median.

If pundit 2 is correct, fem.role's mean will be (Circle one)

lower than its median.   the same as its median.   higher than its median.

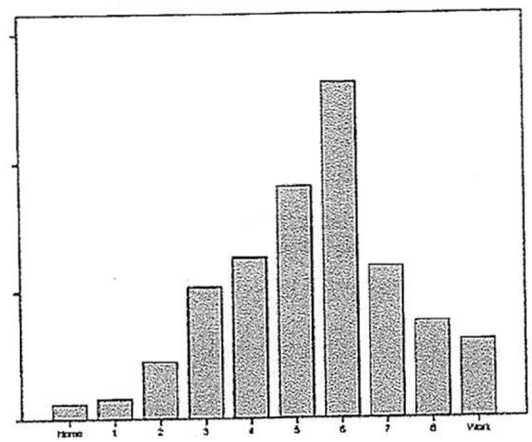
B. Run describe on fem.role. Run freq on fem.role.f. Examine the describe output, the frequency distribution, and the bar chart. Fill in the blanks.

Statistics for fem.role and fem.role.f

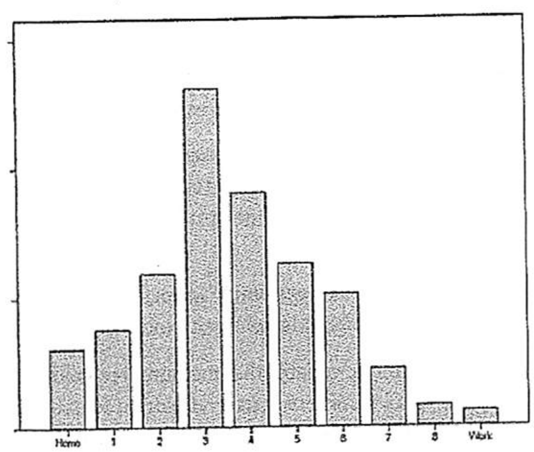
Mean	?
Median	?
Mode	?

C. Which of the following bar charts—X, Y, or Z—most closely resembles the bar chart you obtained? (Circle one)

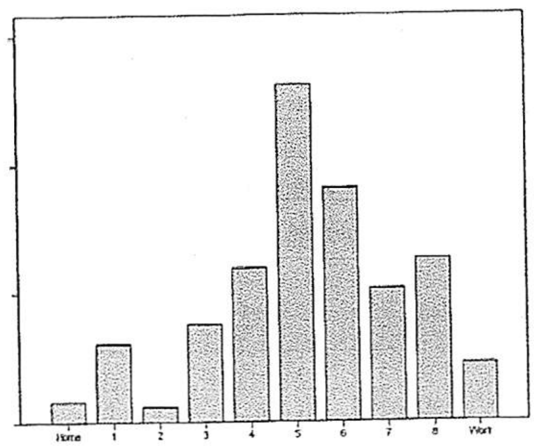
Bar chart X



Bar chart Y



Bar chart Z



D. Based on your analysis, whose assessment is more accurate? (Circle one)

Pundit 1's Pundit 2's

Explain your reasoning.

____

____

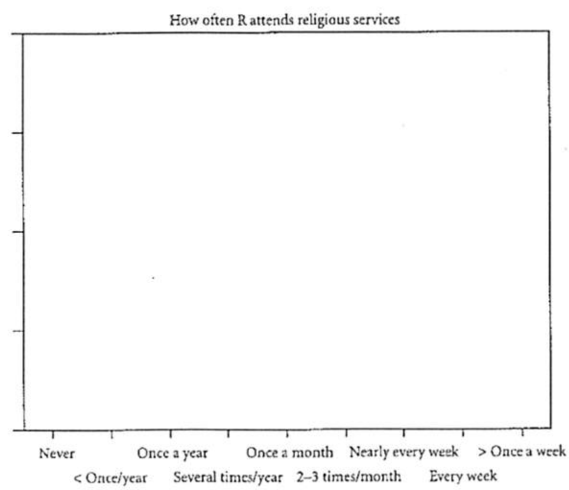
____

____

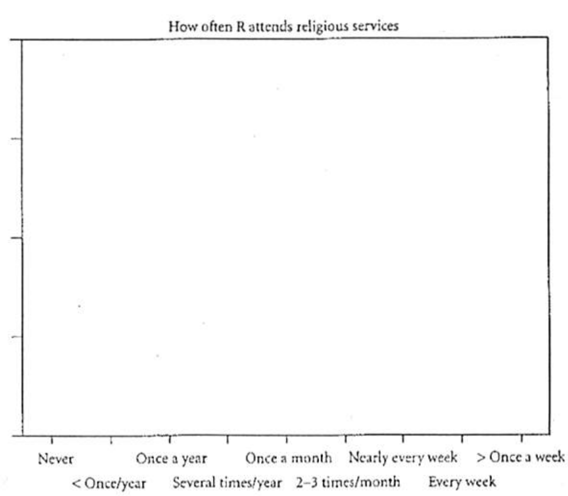
E. Print the bar chart you created in Part B.

4. (Dataset: gss. Variable: attend.f, wtss. Remember that gss must be weighted by wtss.) The General Social Survey provides a rich array of variables that permit scholars to study religiosity among the adult population. Gss contains attend.f, a 9-point ordered factor that measures how often respondents attend religious services. Values can range from 0 ("Never attend") to 8 ("Attend more than once a week").

A. The shell of a bar chart is given below. The categories of attend.f appear along the horizontal axis. What would a bar chart of attend look like if this variable had maximum dispersion? Sketch inside the axes a bar chart that would depict maximum dispersion.



B. What would a bar chart of attend look like if this variable had no dispersion? Sketch inside the axes a bar chart that would depict no dispersion.



C. Run freqC on attend.f. (The freqC function will accommodate attend.f's long value labels.)

The mode of attend.f is ______________. The median of attend.f is ______________.

D. Based on your examination of the frequency distribution and bar chart, you would conclude that attend.f has (Circle one)

low dispersion. high dispersion.

E. Print the bar chart you created in Part C.

F. Exercise a skill you learned in Chapter 1. Use printC to produce a nicely edited table of the frequency distribution you created in Part C. Edit the table in Word. Print the edited table.

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 3

Instructions: Review Chapter 3-4 of the Pollock (2014) book, *An R Companion to Political Analysis*. Answer questions 1 and 4 from Chapter 3 and question 1 from Chapter4. Include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Wednesday, March 4, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 3, which may be found on the class Blackboard page. No late work will be accepted.

Another beautiful four-category ordinal has been added to nes. Remember, before starting the exercises, be sure to save the workspace.

EXERCISES

1. (Dataset: gss. Variables: polviews, wtss.) The dataset gss contains polviews, a seven-category factor that measures political ideology—the extent to which individuals "think of themselves as liberal or conservative."

   A. Run freq on gss$polviews, making sure to weight by gss$wtss. Add up the percentages of all liberals. The percentage of respondents who are either "extremely liberal," "liberal," or "slightly liberal" is (Fill in the blank) ______________ percent.

   B. The percentage of "moderates" is equal to (Fill in the blank) ______________ percent.

   C. Now add up the percentages of all conservatives. The percentage of respondents who are either "slightly conservative," "conservative," or "extremely conservative" is (Fill in the blank) ______________ percent.

   D. Use cut2 to create a new variable, gss$polview3, which combines the three liberal categories, keeps moderates in their own category, and combines the three conservative categories. Cut2 does not work on factors, so use the as.numeric function to create a numeric version of gss$polviews. Name the numeric version gss$polviews.n. Run freq on gss$polviews.n (again weighting by gss$wtss).

   On gss$polviews.n, liberals have the following three numeric codes: _____, _____, _____.

   On gss$polviews.n, moderates have the following numeric code: _____.

   On gss$polviews.n, conservatives have the following three numeric codes: _____, _____, _____.

   E. Use cut2 to create a new variable, gss$polview3. Collapse gss$polviews.n's three liberal codes, keep moderates in their own category, and collapse the three conservative codes. Remember to apply cut2's "plus-1" rule. Run freq on gss$polview3, weighting by gss$wtss. Record the percentages in each category (Fill in the blanks):

   [1, 4) _______ 4 _______ [5, 8) _______

   Before proceeding, make sure that the percentages you just recorded match the percentages (rounded to two decimals) you calculated in Parts A, B, and C. If they match, move on to Part F. If they do not match, check the cut2 expression that you typed. Make sure that you followed the plus-1 rule. Also, check to ensure that your freq runs in Parts A and E included gss$wtss.

   F. Supply gss$polview3 with the following level names: "Liberal," "Moderate," "Conservative." Use as.ordered to define gss$polview3 as an ordered factor. To check your work, run freq on gss$polview3. Run freq on gss$polview3 once more, this time wrapping it inside printC. Copy/paste the frequency distribution from Table.Output.html into your word processor. Print the table.

2. (Dataset: gss. Variables: income06, wtss.) The gss dataset contains income06. In this exercise you use cut2 (with the "g=" argument) to collapse gss$income06 into four equal-size groups. Follow these steps: (i) Create a numeric version of gss$income06. Name the numeric version, gss$income06.n. (ii) Collapse gss$income06.n into four equal-size groups. Name the collapsed variable, gss$income06.n4. (iii) Give gss$income06.n4 these level names: "Low," "MedLow," "MedHigh," "High", (iv) Redefine gss$income06.n4 as an ordered factor.

   A. Run freq on gss$income06.n4. Print the bar chart of gss$income06.n4.

   B. Run printC on the frequency distribution of gss$income06.n4. Copy/paste the frequency distribution into a word processing document. Print the frequency table.

3. (Dataset: gss. Variables: grass, wtss.) The gss dataset contains the factor, grass, which taps respondent opinions about the legalization of marijuana. In this exercise, you use gss$grass to create an indicator, named gss$grass.yes, which is coded 1 for respondents favoring legalization and coded 0 for those who do not favor legalization.

   A. Run freq on gss$grass, making sure to weight by gss$wtss. What percentage favors legalization? (Fill in the blank) ______ percent.

B. Run levels on gss$grass. Using the level name for prolegalization, enlist as.numeric to create an indicator, named gss$grass.yes, coded 1 for those who favor legalization and coded 0 for those who do not favor legalization. Run freq on gss$grass.yes. What percentage is coded 1 on gss$grass.yes? (Fill in the blank) ______ percent.

Before proceeding, make certain that the percentage you just recorded in Part B matches the percentage you recorded in Part A. If they do not match, review the procedure for creating indicator variables.

4. (Dataset: gss. Variables: mslm.col.n, mslm.lib.n, mslm.spk.n, wtss.) The gss dataset contains three numerics that gauge tolerance toward "anti-American Muslim clergymen"—whether they should be allowed to teach in college (gss$mslm.col.n), whether their books should be removed from the library (gss$mslm.lib.n), and whether they should be allowed to preach hatred of the United States (gss$mslm.spk.n). For each variable, a less-tolerant response has a numeric code of 1, and a more-tolerant response is coded 2.

A. Imagine creating an additive index from these three variables. The additive index would have scores that range between what two values?

Between a score of ________________ and a score of ________________.

B. Suppose a respondent takes the more-tolerant position on two questions and the less-tolerant position on the third question. What score would this respondent have?

A score of ________________.

C. Create an additive index, named gss$muslim.tol, by summing gss$mslm.col.n, gss$mslm.lib.n, and gss$mslm.spk.n. Run freq on gss$muslim.tol, weighting by gss$wtss. Record the scores, frequencies, and valid percentages in the table below.

Score on gss$muslim.tol	*Frequency*	*Valid Percent*
?	?	?
?	?	?
?	?	?
?	?	?

D. Use cut2 to recode gss$muslim.tol into three defined categories, as follows: Make the lowest code its own category; combine the middle two codes into one category; make the highest code its own category. Make sure to apply the plus-1 rule of cut2. You can give the collapsed variable a new name, or you can continue with the name gss$muslim.tol. (If the cut2 run goes badly, you can easily go back and re-create gss$muslim.tol from the three original variables.) Run freq on the newly collapsed variable. Make sure that your cut2 run correctly collapsed the middle two codes and left the lowest and highest codes unchanged. If the collapsed variable is incorrect, review the cut2 procedure for collapsing a numeric into defined groups.

E. Define gss$muslim.tol as an ordered factor. Supply these level names: "Low," "Middle," "High." To check your work, run freq on gss$muslim.tol. Run printC on the freq command. Copy/paste from Table.Output.html. Print the frequency distribution.

That concludes the exercises for this chapter. Before quitting R, be sure to save the workspace.

SUMMARY

In this chapter, you learned to perform cross-tabulation analysis and mean comparison analysis with a control variable, and you learned to produce multiple-line charts for numeric dependent variables and for indicator dependent variables. All guided examples have used weighted data. As you now know, because the plotting functions do not allow weights, getting from xtp or compmeans to interaction.plot required a side-trip through ddply. However, even if you are analyzing and graphing unweighted data, such as states or world, this book strongly recommends that you first define a ddply object containing the unweighted means.[5]

Following is a summary of the procedures for obtaining analytic output and graphic support, depending on whether the dependent variable is numeric (interval-level) or factor-class (nominal/ordinal).

Scenario A: Numeric dependent variable

1. Use subset to create separate datasets for each value of the control variable.
2. Run compmeans on the subsets. For weighted data, specify a weight variable.
3. Use ddply to create an object containing the means of the dependent variable.
4. Run interaction.plot, graphing the means from the ddply object.

Scenario B: Nominal or ordinal dependent variable

1. Use subset to create separate datasets for each value of the control variable.
2. Run xtp on the subsets. For weighted data, specify a weight variable.
3. Create an indicator variable, coded 1 for the category of the dependent variable you want to graph.
4. Use ddply to create an object containing the means of the indicator variable.
5. Run interaction.plot, graphing the means from the ddply object. In the interaction.plot syntax, multiply the dependent variable times 100 ('*100') to display percentages.

EXERCISES

1. (Dataset: world. Variables: democ.regime, frac.eth3, gdp.cap2.) Some countries have democratic regimes, and other countries do not. What factors help to explain this difference? One idea is that the type of government is shaped by the ethnic and religious diversity in a country's population. Countries that are relatively homogeneous with most people sharing the same language and religious beliefs are more likely to develop democratic systems than countries having more linguistic conflicts and religious differences. Consider the ethnic heterogeneity hypothesis: Countries with lower levels of ethnic heterogeneity are more likely to be democracies than countries with higher levels of ethnic heterogeneity.

   A. According to the ethnic heterogeneity hypothesis, if you compare countries having lower heterogeneity with countries having higher heterogeneity, you should find (Check one)

   - ☐ a lower percentage of democracies among countries having lower heterogeneity.
   - ☐ a higher percentage of democracies among countries having lower heterogeneity.
   - ☐ no difference between the percentage of democracies among countries having lower heterogeneity and the percentage of democracies among countries with higher heterogeneity.

   B. World contains the variable democ.regime, which classifies each country as a democracy ("Yes" on democ.regime), or a dictatorship ("No" on democ.regime). This is the dependent variable. World also contains frac.eth3, which classifies countries according to their level of ethnic heterogeneity: "Low", "Medium", or "High". This is the independent variable. Run xtp, testing the ethnic heterogeneity hypothesis. Fill in the percentages of democracies:

	Ethnic Heterogeneity		
	Low	*Medium*	*High*
Percentage of democracies	?	?	?

5. Interaction.plot does not perform reliably when a numeric dependent variable has many missing values.

C. Based on these results, you could say that (Check one)

- ❑ as ethnic heterogeneity increases, the percentage of democracies increases.
- ❑ as ethnic heterogeneity increases, the percentage of democracies decreases.
- ❑ as ethnic heterogeneity increases, there is little change in the percentage of democracies.

D. A country's level of economic development also might be linked to its type of government. According to this perspective, countries with higher levels of economic development are more likely to be democracies than are countries with lower levels. The world dataset contains the variable gdp.cap2. Based on gross domestic product per capita, this variable is an indicator of economic development. Countries are classified as "Low" or "High." Obtain an xtp analysis of the democ.regime-frac.eth3 relationship, controlling for gdp.cap2. To accomplish this, you must first subset the world into two datasets, one for low-GDP countries (which you could name "world.low.gdp") and one for high-GDP countries ("world.high.gdp"). [*Hint:* The following expression creates world.low.gdp: world.low.gdp= subset(world, gdp.cap2== "Low").]

Fill in the percentages of democracies:

	Ethnic Heterogeneity		
	Low	*Medium*	*High*
Low GDP per capita percentage of democracies	?	?	?
High GDP per capita percentage of democracies	?	?	?

E. Consider the democ.regime-frac.eth3 relationship for low-GDP countries. Examine the difference between the percentage of democracies for "Low" heterogeneity and the percentage of democracies for "High" heterogeneity. This difference shows that the percentage of democracies among "Low" heterogeneity countries is (Fill in the blank) ________ percentage points lower or higher (circle one) than the percentage of democracies among "High" heterogeneity countries.

Now consider the democ.regime-frac.eth3 relationship for high-GDP countries. Examine the difference between the percentage of democracies for "Low" heterogeneity and the percentage of democracies for "High" heterogeneity. This difference shows that the percentage of democracies among "Low" heterogeneity countries is (Fill in the blank) ________ percentage points lower or higher (circle one) than the percentage of democracies among "High" heterogeneity countries.

F. Think about the set of relationships you just analyzed. How would you describe the relationship between ethnic heterogeneity and democracy, controlling for GDP per capita? (Circle one)

Spurious Additive Interaction

Explain your reasoning.

____

____

____

____

G. Obtain a multiple-line chart depicting the percentage of democracies for each value of frac.eth3, controlling for gdp.cap2. To accomplish this, follow these steps:

(i) Create an indicator variable, coded 1 for democracies and 0 for non-democracies. Name the indicator, "world$democ.yes."

(ii) Define a ddply object containing the means of world$democ.yes. [*Hint:* The following expression creates a ddply object containing "democ.mean," the means of world$democ.yes: obj1=ddply(world, .(gdp.cap2,frac.eth3), summarise, democ.mean = wtd.mean(democ.yes)).]

(iii) Run interaction.plot, graphing obj1$democ.mean across values of obj1$frac.eth3, separately for each value of obj1$gdp.cap2. Supply the main title: "Percentage of Democracies, by Ethnic Fractionalization and GDP Per Capita." In the interaction.plot syntax, suppress the default legend ("legend=F"). To specify appropriate plotting parameters, refer to this chapter's discussion.

(iv) Run the legend function, specifying an appropriate location (in this case, "topright"), and other legend parameters. Print the multiple-line chart.

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 4

Instructions: Review Chapter 6 of the Pollock (2014) book, *An R Companion to Political Analysis*. Answer questions 1 and 4 from Chapter 6. Include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Wednesday, March 25, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 4, which may be found on the class Blackboard page. No late work will be accepted.

the .05 level. It occurs by chance more frequently than 5 times out of 100. If the confidence interval does not include 0, then the mean difference is statistically significant at the .05 level. The difference occurs by chance fewer than 5 times out of 100.

In the gay.therm example, we have seen that males have a lower mean rating than do females. Does the confidence interval, the interval between −13.65 and −8.91, include 0? No, it doesn't. Conclusion: If the null hypothesis is correct, the probability of observing a mean difference of −11.28 is less than .05. Reject the null hypothesis. Just as before, wtd.t.test reports a t-statistic (t.value= −9.32) and an associated probability, or P-value: 0.0000. A more precise conclusion is that if the null hypothesis is correct, then random sampling error produces a mean difference of −11.28 virtually 0 percent of the time. The alternative hypothesis is on safe inferential ground. Reject the null hypothesis.

All right, so now we know that men score significantly lower on gay.therm than do women. But, as hypothesis 2 suggests, are men more internationally minded? Let's run the script and find out:

```
> wtd.t.test(nes.m$global.goals,nes.f$global.goals,
+  weight=nes.m$wt, weighty=nes.f$wt, samedata=F)
$test
[1] "Two Sample Weighted T-Test (Welch)"

$coefficients
     t.value           df      p.value
   1.3220021 1866.3582216    0.1863295

$additional
Difference      Mean.x      Mean.y    Std. Err
  1.526267   55.064567   53.538300    1.154512

> CI95(1.53,1.15)
[1] -0.7239586
[1] 3.783959
```

Again, the information on global.goals appears consistent with the hypothesis. Men, with a mean of 55.07, score higher than do women, who averaged 53.54, a mean difference of 1.53. Does this difference pass muster with the null hypothesis? Consider the 95 percent confidence interval: −0.72 at the low end and 3.78 at the high end. Does this interval include 0, the null's talisman? Yes, the confidence interval brackets 0. Thus, we know that the mean difference occurs by chance more frequently than 5 times out of 100. There is indeed a fairly high probability of observing a difference of 1.53, if in fact the null hypothesis is correct: .186. If the null hypothesis is correct, then we would observe by chance a sample difference of 1.53 about 19 times out of 100. Accept the null hypothesis.

EXERCISES

1. (Dataset: gss. Variables: egalit.scale, wtss.) The General Social Survey asks people a series of questions designed to measure how egalitarian they are—that is, the extent to which they think economic opportunities and rewards should be distributed more equally in society. The gss variable egalit.scale ranges from 0 (low egalitarianism) to 12 (high egalitarianism). In this exercise, you analyze egalitarianism using wtd.t.test and CI95. You then draw inferences about the population mean.

   A. Run wtd.t.test on gss$egalitarianism, with the test value set to 0. (Be sure to weight by gss$wtss). Egalitarianism has a sample mean of (Fill in the blank) ________________.

   B. Based on the results you obtained in Part A, run CI95. There is a probability of .95 that the true population mean falls between an egalitarianism score of (Fill in the blank) ______________ at the low end and a score of (Fill in the blank) ______________ at the high end.

   C. A student researcher hypothesizes that social work majors score significantly higher on the egalitarianism scale than the typical adult. The student researcher also hypothesizes that business majors score significantly lower on the egalitarianism scale than the average adult. After administering the scale to a number of social work majors and a group of business majors, the researcher obtains these results: social work majors' mean, 7.12; business majors' mean, 6.85. Run wtd.t.test, specifying a test value of 7.12. Based on the results, run CI95. Run wtd.t.test again, specifying a test value of 6.85. Run CI95.

Based on your analyses, and applying the .05 test of significance, you can infer that (Check one)

- ☐ Social work majors probably are not more egalitarian than most adults.
- ☐ Social work majors probably are more egalitarian than most adults.

Explain your answer.

____

____

____

____

D. Based on your analyses, and applying the .05 test of significance, you can infer that (Check one)

- ☐ Business majors probably are not less egalitarian than most adults.
- ☐ Business majors probably are less egalitarian than most adults.

Explain your answer.

____

____

____

____

E. Refer to the P-value you obtained from your analysis of the business majors' mean. (Fill in the blanks)

If in the population there is no difference between the mean of egalitarianism and the business majors' mean of 6.85, then the observed difference of ____________ occurs ____________ of the time by chance.

2. (Dataset: gss. Variables: int.info.scale, age, wtss.) Are older people interested in a wider variety of social, economic, political, and scientific issues than younger people? Or, do younger people and older people not differ significantly in the scope of their interests? In this exercise, you subset the gss dataset into two datasets: one composed of respondents who are younger than 30, and one composed of respondents who are 30 or older. You then enlist wtd.t.test and CI95 to test this hypothesis: In a comparison of individuals, people who are 30 or older are interested in a wider range of current issues than are people who are younger than 30.

   Gss contains int.info.scale, which measures respondents' level of interest in ten different issue areas. Scores on int.info.scale range from 0 to 20, with higher scores denoting higher levels of interest. This is the dependent variable. Use the independent variable, age, and the subset function to define two groups: a group of respondents who are younger than 30 ('age < 30') and a group who are 30 or older ('age >= 30'). Don't forget to include the gss weight variable, wtss, in the select argument. [*Hint:* The following expression defines the younger subset, named gss.Y: 'gss.Y=subset(gss, age < 30)'.] Run wtd.t.test to evaluate the difference between the mean for younger people and the mean for older people. In the wtd.t.test expression, type the younger mean first ("gss.Y$int.info.scale") and the older mean second. In the output, the younger mean will be Mean.x and the older mean will be Mean.y. Plug the mean difference and the standard error of the difference into CI95.

   A. Examine the results of your wtd.t.test and CI95 analyses. Fill in the table below.

Statistics for int.info.scale	*Level of interest in current issues*
Mean for younger group (<30 years old)	?
Mean for older group (≥=30 years old)	?
Mean difference	?
Lower 95 percent confidence boundary of mean difference	?
Upper 95 percent confidence boundary of mean difference	?
Does confidence interval contain 0? (Yes or No)	?

C. Focus your analysis on a comparison between respondents with low attendance ("Low") and respondents with high attendance ("High"). Use subset to create two datasets: one for "Low" attenders and one for "High" attenders. Run wtd.t.test and CI95 to find out if high attenders have significantly more children than do low attenders. Apply the .05 level of significance. (Fill in the blanks)

"Low" attendance mean: ______________

"High" attendance mean: ______________

Mean difference: ______________

95 percent confidence interval of the mean difference: Between ______________ and ______________.

D. Does the statistical evidence support the hypothesis that people who are more religious have significantly more children than do people who are less religious? (Check one)

- ☐ Yes, the statistical evidence supports the hypothesis.
- ☐ No, the statistical evidence does not support the hypothesis.

E. Explain your answer in Part D:

____

____

____

____

4. (Dataset: gss. Variables: sibs, relig, authoritarianism, sex, wtss.) Here are two bits of conventional wisdom, beliefs that are widely accepted as accurate descriptions of the world. Conventional wisdom 1: Catholics have bigger families than do Protestants. Conventional wisdom 2: Men have stronger authoritarian tendencies than do women. In this exercise, test these ideas and see how well they stand up to the statistical evidence. Test conventional wisdom 1 using subsetting by religious denomination (relig), and comparing the average number of siblings (gss variable sibs) for Protestants and Catholics. Test conventional wisdom 2 by subsetting by sex (sex) and comparing mean authoritarianism scale scores (authoritarianism) for males and females. (*Note:* For relig, the level name for Protestant is "PROTESTANT" and for Catholic is "CATHOLIC". For sex, the levels are "Male" and "Female".) The authoritarianism scale ranges from 0 (low authoritarianism) to 7 (high authoritarianism). Run the analyses. Record the results in the following table.

	Conventional Wisdom 1	*Conventional Wisdom 2*
Mean difference	?	?
Lower 95 percent confidence boundary of mean difference	?	?
Upper 95 percent confidence boundary of mean difference	?	?
Does confidence interval contain 0? (Yes or No)	?	?
t-statistic	?	?
P-value	?	?
Does statistical evidence support conventional wisdom? (Yes or No)	?	?

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 5

Instructions: Review Chapter 7 of the Pollock (2014) book, *An R Companion to Political Analysis*. Answer questions 2 and 4 from Chapter 7. Include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Wednesday, April 8, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 5, which may be found on the class Blackboard page. No late work will be accepted.

D. Using the states dataset, run xtp and xtp.chi2 analyses on the abortlaw3-cook.index3 relationship. Run xtp and xtp.chi2 on the gunlaw.rank3.rev-cook.index3 relationship. Using the statesD dataset, obtain Somers' d for both relationships. Browse the output. In the table below, enter the percentage of Democratic states, even states, and Republican states having more-restrictive policies. In the abortlaw3 row, for example, record the percentage of states having "9–10" abortion restrictions. For the gunlaw.rank3.rev row, enter the percentage of states in the "More restr" category. For each relationship, record chi-square, chi-square's P-value, and Somers' d.

Dependent Variable	*More Rep*	*Even*	*More Dem*	*Chi-square*	*P-value*	*Somers' d*
abortlaw3 % more restrictive	?	?	?	?	?	?
gunlaw.rank3.rev % more restrictive	?	?	?	?	?	?

E. Consider Somers' d for the gunlaw.rank3.rev-cook.index3 relationship. This value of Somers' d means that, compared to how well we can predict gunlaw.rank3.rev without knowing cook.index3 (Fill in the blank) ______________________.

F. Consider the chi-square P-value for the abortlaw3-cook.index3 relationship. This P-value means that, under the assumption that the null hypothesis is correct, (Fill in the blank)

______________________________________. Therefore, you should (Circle one)

reject   not reject   the null hypothesis.

G. Consider all the evidence from your analysis. The evidence suggests that pedantic pontificator is (Circle one)

correct.   incorrect.

Explain your reasoning.

____

____

____

____

2. (Dataset: gss. Variables: abany, femrole2, sex, wtss.) An interested student has joined pedantic pontificator in a discussion of the gender gap in U.S. politics.

Interested student: "On what sorts of issues or opinions are men and women most likely to be at odds? What defines the gender gap, anyway?"

Pedantic pontificator: "That's easy. A couple of points seem obvious, to me anyway. First, we know that the conflict over abortion rights is the defining gender issue of our time. Women are more likely than men to take a pro-choice position on this issue. Second—and pay close attention here—on more mundane cultural questions, such as whether women should be homemakers or pursue careers outside the home, men and women do not significantly differ."

A. Pedantic pontificator suggested the following two hypotheses about the gender gap: (Check two; Continued on next page)

- ☐ In a comparison of individuals, women are less likely than men to think that abortion should be allowed.
- ☐ In a comparison of individuals, women and men do not differ in their abortion opinions.
- ☐ In a comparison of individuals, women are more likely than men to think that abortion should be allowed.
- ☐ In a comparison of individuals, women are less likely than men to think that women should pursue careers outside the home.

- ☐ In a comparison of individuals, women and men do not differ in their opinions about female roles outside the home.
- ☐ In a comparison of individuals, women are more likely than men to think that women should pursue careers outside the home.

B. The gss dataset contains two variables that serve as dependent variables: abany, which records whether the respondent thinks that a woman should be able to obtain an abortion for any reason ("Yes" or "No"), and femrole2, which gauges respondents' opinions on the appropriate female role ("Home" or "Work"). The independent variable is sex. Perform xtp and xtp.chi2 analyses on the abany-sex and femrole2-sex relationships. (Be sure to include the gss weight variable, wtss.) Obtain Cramer's V for both relationships. In the abany-sex cross-tabulation, focus on the percentage saying "Yes." In the femrole2-sex cross-tabulation, focus on the "Work" category. Record your results in the table that follows.

Dependent Variable	*Male*	*Female*	*Chi-square*	*P-value*	*Cramer's V*
Percent "Yes" (abany)	?	?	?	?	?
Percent "Work" (femrole2)	?	?	?	?	?

C. Based on these results, you conclude that (Check three)

- ☐ a statistically significant gender gap exists on abortion opinions.
- ☐ pedantic pontificator's hypothesis about the femrole2-sex relationship is not supported by the analysis.
- ☐ under the assumption that the null hypothesis is correct, the abany-sex relationship occurs by chance more frequently than 5 times out of 100.
- ☐ pedantic pontificator's hypothesis about the abany-sex relationship is supported by the analysis.
- ☐ a higher percentage of females than males think that women belong in the workplace.

D. Under the assumption that the null hypothesis is correct, the P-value of the chi-square statistic in the femrole2-sex cross-tabulation tells you that (Fill in the blank) ______________________________

____.

3. (Datasets: gss/gssD. Variables: muslim.tol, relig.t013, educ2, wtss.) Think for a moment about the general principle of religious tolerance—the willingness of individuals to tolerate diversity in religious faiths and practices. A person who has a high level of religious tolerance agrees that "all religious groups in the United States should have equal rights" and that "we must respect all religions." A person with low religious tolerance tends to disagree with these ideas. Gss contains relig.t013, which measures respondents' level of agreement with these general principles. Relig.t013 is coded "Low", "Mid", and "High". Now recall muslim.tol, which you created in one of the exercises in Chapter 3. Muslim.tol does not measure general religious tolerance; rather, it measures respondents' level of tolerance toward "anti-American Muslim clergymen." Muslim.tol is coded "Low", "Mid", and "High".

Do people apply general principles of religious tolerance to instances involving radical Muslims? As the literature on political tolerance suggests that individuals' willingness or ability to apply general democratic principles to unpopular groups depends on their level of education, perhaps we see the same phenomenon in the study of religious tolerance. Consider three hypothetical expectations.

Expectation 1: When dealing with low-level education individuals, those with higher levels of general religious tolerance are significantly more likely to express high levels of tolerance toward anti-American Muslim clergymen than those with lower levels of general religious tolerance.

Expectation 2: For high-level education individuals, those with higher levels of general religious tolerance are significantly more likely to express high levels of tolerance toward anti-American Muslim clergymen than those with lower levels of general religious tolerance.

4. Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press. Inglehart writes extensively about cultural change in postindustrial societies.

Expectation 3: The relationship between religious tolerance and tolerance toward Muslim clergy is stronger for high-education individuals than for low-education individuals.

A. To test expectations 1 to 3, subset gss by educ2, and run xtp and xtp.chi2 on each subset, specifying muslim.tol as the dependent variable, relig.t013 as the independent variable, and wtss as the weight variable. Subset gssD on educ2, and run somersD on each subset. Record the results of your analysis in the table that follows. For respondents at both education levels, write down the percentage with "High" tolerance toward Muslim clergy. Record chi-square and its P-value, and Somers' d.

	Religious Tolerance (relig.t013)					
	Low	*Mid*	*High*	*Chi-square*	*P-value*	*Somers' d*
Low-level education respondents: Percent "High" tolerance toward Muslim clergy	?	?	?	?	?	?
High-level education respondents: Percent "High" tolerance toward Muslim clergy	?	?	?	?	?	?

B. Does your analysis support expectation 1? Answer yes or no and explain.

C. Does your analysis support expectation 2? Answer yes or no and explain.

D. Does your analysis support expectation 3? Answer yes or no and explain.

4. (Datasets: world/worldD. Variables: protact3, gender.equa13, vi.re13, pmat12.3.) Ronald Inglehart offers a particularly elegant and compelling idea about the future of economically advanced societies. According to Inglehart, the cultures of many postindustrial societies have been going through a value shift—the waning importance of materialist values and a growing pursuit of postmaterialist values. In postmaterialist societies, economically based conflicts—unions versus big business, rich versus poor—are increasingly supplanted by an emphasis on self-expression and social equality. Postmaterialist societies also are marked by rising secularism and elite-challenging behaviors, such as boycotts and demonstrations. In this exercise, you investigate Inglehart's theory.[4]

The world variable pmat12.3 measures the level of postmaterial values by a three-category ordinal measure: low postmaterialism ("Low post-mat"), moderate postmaterialism ("Moderate post-mat"), and high postmaterialism ("High post-mat"). Use pmat12.3 as the independent variable. Here are three dependent variables, all of which are three-category ordinals: gender.equal3, which captures gender equality ("Low", "Medium", "High"); protact3, which measures citizen participation in protests ("Low", "Moderate", "High"); and vi.rel3, which gauges religiosity by the percentage of the public saying that religion is very important in their lives ("<20%", "20-50%", or ">50%"). You can see that higher codes on the dependent variables denote greater gender equality (gender.equal3), more protest activity (protact3), and higher levels of religiosity (vi.rel3).

A. Using pmat12.3 as the independent variable, three post-materialist hypotheses can be framed:

Gender equality hypothesis (Fill in the blanks):

In a comparison of countries, those with higher levels of postmaterialism have ____________________ levels of gender equality than countries with lower levels of postmaterialism.

Protest activity hypothesis (Fill in the blanks):

In a comparison of countries, those with ____________________ levels of postmaterialism have ____________________ levels of protest activity than countries with ____________________ levels of postmaterialism.

Religiosity hypothesis (Fill in the blank):

In a comparison of countries, those with____
______________________.

B. Consider how the independent variable is coded and how each dependent variable is coded. In the way that R calculates Somers' d, which one of the three hypotheses implies a negative sign on the measure of association? (Check one)

- ❑ The gender equality hypothesis
- ❑ The protest activity hypothesis
- ❑ The religiosity hypothesis

C. Test the hypotheses by performing the appropriate analyses on world and worldD. In the table that follows, record the percentages of countries falling into the highest level of each dependent variable. Also, report chi-square statistics, P-values, and Somers' d statistics.

	Level of Postmaterialism					
Dependent Variable	*Low*	*Moderate*	*High*	*Chi-square*	*P-value*	*Somers' d*
Percentage high gender equality	?	?	?	?	?	?
Percentage high protest activity	?	?	?	?	?	?
Percentage high religiosity	?	?	?	?	?	?

D. Which of the following inferences are supported by your analysis? (Check all that apply)

- ❑ The gender equality hypothesis is supported.
- ❑ Compared with how well we can predict gender equality by not knowing the level of postmaterialism, we can improve our prediction by 20.03 percent by knowing the level of postmaterialism.
- ❑ The protest activity hypothesis is supported.
- ❑ If the null hypothesis is correct, the postmaterialism-protest activity relationship occurs, by chance, less frequently than 5 times out of 100.
- ❑ The religiosity hypothesis is supported.
- ❑ If the null hypothesis is correct, the postmaterialism-religiosity relationship occurs, by chance, less frequently than 5 times out of 100.

That concludes the exercises for this chapter. Before exiting R, be sure to save the workspace.

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 6

Instructions: Review Chapter 8 of the Pollock (2014) book, *An R Companion to Political Analysis*. Answer questions 4 and 5 from Chapter 8. Include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Wednesday, April 22, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 6, which may be found on the class Blackboard page. No late work will be accepted.

3. (Dataset: states. Variables: to.0408, obama08). An article of faith among Democratic Party strategists (and a source of apprehension among Republican strategists) is that high voter turnouts help Democratic candidates. Why is this the case? According to conventional wisdom, Democratic electorates are less likely to vote than are Republican voters. Thus, low turnouts naturally favor Republican candidates. The reasoning is that as turnouts increase, a larger number of potential Democratic voters go to the polls, creating a better opportunity for Democratic candidates. Therefore, as turnouts go up, so should the Democratic percentage of the vote.[11]

   A. Use lm to test the conventional wisdom. The states dataset contains to.0408, the percentage-point change in presidential election turnout between 2004 and 2008. States in which turnout declined between 2004 and 2008 have negative values on to.0408, while states in which turnout increased have positive values on to.0408. (For example, North Carolina's turnout increased from 57.8 percent to 65.8 percent, giving North Carolina a value of 8.0 on to.0408. Utah's turnout dropped from 58.9 percent to 53.3 percent, giving it a value of −5.6 on to.0408.) The independent variable is to.0408. Another variable, obama08, the percentage of the vote cast for Democratic candidate Barack Obama, is the dependent variable.

   Based on your results, the regression equation for estimating the percentage votes cast for Obama is (Fill in the blanks)

   __________ + ____________ * to.0408.

   (constant) (regression coefficient)

   B. The P-value for the regression coefficient on to.0408 is __________, and adjusted R-square is __________.

   C. Consider your findings in Parts A and B. Your conclusion is that (Fill in the appropriate blank)

   the conventional wisdom is correct because____

   ____.

   the conventional wisdom is incorrect because____

   ____.

4. (Dataset: States. Variables: abortlaw, permit, womleg.2010.) As you are no doubt aware, in its momentous decision in *Roe v. Wade* (1973), the U.S. Supreme Court declared that states may not outlaw abortion. Even so, many state legislatures have enacted restrictions and regulations, which while not banning abortion, make an abortion more difficult to obtain. However, other states have few or no restrictions. What factors might explain these differences in abortion laws among the states? We know that the mass public remains divided on this issue. Public opinion in some states is more favorable toward permitting abortion, and in other states is less favorable. Does public opinion guide state policy on this issue? What about the direct role of state lawmakers? Are state legislatures with fewer female legislators more likely to enact restrictive abortion laws than are legislatures with more female legislators?

   States contains abortlaw, which measures the number of abortion restrictions a state has enacted into law. Values on abortlaw range from 0 (least restrictive) to 10 (most restrictive). This is the dependent variable. States also has the variable permit, the percentage of the mass public saying that abortion should "always" be permitted. And states has womleg.2010, the percentage of female legislators. First, run a bivariate regression with abortlaw and permit. Then, run a multiple regression that includes permit and womleg.2010.

   A. Think about how abortlaw and permit are coded. If you use regression analysis to test the idea that public opinion on abortion affects state abortion policy, then you expect to find (Check one)

   ☐ a negative sign on permit's regression coefficient.

   ☐ a positive sign on permit's regression coefficient.

   B. Analyze the abortlaw-permit relationship using lm. According to the results, the regression equation for estimating the number of abortion restrictions is (Fill in the blanks)

   __________ ____________________ *permit.

   (constant) (regression coefficient)

11. Martinez, M. D., & Gill, J. (2005, November). The effects of turnout on partisan outcomes in U.S. presidential elections 1960–2000. *Journal of Politics, 67*(4), 1248–1274. Martinez and Gill find that the Democratic advantage from higher turnouts has declined over time.

C. Following this chapter's discussion of the scatterplot function, create a scatterplot of the abortlaw-permit relationship. Supply an informative main title and descriptive axis labels. Print the scatterplot.

D. Making certain that the scatterplot you created in Part C is still in the active graphics window, use the symbols function to overlay a bubble plot on the scatterplot. Specify that the plot's circles display womleg.2010, the percentage of women in the legislature. Print the bubble plot.

E. Examine the bubble plot you created in Part D. Based on your examination, does the "bubbled" variable, womleg.2010, appear to be related to the dependent variable, number of abortion restrictions? Answer yes or no and briefly explain your answer.

____

____

____

F. Run a multiple regression analysis of abortlaw, specifying permit and womleg.2010 as independent variables. Based on your results, the multiple regression for estimating the number of abortion restrictions is (Fill in the blanks)

14.21 + ______________ *permit + ______________ *womleg.2010.

G. The P-value for the regression coefficient on permit is (Fill in the blank) ______________, and the P-value for the regression coefficient on womleg.2010 is (Fill in the blank) ______________.

H. After controlling for womleg.2010, suppose someone claimed that the relationship between abortlaw and permit turns out to be spurious. Does your analysis support this claim? Answer yes or no and briefly explain your answer.

____

____

____

5. (Dataset: states. Variables: demstate09, dempct.m, libpct.m.) In this exercise, you use correlation and multiple regression to examine a set of relationships between mass attitudes and the partisan make-up of state legislatures. State legislatures are remarkably varied in this regard—ranging in partisan composition from about 24 percent Democratic to over 90 percent Democratic. What accounts for this variation? Consider two plausible independent variables: the percentage of a state's citizens who are self-identified Democrats and the percentage of citizens who are self-described liberals. Each of these variables should have a positive relationship with the percentage of Democrats in the state legislature.

A. States contains these three variables: demstate09, the percentage of state legislators who are Democrats; dempct.m, the percentage of Democrats in the mass electorate; and libpct.m, the percentage of self-described liberals in the mass public. Run wtd.cor to find the correlation coefficients among demstate09, dempct.m, and libpct.m. Fill in the six empty cells of this correlation matrix:

	Percent Mass Public Democratic	*Percent State Mass Public Democratic*	*Percent Legislators Liberal*
Percent state legislators Democratic correlation	1	?	?
Percent mass public Democratic correlation	?	1	?
Percent mass public liberal correlation	?	?	1

B. According to the correlation coefficient, as the percentage of liberals in the mass public increases, the percentage of Democratic state legislators (Circle one)

increases. decreases.

C. Suppose someone were to make this claim: "Being a Democrat and being a liberal are practically synonymous. The relationship between the percentage of Democratic identifiers and the percentage of liberals, therefore, will be positive and strong." According to the correlation coefficient, this claim is (Fill in the appropriate blank)

correct because ______________________________.

incorrect because ______________________________.

D. Perform a multiple regression analysis to obtain estimates for the partial effects of dempct.m and libpct.m on demstate09. Demstate09 is the dependent variable, and dempct.m and libpct.m are the independent variables. Based on your results, the multiple regression for estimating the percentage of Democratic state legislators is (Fill in the blanks)

−28.375 + ______________ *dempct.m + ______________ *libpct.m.

E. The P-value for the regression coefficient on dempct.m is (Fill in the blank) ______________, and the P-value for the regression coefficient on libpct.m is (Fill in the blank) ______________.

F. As you may know, Nebraska's state legislature is unique in two ways: It is unicameral (all other state legislatures are bicameral), and it is nonpartisan. Candidates do not run for the state legislature using party labels, and the legislature is not organized on the basis of party. Thus, Nebraska has a missing value on the variable demstate09, and it was not included in the regression analysis that you just performed. However, if you peruse states, you find that 29.03 percent of Nebraskans are Democrats and 16.44 percent are self-described liberals.

For the sake of speculation, assume that Nebraska decided that all members of the state legislature should declare a partisan allegiance. Based on your regression model, about what percentage of state legislators would be Democrats? (Circle one)

About 30 percent About 40 percent About 50 percent

G. Based on your interpretation of the multiple regression output, you can conclude that (Check all that apply)

- ☐ controlling for the percentage of the mass public who are Democratic, a 1-percentage-point increase in the percentage of liberals in the mass public is associated with about a 2.6-percentage-point increase in the percentage of Democratic state legislators.
- ☐ both independent variables are significantly related to the dependent variable.
- ☐ taken together, both independent variables explain over one-half of the variation in the dependent variable.

6. (Dataset: gssD. Variables: fem.role, authoritarianism, age).

Two scholars are discussing why some people are more likely than others to accept women in nontraditional roles.

Scholar 1: "The main thing to know about people is how old they are. Older people were socialized in a more traditional time and are not as willing to accept women in roles outside the home. By contrast, younger people are being socialized during an era of greater female empowerment outside the home. I'll bet that if you perform a regression analysis using a variable measuring the level of acceptance of women outside the home as the dependent variable, and use age as the independent variable, the regression coefficient on age is negative and statistically significant."

Scholar 2: "You are on the right track, but you're not really addressing the question of *why* age and gender-role attitudes are negatively related. More specifically, the 'more traditional time' you refer to was a time of greater authoritarianism. Authoritarianism is significantly related to gender-role attitudes: As authoritarianism goes up, support for nontraditional roles goes down. Thus, the older generation's objection to nontraditional gender roles arises from their greater authoritarianism, not simply their age. Go ahead and run your simple bivariate regression using age as the independent variable. Then, run a multiple regression, using age *and* authoritarianism as independent variables. The multiple regression will show that authoritarianism is strongly related to gender-role attitudes. But, the coefficient on age will be statistically insignificant."

The design dataset, gssD, contains fem.role, a metric that runs from 0 (women belong in the home) to 9 (women belong outside the home). (You created gssD in Chapter 7. If you need to re-create gssD, review Chapter 7's discussion of the svydesign function.) So, higher values of fem.role indicate greater acceptance of

Arizona State University
School of Politics and Global Studies
POS 401: Political Statistics
Spring 2015

Assignment 7

Instructions: Review Chapter 9-10 of the Pollock (2014) book, *An R Companion to Political Analysis*. Answer questions 5 from Chapter 9 and question 1 from Chapter 10. Include all tables and figures in your word processing document (not as separate files), and please turn in your R code as well. You may send your R script file, or you may copy and paste your R code into your word processing document at the end of your write-up. Please do not copy and paste the printout from the R console.

Your assignment is due on Friday, May 1, 2015 at 6:00PM. All assignments must be turned in via the digital dropbox for Assignment 7, which may be found on the class Blackboard page. No late work will be accepted.

4. (Dataset: gssD. Variables: abortion.scale, dogmatism, educ2). One of the examples in this chapter discussed the polarization perspective—the idea that political conflict is more pronounced among people who are more knowledgeable about politics than it is among less knowledgeable people. Perhaps the same pattern applies to the relationship between religious dogmatism and abortion opinions. Namely, that dogmatism has a strong effect on abortion attitudes among politically knowledgeable people, but this effect is weaker for people who have lower knowledge about politics. We can use the education variable (educ2) as a surrogate for political knowledge, because we can reasonably assume that people with more education are more politically knowledgeable than less educated people. In this exercise, you rerun the svyglm model from Exercise 3, only this time include the educ2*dogmatism interaction.

   A. Run the regression, using abortion.scale as the dependent variable, and dogmatism, educ2, and educ2*dogmatism as the independent variables. According to the Coefficients table, the multiple regression equation for estimating scores on the abortion scale is as follows (Fill in the blanks, putting the constant in the first blank):

   __________ + __________ *dogmatism + __________ *educ2 + __________ * educ2*dogmatism.

   B. Use the regression coefficients to estimate scores on the abortion scale for respondents with different values on the dogmatism scale, and the education variable. Calculate the estimates for low-education ("0–12 yrs") and high-education ("13 plus yrs") people with very low dogmatism (a score of 0 on the dogmatism scale), moderate dogmatism (a score of 5), and very high dogmatism (a score of 12). Record the abortion.scale estimates next to the question marks in the table that follows.

	Level of Dogmatism		
Level of Education (educ2)	*Very Low Dogmatism (dogmatism = 0)*	*Moderate Dogmatism (dogmatism = 5)*	*Very High Dogmatism (dogmatism = 12)*
Lower education (0–12 yrs)	?	?	?
Higher education (13 plus yrs)	?	?	?

   C. Examine the table of estimates. Think about the polarization perspective. Does the analysis support the idea that religious dogmatism has a larger effect on abortion opinions among people with more education than among people with less education? (Circle one)

   Yes No

   Explain your reasoning.

   ____

   ____

   ____

   ____

5. (Dataset: gssD. Variables: polviews.n, race2, homosex2.) If one is trying to predict ideological self-identification on the basis of opinions on social issues, such as homosexuality, one expects most African Americans to be conservatives. Indeed, blacks are considerably more likely to oppose homosexuality than are whites. For example, according to the gss data, over 70 percent of blacks say that homosexuality is "always wrong," compared with 50 percent of whites. Yet, only about 20 percent of blacks call themselves "conservative," compared with 40 percent of whites. Why? A commonly accepted view is that social issues lack salience for African Americans. Issues like homosexuality may matter for whites—whites who object to homosexuality are more likely to be self-described conservatives than are whites who do not object—but they have no effect for blacks. According to this argument, blacks who think homosexuality is wrong are no more likely to call themselves conservatives than are blacks who do not think homosexuality is wrong. Or, so the familiar argument goes. Is this idea correct? More research needs to be done on this question.[5]

---

5. Kidd, Q., Diggs, H., Farooq, M., & Murray, M. (2007, March). Black voters, black candidates, and social issues: Does party identification matter? *Social Science Quarterly, 88,* 165–176.

You can model salience with interaction. Consider the 7-point ideological scale (polviews.n) as a dependent variable, ranging from "extremely liberal" at 1 to "extremely conservative" at 7. Now bring in two independent variables: race2, with levels "White" and "Black", and homosex2, a factor gauging opposition to homosexuality, with levels "Not always wrong" and "Always wrong". So, for race2, the first-named level is "White", and for homosex2 the first-named level is "Not always wrong". Finally, think about an interaction term, race2*homosex2. Examine the regression model that follows:

$$\text{polviews.n} = a + b_1\text{*race2} + b_2\text{*homosex2} + b_3\text{* race2*homosex2}.$$

A. The interaction term, race2*homosex2, takes on a value of 1 for (Check one)

- ☐ blacks who think that homosexuality is "always wrong."
- ☐ blacks who think that homosexuality is "not always wrong."
- ☐ all respondents.

B. To gauge the effect of homosex2 among whites, you need to compare values of polviews.n for "not always wrong" whites and "always wrong" whites.

Which of the following estimates polviews.n for "not always wrong" whites? (Circle one)

a $\qquad$ $a + b_1$ $\qquad$ $a + b_2$

Which of the following estimates polviews.n for "always wrong" whites? (Circle one)

a $\qquad$ $a + b_1$ $\qquad$ $a + b_2$

C. Remember that higher scores on polviews.n denote stronger conservative self-identifications.

If the salience argument is correct—the idea that heightened opposition to homosexuality leads to stronger conservative ideological leanings among whites, but not blacks—then the sign on the coefficient, b_2, is (Circle one)

negative. $\qquad$ positive. $\qquad$ zero.

If the salience argument is correct, then the sign on the coefficient, b_3, is (Circle one)

negative. $\qquad$ positive. $\qquad$ zero.

D. Use svyglm to run the regression. The regression equation for estimating polviews.n is as follows (Fill in the blanks, putting the constant in the first blank):

polviews.n = ________________ + ________________ *race2 + ________________ *homosex2 + ________________ * race2*homosex2.

E. Which of the variables in the model have statistically significant effects on polviews.n? (Check all that apply.)

- ☐ race2
- ☐ homosex2
- ☐ race2*homosex2

F. Use the model to estimate polviews.n for "not always wrong" whites and "always wrong" whites. For "not always wrong" whites you obtain ________________, and for "always wrong" whites you obtain ________________.

G. Use the model to estimate polviews.n for "not always wrong" blacks and "always wrong" blacks. For "not always wrong" blacks you obtain ________________, and for "always wrong" blacks you obtain ________________.

H. Consider all the evidence you have adduced. Based on the evidence, the salience idea appears to be (Circle one)

correct. $\qquad$ incorrect.

Explain your answer.

____

____

____

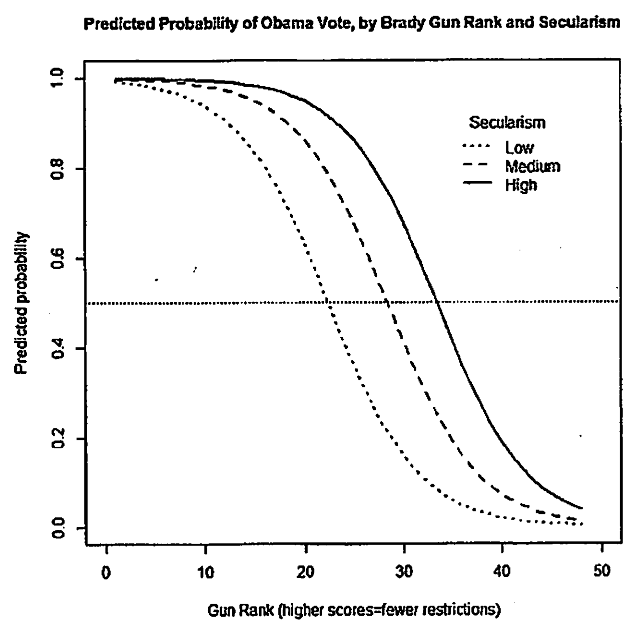
```
matplot(data2$gunlaw.rank, data2[, c("secLow","secMed","secHigh")], type="l", lty=c(3,2,1),
  ylab="Predicted probability",
  main=list("Predicted Probability of Obama Vote, by Brady Gun Rank and Secularism",cex=1),
  xlab="Gun Rank (higher scores=fewer restrictions)", lwd=2, col="black" ,
  xlim=c(0,50))

abline(h=.5,lty=3,lwd=.5)

legend("topright",
  legend=c("Low","Medium","High"),
  lty=c(3,2,1), col="black",
  lwd=2,
  title="Secularism",
  inset=0.1,
  bty="n")
```

Another first-rate R graphic appears (Figure 10.5).

Figure 10.5 Multiple-Line Chart of Three Logistic Regression Curves



EXERCISES

1. (Dataset: world. Variables: democ.regime, frac.eth, gdp.10.thou.) In Chapter 5, you tested this hypothesis: In a comparison of countries, those having lower levels of ethnic heterogeneity are more likely to be democracies than those having higher levels of ethnic heterogeneity. This hypothesis states as heterogeneity goes up, the probability of democracy goes down. Then, you reran the analysis controlling for a measure of countries' economic development, per capita gross domestic product (GDP). For this independent variable, the relationship is thought to be positive: As economic development increases, so does the likelihood that a country is democratic. In the current exercise, you reexamine this set of relationships using interval-level independent variables and a more powerful method of analysis, logistic regression.

   World contains these three variables: democ.regime, frac.eth, and gdp.10.thou. Democ.regime is coded 1 if the country is a democracy and coded 0 if it is not a democracy. This is the dependent variable. One of the independent variables, frac.eth, can vary between 0 (denoting low fractionalization) and 1 (high fractionalization). The other independent variable, gdp.10.thou, measures per-capita GDP in units of $10,000.

   A. Create a new model, modelD, by running glm and using frac.eth and gdp.10.thou to predict democ.regime. Run summary, orci, and logregR2 on modelD. The table on the following page contains nine question marks. Fill in the correct value next to each question mark.

Model Estimates	*Coefficient*	*P-value*	*Odds Ratio*
Constant	.842		
frac.eth	−1.592	?	?
gdp.10.thou	.713	?	?
logregR2 statistics	Statistic	Significance	
Chi-square	?	?	
Cox-Snell index	?		
Nagelkerke index	?		
McFadden's R2	?		

B. Use each odds ratio to calculate a percentage change in the odds. Controlling for gdp.10.thou, a one-unit change in frac.eth, from low heterogeneity to high heterogeneity, (Check one)

- ☐ increases the odds of democracy by about 20 percent.
- ☐ decreases the odds of democracy by about 20 percent.
- ☐ decreases the odds of democracy by about 80 percent.

Controlling for frac.eth, each $10,000 increase in per-capita GDP (Check one)

- ☐ increases the odds of democracy by about 104 percent.
- ☐ increases the odds of democracy by about 204 percent.
- ☐ increases the odds of democracy by about 40 percent.

To respond to Parts C, D, E, and F, you will need to perform these preliminary steps: (i) Use summary or describe to obtain descriptive information about the two independent variables. (ii) Set up a data frame, 'data3,' by using the observed range of frac.eth, from 0 to .93. Specify increments of .01. (If you are not sure how to do this, see Box 10.2.) (iii) Retrieve modelD's coefficients. Use the coefficients to calculate three variables containing the predicted probabilities of democracy at three values of per-capita GDP: low ('data3$LowGDP'), medium ('data3$MidGDP'), and high ('data3$HighGDP'). Use gpd.10.thou's lower quartile boundary as the low value, its median as the medium value, and its upper quartile boundary as the high value.

C. As an empirical matter, the least fractionalized country in the world dataset has a value of 0 on frac.eth, and the most fractionalized country has a value of .93 on frac.eth. According to data3, the predicted probability of democracy for a country with low fractionalization (frac.eth = 0) at the median value of gdp.10.thou (MidGDP) is equal to (Fill in the blank) _____________. The predicted probability of democracy for a highly fractionalized country (frac.eth = .93) at the median value of gdp.10.thou is equal to (Fill in the blank) _______________.

D. Consider this statement: At all levels of fractionalization, richer countries are more likely to be democracies than are poorer countries. Do your findings support this statement? Answer yes or no and explain.

____

____

____

E. Consider this statement: At the median value of fractionalization, all countries—low GDP, middle GDP, and high GDP—are probably democracies. Do your findings support this statement? Answer yes or no and explain.

____

____

____

F. Consider this statement: At the level of fractionalization where low-GDP countries switch over from more likely to be democratic to less likely to be democratic, there is still a probability of about .6 that high-GDP countries are democracies. Do your findings support this statement? Answer yes or no and explain.

___

___

___

2. (Datasets: nes/nesD. Variables: voted, educ.r, age.) In this chapter's guided examples, you used gss/gssD to analyze the relationship between education, age, and turnout in the 2004 election. (The 2008 General Social Survey, on which gss/gssD are based, concluded its last interview in September, so interviewers could not ask about voting behavior in the 2008 election.) Recall that both independent variables, education and age, are significantly related to the probability of voting. Recall also the markedly different effects of education for younger voters and older voters. In this exercise, use the nes/nesD datasets to reexamine this set of relationships and see whether the same pattern of effects occurred in 2008. The nes sets contain voted (coded 1 for voters and 0 for nonvoters), educ.r (number of years of schooling), and age (respondent age in years).

A. Run a svyglm analysis on nesD to estimate a new model, 'modelE.' Write the correct values for modelE next to the question marks in the following table:

Model Estimates	*Coefficient*	*P-value*	*Odds Ratio*
Constant	?		
educ.r	?	?	?
age	?	?	?
logregR2 statistics	Statistic	Significance	
Chi-square	?	?	
Cox-Snell index	?		
Nagelkerke index	?		
McFadden's R2	?		

In order to complete Parts B and C, you need to perform the following steps: (i) Run describe on nes$educ.r to obtain its range. (ii) Define a data frame, 'data4,' containing a row for each value of edu.r. (iii) Retrieve modelE's coefficients. Use the coefficients to calculate the effect of education on the probability of voting for voters who are 26 years old and those who are 65 years old. As in this chapter's guided example, you need to create two expressions, one calculating the predicted probability for 26 year-olds ('data4$pred26') and one for 65-year-olds ('data4$pred65').

B. In this chapter's guided examples, your gss/gssD analyses showed that for older voters, the probabilities switched in favor of voting (from less than .5 to more than .5) between 7 years and 8 years of education. For younger voters, the switchover occurred at a substantially higher level of education—between 14 years and 15 years. In what ways are your nes/nesD findings similar to the gss/gssD findings? In what ways are the two sets of findings different? Write a short paragraph describing the similarities and differences.

___

___

___

___

___

C. Use matplot to obtain a graph of the predicted probability of voting (y-axis) at each value of educ.r, for 26-year-olds and 65-year-olds. Use this chapter's example for guidance. Be sure to add a y-axis reference line and a legend. Print the graph.