POLITICAL SCIENCE 30 Summer 2017 (Draft 7.3.17)
POLITICAL INQUIRY
Lectures   Tuesday (Thursday of Week 1)   WLH 2111 2-4:50
Lab Thursday (starting Week 2) ERC 117
Final Assignment due by Saturday, August 5, 3 PM

Note: minor changes may be made to this syllabus throughout the quarter (especially the first two weeks). You are responsible for any changes mentioned in class, even if you are absent. Changes will also be posted.

Professor Peter Galderisi
Office: 449 Social Sciences Building
Office Hours: Wednesday 12-3

Graders:
 Huchen Liu
 Hernan Picatto

INTRODUCTION

This class will introduce you to the fundamentals of political (actually any) research. We will first deal with the aspects of scientific research--how one can ask questions that can be logically answered--and the use of the 'scientific method' in the study of political events. We will review several common methods of acquiring the necessary data for our answers, discuss the benefits of and the problems with each, and review the basics of standard research design. Throughout the quarter, you will be introduced to the use of elementary statistics as tools sometimes necessary for the analysis of political data.   As we cover statistics, you will apply what you have learned by carrying out analyses, using SPSS software, of a data set or sets that I will provide (more on this later in class—I'm currently working on creating another data set that covers the 2016 election).

REQUIREMENTS:

I. READINGS:

A. Two 'books" (I'll explain this in class) are available for purchase through the campus bookstore:

 Galderisi *Understanding Political Science Statistics: Of Observations and Expectations*
 Galderisi *SPSS Manual to Accompany UPSS*

You will be able to download the SPSS data files later this quarter from the course page. Full instructions will be given.

B. Several readings may also be required and are listed later in this syllabus. They can be found through the course page. Again, full instructions will be given.

We **STRONGLY** suggest reading the material both **before and after** the appropriate class or section.
We **STRONGLY** suggest going through the sample questions at the end of each chapter. Answers to all odd number questions can be found at the end of the text.

C. Several videos, mainly dealing with using SPSS, will be posted. They can be opened by clicking onto the provided link.

II. SOFTWARE

You may use SPSS in an on-campus computer laboratory or you may acquire it for a compatible computer of your own using a campus license. Instructions on acquiring that license will be given.

I am also in the process of hiring an undergraduate peer adviser who will set up shop in one of the computer labs or my office to assist you. I'll let you know when this is done.

III. ATTENDANCE AND EFFORT:

A. **Attendance is MANDATORY!**
Class lectures will proceed in a logical, progressive fashion (much more so than in any other class). One missed lecture (there are only 4 more after the first introductory day) can lead to a total lack of comprehension over the next series of classes. Lab sections will be conducted in a manner intended to maximize your understanding and complete your class project.

B. **Hard work and perseverance.**
Remember, research methodology and especially statistics are like a foreign language. Without constant exposure, practice, and repetition, languages are hard to master, especially in five (or fewer) weeks. The same is true here. Conclusion? Expect to work hard at the beginning of this class, harder in the middle, and harder still at the end. Payoff? You may actually understand this stuff, and we may all maintain what little sanity we have left (well, that ship has sailed for me, but as the immortal albeit still living Jimmy Buffett wrote: "If we all weren't crazy, we'd all go insane").

C. Weekly Preparation guides will be posted covering the major points you need to learn. If you follow them you will get more out of this class.

D. You will learn the most from this class if you actively participate in lecture and, especially, lab sections (again, think about learning a foreign language). Once we start covering statistics, you will need some way to do basic calculations. Nothing fancy here— a basic, $4 calculator or the calculator function on your phone should do.

Speaking of phones: The only reason why you need to ever look at them is to use the calculator function or, for some of you, take notes. Anyone using the phone for personal (other than emergency) reasons will be asked to leave. It distracts me, other students, and is becoming increasingly and annoyingly prevalent.

IV. ASSIGNMENTS: To be discussed in class

A NOTE ON GRADING:

Any request for a grade review must be made to your grader in writing (typed) with a full explanation of why you are requesting the review. Note that any review may result in a higher or lower grade (or no change).

FINAL GRADE DISTRIBUTION:

Understanding the difficulty of this class for many, the grade "curve," especially at the low, passing end, is rather generous. A few A+ grades will be also be given to the top students in class and sections.

Also—do NOT wait until the last minute to start your work. Just like the man (woman, your view of a divine order) upstairs, we can only help those who help themselves.

INCOMPLETES
The university grants us precious little discretion here. To qualify for an incomplete I must demonstrate that you have been doing passable work (which may be difficult if we only have a final assignment) and you must demonstrate a reason for requesting an incomplete that conforms to university guidelines (documented illness, death or emergency in the family, unexpected military deployment, etc.). Again, the university makes this decision—not me nor the graders.
"Going on an early vacation" is not a university sanctioned excuse.

WEB PAGE:

TritonEd (formerly TED, formerly Blackboard) is becoming less useful for me. I am in the process of setting up a Google Drive web page for this class.

A NOTE ON THE USE OF STATISTICS:

I will not attempt in this class to indoctrinate you into believing that only statistically-based research is valid research. Obviously, such an undertaking would be methodologically ludicrous. One begins one's research by asking theoretically important questions. Sometimes, <u>and only sometimes</u>, statistics can help us to answer those questions. Statistics are merely a summary tool. They allow us to test whether what we expect is confirmed by what we observe (thus, the subtitle of the text). They help us with our research, but they are not the driving force behind it. Learning statistics yields some valuable results. First, you will have a greater choice of research questions to ask. You will no longer need to shy away from at least some questions that require statistically-based answers. Second, you will be better able to evaluate others' scholarly research. We sometimes tend to accept others' statistical findings as gospel, or reject them as trivial when, in fact, we make no attempt to try to understand what the researcher tried to accomplish (hear that, politicians). Ignorance may be bliss, it may even help in politics, but it is not academically virtuous. Third, you will acquire the foundation needed to do advanced work in statistical methods if you so choose. I will be more concerned with teaching you the basic how and why of statistical generalization, than in making sure that we cover every statistic available. Last, you will acquire or refine a set of skills sets deemed valuable in the real world that can actually make you more employable.

SCHEDULE:

Week 1: <u>The Language of Scientific Research: Facts, Concepts, Hypotheses, Theories</u>
                Text, Chapter 1
                Martin Kelly (2015), "Predicting the presidential election with baseball."
https://www.thoughtco.com/predicting-the-presidential-election-with-baseball-104619

                For review: Complete odd numbered exercises, Text, Chapter 1 (answers in back)

Make sure you understand and can respond to the following:

- What is the difference between a unit of analysis/fact and a property? Between a property and a category of that property?

- E.G. If our analysis required us to measure the age of citizens legally eligible to vote in California:

    Units of analysis: eligible voters in California
    Total units: all eligible voters in California Property: Age
    Category: dependent on how we aggregate information:
            Any individual year (46 or 60 or 72)
            If combined into groups--e.g., "65 and older" or, if defined elsewhere,"older"

Note: "citizens over 65" is neither a property nor a category. It represents several units of analysis (a complete universe if we are only concerned about these individuals or a subset of all citizens within a certain jurisdiction).

- The nature/format of a hypothesis:

Falsifiability (potential):
        o properties are clearly defined explicit comparison
        o explicit direction of relationship
        o not a value judgment that is "true" based on faith

Not immediately verifiable: *general* statement--not true/false on limited investigation
        o safety "test"--do you have two properties (also called variables once we get to the point of measurement) =one implied "independent" (potential cause) and one implied "dependent" (outcome)?
        o If not, then you probably have a (immediately verifiable) statement -- t/f on limited investigation.

- What role does a "theory" or "theory sketch" play in hypothesis testing?
    - Theories explain WHY two properties SHOULD BE related the way our hypothesis suggests. Without a useful theory or theory sketch, we may just have an accidental or coincidental correlation. For example, is there any theoretical reason to believe that presidential victories should be influenced (as specified in the Kelly essay) by which league wins the World Series?

Week 2: <u>Variables, Measurement, and Beginning Statistics</u>
        Text, Chapter 2, 3, 4 (skip section on the IQV)
        McDonald, Michael P., Samuel L. Popkin. 2001. "The Myth of the Vanishing Voter."
            American Political Science Review. 95:4 (December) 963-74.
            www.uvm.edu/~dguber/POLS234/articles/mcdonald.pdf

For Review: Complete several odd numbered exercises, Text, Chapters 2-4 (answers in back)

Measurement Issues:

- Generally, what is an internal validity measurement problem?
- What is the 'Bradley effect'? Is it an internal or external validity problem?
- Should we now call this a "Trump" effect?
- What is the 'Chicago effect'?
- With surveys, what are the potential problems with question wording and ordering?
- Why might Political Scientists have overestimated the decline in turnout from 1960 on (see Popkin)?
- What is the difference between the VAP and VEP?
- Why is the difference relevant over the time period in the essay?

Levels of Measurement and Frequency Distributions:

- The different assumptions of nominal/ordinal/interval data--examples of each
- The importance of standardization when looking at frequency distributions
- How nominal data can be aggregated into interval data
- How categorization can influence our interpretation of frequency distributions

Central tendency:

- Differences among the mode, median and mean--which type(s) of data can each summarize?
- Skewness and what can be inferred from the difference between median and mean values.
- Why the mean of state figures in the text is not necessarily the same as national results.

Dispersion/variation (if we get to it):

- The meaning of the variation ratio and when it reaches 0, when it reaches 1
- *Do not worry about the IQV—skip this except for your own enlightenment*
- The Range
- The MAD
- Variance and Standard Deviation

READINGS AND VIDEOS for WEEK 3:

Chapter 5

Class Objectives:

- Understand the concept of standardization using Z-Scores
- Learn how to determine areas under a normal curve distribution
- Learn how to derive Z-Scores and translate them into percentiles
- Begin an understanding of what we can and cannot infer from a distribution

Chapter 6:

Class Objectives:

- Learn how to apply the normal curve to determine confidence intervals.
- Understand the "central limit theorem"
- Given a population mean, a standard deviation, and a sample size, how do we determine the range of values (confidence interval) for the means of 95% of a very large number of samples randomly drawn from a given population?
- What is the difference between a standard deviation and a standard error of the mean?
- Inference with SPSS--Examples

Chapter 7:

Objectives:

- Understand how the central limit theorem applies to mean differences
- Given differences in sample means, what can we infer about the difference in the means of the population from which the sample was drawn?
- SPSS examples to be used

You may also want to review Section 4.3 in the *SPSS Manual*

You may also want to view an SPSS video on T-Tests that I created a while back. Definitely review this before lab on Thursday:

<div align="center">

http://www.screencast.com/t/7kp0cWvCY

</div>

For a review of what we will cover in class, you may want to review these narrated PowerPoints that I prepared for my online class:

- Z-Score I: http://www.screencast.com/t/OEYO8UmO3y
- Z-Score II: http://www.screencast.com/t/gq3IHihlPC
- Z-Score III: http://www.screencast.com/t/6Fc6qi8JUt

READINGS AND VIDEOS for WEEK 4:

Chapter 9

Class Objectives:

- Understand how to test for the association between two nominal variables
    - Descriptively—relevant percentage differences
    - Descriptively—lambda
        - Be able to Interpret a PRE Measure
    - Inferentially—chi-square
        - How to determine degrees of freedom
        - How to interpret a chi-square value
        - When to confidently reject Ho, when not to
- Recognize that statistics carry different assumptions that can lead to different results
    - Statistical Independence
    - Accord

Chapter 10:

Class Objectives:

- Understand the requisites for a true experimental design
- Understand design parameters when a true experiment is not possible
- How to use controls with nominal data and for which purpose
    - Spurious relationship
    - Explanatory relationship
    - Specification/interaction effects

You may also want to review Section 4.5 in the *SPSS Manual*

You may also want to view an SPSS video on Crosstabs that I created a while back. Definitely review this before lab on Thursday. I have it set up in two formats (very large file). I recommend opening it into a full screen. Choose the format that works best for you.

https://drive.google.com/file/d/0B9Enziq2Gbb1WUExMGN4dUdzMG8/view?usp=sharing
https://drive.google.com/file/d/0B9Enziq2Gbb1UUZXM2xhelc5azA/view?usp=sharing

For a review of what we will cover in class, you may want to review these narrated PowerPoints that I prepared for my online class:

Chi-square: https://www.screencast.com/t/3ujxBrGncKGf

Lambda: https://www.screencast.com/t/Ps0qhfupJ

READINGS AND VIDEOS for WEEK 5:

Chapters 11-12 REGRESSION

Videos:

REGRESSION 1:

http://www.screencast.com/t/pwZYlnbo6P (Links to an external site.)Links to an external site.

REGRESSION 2:

http://www.screencast.com/t/1jxTsyJUx9S (Links to an external site.)Links to an external site.

REGRESSION 3:

http://www.screencast.com/t/iKOrbVVU (Links to an external site.)Links to an external site.

FOR LAB 1—July 13

Before Thursday:

      1. Review *SPSS Manual* Sections 1.1-3.2, 4.1-4.2
      2. Review Trial Run Lab Instructions (posted on TritonEd)
      3. Review RECODE and COMPUTE videos (urls posted)

      New procedures:

      RECODE – SPSS, Section 3.1

      COMPUTE –SPSS, Section 3.2

      FREQUENCIES --SPSS Manual, Section 4.1

      MEANS --SPSS Manual, Section 4.2

At Lab:

1. Familiarize Yourself with Layout/opening SPSS

2. Run "trial Run" if you haven't yet

3. WEIGHT by PW2016_FULL.

4. RECODE V36 into 3 new variables

    • One with 3 categories, leaners as partisans
    • One with 3 categories, leaners as independents
    • One with 3 categories measuring strength of partisanship

5. Run a FREQUENCIES of the original and recoded variables.

6. COMPUTE two new variables:

    • One comparing the feeling thermometer for Clinton with that for Trump

7. Run a FREQUENCIES of V16, V17 and your new variable. Ask for the following statistics: MEDIAN, MEAN, RANGE and STDEV

8. Use the MEANS procedure to compute the same statistics, but separately for men and for women

SYNTAX for SPSS:

```
WEIGHT BY PW2016_FULL.
RECODE v36 (1 2 3=1)(4=2)(5 6 7=3) into PIDP.
RECODE v36 ( 1 2=1)(3 4 5=2)(6 7=3) into PIDL.
RECODE v36 (1 7=1)(2 6=2)(3 4 5=3) into PIDS.
FREQUENCIES VARIABLES=V36 PIDP PIDL PIDS/BARCHART.
COMPUTE CANDDIFF=v16-v17.
FREQUENCIES variables=v16 v17 canddiff/statistics=mode median mean range stddev/barchart.
MEANS TABLES=V16 V17 CANDDIFF BY V1
 /CELLS COUNT MEDIAN MEAN STDDEV.
```

Project Write-Up Part I:

- Start with a hypothesis or hypotheses that link gender with your feeling thermometers. Remember that hypotheses are generalizations, and not just specific to these variables from this survey in that year.

- Do your results confirm your hypotheses?

  - Are women (on mean/median average) more warmly disposed towards the Democratic presidential candidate than are men?
  - Are women (on mean/median average) less warmly disposed towards the Republican presidential candidate than are men?
  - In comparing thermometer ratings for both candidates, are women (on mean average) more warmly disposed towards the Democratic presidential (as opposed to the Republican candidate) then are men?

- Feel free to use tables and/or graphs to help present your discussion.

FOR LAB 3—July 27 (draft 7.23.17)

Before Tuesday:

1. Review text—Chapters 9-10

Before Thursday

1. Review *SPSS Manual* Section 4.5
2. Familiarize yourself with CCES2016 data set (available soon)

   Old procedure: RECODE

At Lab:

1. Open up SPSS data set ANES2016.SAV as you did before
2. WEIGHT by the same weight-- PW2016_FULL
3. Once again, use variables V16 and V17
4. RECODE both variables into just 3 categories: negative (<50), neutral (=50), positive (>50)
5. To make sure that you do not lose the original 0-100 variation in V16 and V17 RECODE each into a new variable (V16A and V17A)
6. Review Example 3.4 (starting on p. 22) to determine how this is done.
7. Remember: you need to have two RECODE commands here, each ended by a period.

   New Procedure –
   CROSSTABS   section 4.5

1. Produce a crosstabulation with the new variables V16A and V17A as your *dependent* variable (defining the rows) and V1 (gender) as your *independent* variable.
2. Ask for column percentages, chi-square and lambda.
3. Describe what each of these 3 statistics tell you for *both* dependent variables (V16A and V17A)

   1. Relevant Percentage Differences (descriptive):
   - Are men more/less positive/negative than women on their feelings towards Hillary Clinton?
   - Are men more/less positive/negative than women on their feelings towards Donald Trump?

   2. Lambda (descriptive):
   - By what proportion do we reduce our error in guessing feelings with knowledge of the respondents' gender?
   - Use the V16A and V17A dependent values

c. <u>Chi-square</u> (inferential):
- If different (%), are the differences (V16A, V17A) "statistically significant"?
- Can we confidently reject the possibility that, in the population from which this sample was drawn, gender and feelings are "statistically independent" of each other?

d. Do you notice any interpretive differences between lambda and chi-square?

Now choose <u>*either*</u> the health care bill or position on gay marriage as you did last week:
- Run a controlled crosstab with either V22A (health care) or V59 (gay marriage) as your control variable. The independents variable I still V1, the dependent variables are still V16A and V17A.

Project Write-Up Part II-III:

- Do your results confirm your hypotheses, both descriptively (last week) and *inferentially*?

  o Are women (on mean average) more warmly disposed towards the Democratic presidential candidate than are men? (means and percentages): means, t-test, relevant percentage difference, lambda, chi-square
  o Are women (on mean average) less warmly disposed towards the Republican presidential candidate than are men? (means and percentages): means, t-test, relevant percentage difference, lambda, chi-square

- Now, using V22A or V59, does your control help to explain the gender gap in candidate sentiment? Does it extend it? More on this in lab.

CCES2016 instructions—next page

You will also, for this exercise, run a test to confirm/disconfirm your general 'gender gap' hypothesis with the CCES data for your state.

- Open up the CCES2016.sav data set from Tritoned. It is rather large.
- WEIGHT by POSTW.
- Select (use SELECT IF) your state by the associated number at the back of the codebook. For the rest of your analysis run, only respondents from your state will be included.
- There is no equivalent to the feeling thermometer in the CCES data set. Instead of feeling thermometers, you will use how one voted in the 2016 presidential election as your DV. For this purpose, not listed in your codebook, I have created a new variable, V56a, that only includes those who voted for either Trump or Clinton (2-party vote).
- Repeat, but this time with the CCES data set for your state, the entire CROSSTABS analysis that you performed with the ANES2016 data set:
  - Your IV is gender (V1)
  - Your DV is 2-party presidential vote (V56a)
  - Your control variable is either V41 (gay marriage) or V46 (repeat ACA).

Compare your results with what you determined nationally with the ANES2016 study. A few considerations:

- You are only looking at *one* state. The gender-party choice relationship may be different from what you produce for the entire country and for other states (note the difference in support for gay marriage and the ACA listed in your codebook).
- This is a *different* survey with somewhat different approaches to obtaining a fairly representative sample of the U.S. eligible voting population.
- As presidential vote choice is your DV, only those who *actually voted* (or claim to have voted) are included in your CCES analysis. When using feeling thermometers (ANES), all respondents are included, even those who didn't vote.
  - Can you figure out how to run the ANES analysis with only those who voted?

POLI 30 SUMMER-I 2017 DRAFT
CLASS PROJECT WRITE-UP (but one size DOES NOT fit all)

As I've been telling you, the final write-up should be based heavily on the individual weekly write-ups that you (well, some of you) have submitted.

But, to alleviate some stress, allow me to offer a reasonable, base outline of what you should have in your final analysis. I hesitate to be too detailed. Consider this a base and only a base. The university is pushing a new motto—"promote individuality" (or something like that). I wish to do so, but will settle for a stock final presentation. If yours reads differently, fine—you are promoting individuality.

STEP 1: What is this all about?

- This is where you lay out your hypothesis and, perhaps, a little piece on why it is important in current U.S. politics (you may add my 'gender gap' graph from class for dramatic effect—especially in describing that a "gap" doesn't always mean "on different sides"). You can state why you might think the "gender gap" to be greater in 2016 given the nature of the candidates. Remember, the hypothesis should be as general as possible. We started off with expressing a "relationship" between gender and how one feels about candidates, but it can (should) start off more broadly:

    > Example: *Men are more likely tobe favorably disposed towards (feel comfortable with, agree with, side with, and vote for Republican candidates) than women.*

STEP 2: General Discussion of How You Will carry Out Your Analysis

- Specify your broadly defined IV and DV.
- Specify how you will operationalize (fancy for measure) your variables.
- Gender will be based on self-assessment (but with only a binary choice provided due to survey constraints)
- DV can be measured in several ways:
    - How warmly disposed one is towards the two major presidential candidates as indicated by responses to a stock feeling thermometer item (perhaps you can describe what a feeling thermometer is).
    - How one voted in the presidential election (2-party only)

- Sources of the variables:
    - ANES2016
    - CCES2016 for one state
    - Differences between them

STEP 3: Step by Step (Week by Week)

Here you will take your general hypothesis and break it down into how you will test for it (confirm or disconfirm). This is usually of the form….

> If [hypothesis] is true, then we should expect to find that, in 2016, men were more ….. than were women.

> Example: "If men are more favorably disposed towards Republicans than are women, then, in investigating feeling thermometer differences between men and women, we should find that men, on mean average, were more warmly disposed towards the 2016 Republican presidential candidate, Donald Trump, than were women. Opposite for Hillary Clinton.

> Evidence (use and order as you see fit):

- Descriptive—what was the mean ft for men, for women for both Trump and Clinton
  (ANES 2016)?
  - o Is the difference as hypothesized?
  - o Anything noteworthy about the differences—which differences are more dramatic and for which gender?
- Inferential-
  - o Using one variable t-test, what can we infer about men and women's ft ratings in the population from which this sample was drawn (is neutrality part of our confidence projection?)
  - o Using an independent (grouped) sample t-test are those differences (m-f) statistically significant—what does that mean?

Next: Discuss the 'tests' that you will conduct with CROSSTABS:

    1. I'll start you off: If men are more favorably disposed towards Republicans than are women, then we should find that a higher percentage of men were positively disposed towards (had a positive feeling thermometer rating of) Donald Trump in 2016 than did women.

- Discuss how you RECODED the fts in order to allow for a decent crosstabular analysis.
- Descriptive—do the relevant percentage differences confirm your hypothesis?
  --does lambda
- Inferential—what does the chi-square tell us? Can we confidently reject statistical independence in the population from which the sample was drawn?

    2a. Now, substitute the 2016 presidential vote (2-party) for your DV. Re-specify the test (different DV) and go through the descriptive and inferential analysis above.

- Here are the results for the ANES2016 data set (I removed 3rd party voters):

| | | V1 R self-identified gender | | |
		male	female	Total
V66a2	Trump	563	742	1305
		49.3%	55.5%	52.7%
	Clinton	578	595	1173
		50.7%	44.5%	47.3%
Total		1141	1337	2478
		100.0%	100.0%	100.0%

			Value
Nominal by Nominal	Lambda	V66a2 Dependent	.013

Chi-Square Test

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Continuity Correction[b]	9.109	1	.003		
N of Valid Cases	2478				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 540.11.

b. Computed only for a 2x2 table

2a. After interpretation of 2016, produce output and discuss same results from CCES2016 (all states+DC). Any differences? If so, any conjecture as to why?

2c. Produce output and discuss same results from CCES2016 (just your state)
     Any differences from all of CCES?

I will produce a comparative table for all states after class is over and I re-celebrate "National Scotch Day"

3. Using an explanatory control:

T-TEST: You selected those who were in favor of gay marriage/the ACA and, separately, opposed from the ANES2016 data set. Did views on your issue help to reduce (i.e. explain) the mean differences between males and females?

CROSSTABS: You controlled for a third variable (gay marriage or ACA). What happened to your percentage differences? Lambdas? Chi-squares?

Note: you need only do this for your state, but we won't reject a full CCES analysis here

Note: a t-test and/or chi=square can go from significant to insignificant for two reasons:
- Descriptive difference is smaller
- Sample size is reduced by splitting the sample in two

REGRESSION –WEEK 5 coming soon to an email near you!

STEP 4: Summarize—what do your data say (confirm/disconfirm) about your original hypothesis?