

SOC 3050—Statistics for Sociology

Spring 2021

Tuesdays and Thursdays, 11:30-12:45, Online synchronous meetings

Professor: Tim Bartley

Office Hours: Tuesday 3:30-4:30 pm, Wednesday 9:15-10:15 am

During these times, you can drop in and find me on Zoom using the links on our Canvas page. Other times are possible by appointment.

Email: bartleyt@wustl.edu

Teaching Assistant: Rachel Sznajderman

Office Hours: TBA

Email: rachelsznajderman@wustl.edu

Course Description

Quantitative data analysis has become a structuring force in society. It powers scientific discovery, policy decisions, and algorithms directing flows of information, resources, and social control. Interpreting and conducting quantitative data analyses are also essential skills for sociology majors, among others. They are also marketable skills. This course is designed to clarify the logic behind statistical analyses, develop your quantitative analytic skills, and de-mystify the varied uses of statistics in society and in the social sciences.

We will meet online via Zoom at the scheduled class time and seek to mimic an in-class experience as much as possible. You should be at your computer, actively taking notes, asking questions, and engaging in discussions (both among the whole class and in small breakout groups) during our class meetings. Your camera should be on so that we can make this format work for intensive class-sessions. I understand that this is an unusual time, and circumstances may arise on occasion that make it difficult to be fully present in these ways. If this happens on more than a few occasions, we will need to talk about whether and how you can commit to the course at this time. If things start to go haywire in your life, please let me know as soon as you can.

The first part of the course focuses on *descriptive statistics*—that is, techniques for summarizing quantitative data within a particular sample or set of cases. We will start by describing a single variable—that is, some aspect of the social world that varies from case to case or over time. We will then look at relationships between two or more variables in order to understand how one part of the social world shapes or causes another. The second part of the course focuses on *inferential statistics*. The goal here is to make conclusions (inferences) about an entire population based on observing only a small part of that population (a sample). Once we develop the tools for making inferences, we will use those to extend the material from the first part of the course, in order to make sense of large-scale social trends, patterns, and their possible causes.

The course will also provide some practical experience doing data analysis in Stata, a software package that is widely used in social scientific settings. You will gain valuable skills and experience using Stata, but this is not a course that is focused on programming. Instead, the course provides a strong

conceptual foundation for understanding the *what* and *why* of statistics, along with some initial programming skills (the *how* of data analysis) that you can apply and extend beyond this course.

In addition to learning the principles of social statistics and some data analysis skills, this course seeks to help you make sense of quantitative data and analytics in the real world. This should help you be a smart, critical consumer of quantitative information.

I assume no prior knowledge of statistics, and the course is not particularly math-intensive. Instead, I will emphasize the logic of describing variation, making comparisons, moving from samples to populations. In addition, there will be an emphasis on developing—and writing—clear, substantive interpretations of results. If you are a bit math-phobic, don't panic—you will have lots of opportunities to get more comfortable working with quantitative data.

If you work hard in this class, I believe you can gain the following:

- A strong foundation in the conceptual logic of statistics, which should serve you well if you decide to seek further expertise in this area;
- Experience and practice in talking about and writing about quantitative analyses;
- Practical experience organizing and analyzing data in Stata, which you can later apply in a variety of organizational and/or research settings;
- A generalizable set of tools for critical thinking.

Required Texts

Bergstrom, Carl T. and Jevin D. West. 2020. *Calling Bullshit: The Art of Skepticism in a Data-Driven World*. Random House.

Available for purchase (at a low price) at the Wash U bookstore and elsewhere. Alternatively, one user at a time can read online via the Wash U library catalog. Check the library catalog for details.

A few other short articles or chapters on uses or misuses of statistics , which will be available on Canvas

Additional resource:

Online Statistics Education: A Multimedia Course of Study (<http://onlinestatbook.com/2>). Project Leader: David M. Lane, Rice University.

This is a free resource that provides basic foundations and several useful simulations of core statistical concepts and theorems. You should use it in the following ways: (1) When you are confused about something in class, look here for additional explanation and examples. The approach in this resource (unlike some others online) is very similar to the approach of this class. (2) As we get into principles of sampling, sampling distributions, and regression analysis, play with the simulations that are available here and on the associated Rice Virtual Lab in Statistics. (Some will require installing Java.)

Recommended Texts

If you are interested in doing quantitative analyses beyond this class (e.g., for a thesis or other research opportunity), I highly recommend getting a copy of the following books:

Pevalin, David and Karen Robson. 2009. *The Stata Survival Manual*. Open University Press. (Available as an ebook through the Wash U library via [this link](#).)

Miethe, Terance D. and Jane Florence Gauthier. 2008. *Simple Statistics: Applications in Social Research*. Oxford University Press.

Required Software

Stata: This is a widely used and user-friendly program for advanced statistical analysis. You can buy a 6-month student license for Stata/IC for \$48.

(<https://www.stata.com/order/new/edu/gradplans/student-pricing/>) It is also possible to use Stata for free through Wash U's virtual lab: <https://wustl.apporto.com/> and then choose ArtSci. If you have a reliable internet connection, then this method should work well enough for our purposes, but it will be a bit slower than using Stata on your own computer. If you would like to build your skills beyond what is required for this course, then buying a copy of Stata would be wise.

Microsoft Excel: This widely used spreadsheet program is useful for building a dataset. (It can also be used to do many basic statistical analyses and make some useful graphs, though we will mostly do these tasks in Stata.) If you do not already have Excel installed on your computer, you can access and download it through WashU's Office 365 program via these links:

Windows: <https://it.wustl.edu/items/microsoft-office-for-windows/>

Mac: <https://it.wustl.edu/items/microsoft-office-for-mac/>

Other Necessities

-A laptop or desktop computer with a decent internet connection and a large enough screen to see the lecture presentations. Most any laptop or desktop will be fine, but it will definitely not work to use a smartphone. An iPad/tablet will probably be too small too.

-A space to work where you can both see the screen and *take notes by hand* (pen-and-paper). Note taking is important in this class, and given the character of the lectures, it will not work to try to type notes out. Any kind of table or desk should be fine, but trying to have everything on your lap would probably be quite difficult.

-A basic calculator (that can add, subtract, multiply, divide, and *take square roots*). Using the calculator on your computer should be okay, but please familiarize yourself with its functions so that you can use it quickly when necessary. The calculator on your phone may not take square roots, which would make it an inappropriate choice.

-The ability to scan or take a picture of your written work and upload it to our Canvas site.

If any of these things is likely to be difficult for you this semester, please let me know. There should be some assistance available if you need help with hardware or your internet connection. You may also contact Student Financial Services for information on the availability of technology grants:

<https://financialaid.wustl.edu/contact-us/>

Course Requirements

Problem sets

I will distribute problem sets regularly, typically one for each lecture day. It is important to find a time to work on these soon after our class meetings. This is not a reading-centered class but rather a problem-centered class, so you should plan to invest time in working through the problem sets. Use your lecture notes as a guide when doing the problem sets. Then compare your work to the answer key and revise whatever parts need correction or improvement.

Visible edits and additions are fine—and indeed encouraged, since they demonstrate that you have been conscientious about completing the problems based on your notes and then using the answer key to improve.

You should scan/photograph your work on each problem set and upload it to our Canvas site no later than 5:30 pm on the day before the next class. (In other words, problem sets distributed on Tuesday should be completed and uploaded no later than 5:30 pm on Wednesday. Those distributed on Thursday should be completed and uploaded no later than 5:30 pm on Monday.) This schedule will normally allow us to provide feedback before the next class.

If your schedule allows it, I highly recommend completing the problem sets very soon after you receive them, so that the material is fresh. This also gives you a cushion if a particular problem set takes longer than expected.

Much of your work on the problem sets will be done by hand (i.e., with paper and pen/pencil), but sometimes you will also be asked to complete tasks in Stata, in which case you should take and upload screenshots of the completed tasks.

The problem sets will not be graded, but they will be assessed for completeness and carefulness using the following rating system:

3: Good (meaning that the problems are complete and carefully done/interpreted. It is up to you to ensure that your answers are of the highest quality and that you are absorbing the material. This rating only indicates that you appear to be taking that task seriously.)

2: OK but needs revision (meaning that you are on the right track but should go back and re-check

the answer key and revise accordingly. You do not need to submit your revision, but you should check it yourself and let us know if you need help.)

1: Needs Significant Additional Work (meaning that the problems are incomplete or that you appear not to have taken full advantage of the resources available in completing them. In this case, you should finish and/or re-do the problem set. You do not need to submit your revision, but you should check it yourself and let us know if you need help.)

0: Missing (not received by the deadline)

Your average rating across all problem sets (approx. 18-20 in total) will then be divided by 2.65 and translated onto a 100 point scale (max. 100). For example, if your average were 2.65 or more, you would receive all 100 points. If your average were 2.5, you would receive 94.3 points $((2.5/2.65)*100=94.3)$. If your average were 2.0, you would receive 75.5 points $((2.0/2.65)*100=75.5)$. Etc. This provides some significant leeway for mistakes and days when things don't go your way. (Even an occasional "1" rating won't hurt much.) But it is important not to take this for granted. You should plan to complete all problem sets carefully and on time.

Most importantly, doing the problem sets carefully is by far the best way to gain a solid understanding of the material and to develop your skills. In a course like this, learning by doing is essential. Doing the problem sets carefully will prepare you for the projects described below.

First Project

In this project, you will use Stata and the tools of descriptive statistics to analyze data on the 50 U.S. states. Your job will be to choose variables of interest, conduct the appropriate analyses, and interpret your results carefully in order to describe patterns among the 50 states. All of this should follow the procedures learned in class, as well as the specific instructions for this assignment, which will be distributed roughly 1.5 weeks before the due date. Due date: March 16

Final Project

The course will culminate in a major data analysis project, in which you use will not only describe patterns in a sample of American adults but also make inferences about the American population as a whole. This will be our main focus during the last part of the semester, and it will utilize the skills you develop over the course of the semester—including Stata skills but also skills related to choosing the right method of analysis and interpreting the results carefully.

The final version of your report is due during the final exam period, no later than May 12. You will present the preliminary work on your project to the class in late April/early May. Some parts of the project may be done in small groups of two. Full instructions and guidelines will be provided by mid-late April.

Reading Response Essays

On two occasions, you will be asked to write a short essay (less than 1 single-spaced page) summarizing and responding to the course readings, such as chapters from the *Calling Bullshit* book and/or other course readings. This is not a reading-intensive course, but the readings will nevertheless provide important guidance as we work to unpack the practical uses and social significance of quantitative data/analytics.

First essay: Due Feb. 18, covering the first 5 chapters of *Calling Bullshit* as well as the excerpts from *Stat-Spotting* and the article on “Race and America: Why Data Matters.”

Second essay: Due April 6, covering the final 6 chapters of *Calling Bullshit*.

“Stats in Practice”

On one occasion, you should post and present some “real world” example of statistics or quantitative data analysis. Your item could come from (1) news and media sources, (2) a reading from one of your other courses (such as a research article using quantitative analysis), (3) online materials (such as data visualizations, data dashboards, etc.). Your goal should be to briefly describe the example and raise questions about it, ideally drawing links to material from class, *Calling Bullshit* book or other course readings. Your questions may be analytical, critical, simply descriptive (i.e., “What did they actually do to reach this conclusion?”)

Keep it short. Once you decide on the example you want to use, find a way to share it in no more than 5 minutes and just 1-4 images/slides. You don’t need to cover everything. Instead, hone in on what you see as the most interesting or relevant part of a news story, a table of results in a research article, a data visualization, etc.

You will be assigned a roughly two-week period when you are responsible for posting and presenting your item. If you see something relevant prior to your assigned period, you are welcome to present/post it at that point instead.

Please have the item ready in a format that can be shared—by sharing your screen in Zoom. Your item should also be posted to Canvas before or shortly after class. You will earn 30 points for presenting and posting your item in the assigned period (or prior to that if you prefer).

Participation and Engagement (including contributing to a “Stata Wiki” page)

You can earn up to 30 points for demonstrating consistent engagement in the course. At a minimum, this means being present, prepared, and involved in our class meetings throughout the semester. In addition, you should contribute to a “Stata Wiki” page in which we will collectively catalog the Stata commands we are using and add notes about tips, tricks, and challenges encountered. If circumstances arise that hinder your ability to participate in the class sessions, please let me know as soon as possible so that we can discuss alternative arrangements. I understand that these are unusual times, which come with a range of new stressors and unexpected events. I am happy to talk with you at any point and/or help you access other resources as needed.

How to do well in this course: Some courses demand only minimal attention during most of the semester and intense cramming at a few points. This is not one of those courses. This is a “doing” class more than a “reading” class, and you will be expected to demonstrate your budding statistical skills regularly. The lectures will be the primary source of information about the material covered in this course, and the online textbook will be used primarily as a reference and a way to fill in some of the gaps. Taking detailed notes is essential.

If you are engaged in the course sessions, take good notes, do the problem sets carefully, and follow along with the Stata instructions, you will find yourself learning rewarding new skills—and even enjoying it!

Rachel and I are here to help, so please do not hesitate to talk with us. If you have questions or find yourself starting to get confused, please let us know before it is too late.

Grading: Grading will be based on the number of points you earn as a percentage of the total possible points. The table below shows the number of points possible for each component.

<u>Item</u>	<u>Points</u>
Problem sets	100 points
First project	100 points
Final project	200 points
Reading response essays (2 @ 40 points each)	80 points
Stats-in-the-news presentation	30 points
Participation	<u>30 points</u>
	540 total points possible

Final grades will be based on the number of points you earn as a percentage of the total points possible (rounding up to the nearest percentage).

A+ 97-100%	B 83-86%	C- 70-72%
A 93-96%	B- 80-82%	D+ 67-69%
A- 90-92%	C+ 77-79%	D 60-66%
B+ 87-89%	C 73-76%	NCR 59% or less

Policy on late work

In general, you are expected to complete all work at the scheduled time. If (1) truly extraordinary and unavoidable circumstances delay your submission AND (2) you notify me in advance of the due date, then you will be eligible for an adjusted timeline. Otherwise, papers and projects that are submitted late will be subject to a 10% penalty for each 24-hour period beyond the due date.

Policy on Academic Integrity

Ethical behavior is an essential component of learning and scholarship. Students are expected to understand, and adhere to, the University's academic integrity policy: wustl.edu/policies/undergraduate-academic-integrity.html. Plagiarism or unauthorized collaboration will result, at a minimum, in a loss of credit for the assignment. Students who violate the academic integrity policy will also be referred to the Academic Integrity Policy Committee. If you have any doubts about what constitutes a violation of the Academic Integrity policy, or any other issue related to academic integrity, please talk to me.

Inclusive Learning Environment Statement

The best learning environment—whether in the classroom, studio, laboratory, or fieldwork site—is one in which all members feel respected while being productively challenged. At Washington University in St. Louis, we are dedicated to fostering an inclusive atmosphere, in which all participants can contribute, explore, and challenge their own ideas as well as those of others. Every participant has an active responsibility to foster a climate of intellectual stimulation, openness, and respect for diverse perspectives, questions, personal backgrounds, abilities, and experiences, although instructors bear primary responsibility for its maintenance.

A range of resources is available to those who perceive a learning environment as lacking inclusivity, as defined in the preceding paragraph. If possible, we encourage students to speak directly with their instructor about any suggestions or concerns they have regarding a particular instructional space or situation. Alternatively, students may bring concerns to another trusted advisor or administrator (such as an academic advisor, mentor, department chair, or dean). All classroom participants—including faculty, staff, and students—who observe a bias incident affecting a student may also file a report (whether personally or anonymously) utilizing the online Bias Report and Support System (brss.wustl.edu).

Resources for Students

1. **DISABILITY RESOURCES:** If you have a disability that requires an accommodation, please speak with instructor and consult the **Disability Resource Center** (<https://students.wustl.edu/disability-resources/>)
2. **WRITING ASSISTANCE:** For additional help on your writing, consult the expert staff of **The Writing Center** (writingcenter.wustl.edu). It can be enormously helpful to ask someone outside a

course to read your essays and to provide feedback on strength of argument, clarity, organization, etc.

3. THE UNIVERSITY'S PREFERRED NAME POLICY FOR STUDENTS, with additional resources and information, may be found here:
registrar.wustl.edu/student-records/ssn-name-changes/preferred-name-policy/preferred-name-policy-student/.
4. SEXUAL ASSAULT: The University is committed to offering reasonable academic accommodations to students who are victims of sexual assault. Students are eligible for accommodation regardless of whether they seek criminal or disciplinary action. Depending on the specific nature of the allegation, such measures may include but are not limited to: implementation of a no-contact order, course/classroom assignment changes, and other academic support services and accommodations. If you need to request such accommodations, please direct your request to Kim Webb (kim_webb@wustl.edu), Director of the Relationship and Sexual Violence Prevention Center (<https://students.wustl.edu/relationship-sexual-violence-prevention-center/>). Ms. Webb is a confidential resource; however, requests for accommodations will be shared with the appropriate University administration and faculty. The University will maintain as confidential any accommodations or protective measures provided to an individual student so long as it does not impair the ability to provide such measures.

SEXUAL ASSAULT REPORTING: If a student comes to me to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if I otherwise observe or become aware of such an allegation, I will keep the information as private as I can, but as a faculty member of Washington University, I am required to immediately report it to my Department Chair or Dean or directly to Ms. Jessica Kennedy, the University's Title IX Director. For information on the Title IX office, reporting options, and confidentiality, see <https://titleix.wustl.edu/students/confidentiality-resources-support/>.

5. MENTAL HEALTH: Mental Health Services' professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See:
<https://students.wustl.edu/habif-health-wellness-center/>

Outline of topics to be covered

This updated outline provides information on the topics to be covered and links to the most relevant sections of the online textbook. Please use this as a reference if you want to see where we are headed or where to look for additional guidance.

Topic

Sections of online textbook

I. Introduction: Key concepts

<http://onlinestatbook.com/2/introduction/introduction.html> (Sections: What are statistics? Importance of statistics, Descriptive Statistics, Inferential Statistics, Variables, Levels of Measurement)

II. Describing the Social World: Univariate Descriptive Statistics

A. Distributions

<http://onlinestatbook.com/2/introduction/distributions.html>
http://onlinestatbook.com/2/graphing_distributions/graphing_distributions.html (Sections: Qualitative variables, Quantitative variables, Histograms)
http://onlinestatbook.com/2/summarizing_distributions/shapes.html

B. Central tendency of a distribution

http://onlinestatbook.com/2/summarizing_distributions/central_tendency.html
(Sections: What is central tendency? Measures of central tendency, Balance scale simulation, Mean and Median, Comparing measures)

C. Dispersion/variability of a distribution

http://onlinestatbook.com/2/summarizing_distributions/variability_ov.html
(Sections: Measures of variability, Variability demo)

III. Analyzing Relationships: Bivariate Descriptive Statistics

A. Two qualitative variables: Contingency tables and related measures of association

http://onlinestatbook.com/2/chi_square/contingency.html
http://onlinestatbook.com/2/chi_square/one-way.html
http://onlinestatbook.com/2/case_studies/diet.html
(caveat: These chapters use contingency tables to make inferences, so some content is premature for our purposes.)

B. Comparing groups: Qualitative independent and quantitative dependent variable

http://onlinestatbook.com/2/graphing_distributions/boxplots.html
http://onlinestatbook.com/2/case_studies/ratings.html

C. Two quantitative variables: Regression and correlation

http://onlinestatbook.com/2/describing_bivariate_data/bivariate.html
(Sections: Introduction to bivariate data, Values of the Pearson correlation, Guessing correlations, Properties of Pearson's r , Computing Pearson's r)
<http://onlinestatbook.com/2/regression/regression.html>
<http://onlinestatbook.com/2/regression/partitioning.html>
http://onlinestatbook.com/2/case_studies/tv_violence.html
http://onlinestatbook.com/2/regression/regression_toward_mean.html

D. Introduction to multivariate analysis

Reading on Blackboard: "Why do multivariate analysis?"

IV. Foundations of Statistical Inference, part 1

- A. Probability and random variable distributions <http://onlinestatbook.com/2/probability/probability.html> (Sections: Introduction, Basic concepts, Conditional p demo, Monty Hall problem)
- B. Sampling distributions http://onlinestatbook.com/2/sampling_distributions/sampling_distributions.html (Sections: Introduction, Basic demo, Sample size demo, CLT demo, Sampling distribution of the mean)
- C. Normal and standard normal distributions http://onlinestatbook.com/2/normal_distribution/normal_distribution.html (Sections: Introduction, History, Areas of normal distributions, Standard normal)

V. Foundations of Statistical Inference, part 2

- A. Estimation <http://onlinestatbook.com/2/estimation/estimation.html> (Sections: Introduction, Confidence intervals intro)
- (1) Confidence interval for a mean http://onlinestatbook.com/2/estimation/confidence_ov.html (Sections: Confidence interval for the mean, t distribution, Confidence interval simulation)
- (2) Confidence interval for a proportion http://onlinestatbook.com/2/sampling_distributions/samp_dist_p.html
http://onlinestatbook.com/2/estimation/proportion_ci.html
- B. Hypothesis testing
- (1) Logic of hypothesis testing http://onlinestatbook.com/2/logic_of_hypothesis_testing/logic_hypothesis.html (All sections)
- (2) Hypothesis test about a single mean http://onlinestatbook.com/2/tests_of_means/single_mean.html

VI. Testing Hypotheses about Bivariate and Multivariate Analyses

- A. Comparing groups: Difference of means test and its extensions http://onlinestatbook.com/2/tests_of_means/difference_means.html
http://onlinestatbook.com/2/tests_of_means/robust_sim.html
http://onlinestatbook.com/2/tests_of_means/pairwise.html
- B. Hypothesis tests about contingency tables http://onlinestatbook.com/2/chi_square/distribution.html
http://onlinestatbook.com/2/chi_square/contingency.html
- C. Hypothesis tests about regression coefficients
- (1) Simple regression <http://onlinestatbook.com/2/regression/accuracy.html>
<http://onlinestatbook.com/2/regression/inferential.html>
- (2) Multiple regression http://onlinestatbook.com/2/regression/multiple_regression.html
- D. Extensions and additional tools, as needed for your projects

Reminder of due dates

Note that problem-sets will be distributed immediately after most class sessions. As explained above, these are due by 5:30 pm on the day before the next class.

Please also remember that you will be assigned an approximately 2-week period in which to present and post your Stats in the News item. Other important due dates are shown below:

Jan. 28-Feb. 2: Be ready to discuss the excerpts from *Stat-Spotting* and the article, "Race and America: Why Data Matters," both posted on Canvas

Feb. 4-16: Read the first five chapters of *Calling Bullshit* (on your own schedule) and be ready to share your reactions in class

Feb. 18: Reading response essay #1 due

March 16: First project due

March 18-April 1: Read the last six chapters of *Calling Bullshit* (on your own schedule) and be ready to share your reactions in class

April 6: Reading response essay #2 due

April 27-May 4 (approx.): Presentations of preliminary work on the Final Project

May 12: Final project due