

JSGS 807 STATISTICS FOR PUBLIC SECTOR MANAGERS

Fall 2020	University of Regina
Instructor:	Justin Longo , Assistant Professor and Director, Digital Governance Lab
Phone and Email:	306-450-5345 (sms preferred; phone only if necessary) justin.longo.phd@gmail.com
Office Hours:	Email or Slack me to request time
Office Location:	All office hours & individual meetings will be held online via video or telephone
Classroom:	All classes will be held online via Zoom with recordings available
Dates:	Fall 2020 — September 8 to December 1 (no synchronous class times)
Group Communications:	Twitter Class List Google Drive folder Class Recordings Slack
Teaching Assistant:	

COURSE CONTENT AND APPROACH

Official Calendar Description: Administrative decision making and policy development often require the analysis of quantitative data. This course will introduce students to descriptive and inferential statistics often used in policy environments so that they will be effective data users and interpreters. Students will be taught how to use and present descriptive statistics.

Justin's Description: Get ready for a stats class like you've never had before! This is a course for future public sector leaders with limited background in (and even an aversion to) mathematics, statistics, or economics, providing an overview of statistical concepts, principles, and techniques that are relevant and useful in public administration and policymaking. The central aim of this course is to make you an intelligent consumer of statistical analysis, and a strategic user of statistics in policy and administrative processes, through critical assessment and working first-hand with datasets and spreadsheets. The core argument that this course advances is that policy analysts and public sector managers responsible for policy development and delivery have a special responsibility to manage and use data and statistics accurately and ethically. To exercise this responsibility requires a conceptual understanding of data and analysis beyond the techniques of statistical analysis. In response, this course offers an introduction to the language, techniques, and tools involved in the collection, analysis, interpretation, and presentation of data and statistical analysis in support of policy making and administration, an exploration of critical questions in the use of statistical analysis in the public sector, and an assessment of the best ways to support informed policy and administrative decisions using data and statistical analysis.

Nothing that you will learn in the course of your studies will be of the slightest possible use to you in after life – save only this – that if you work hard and diligently you should be able to detect when [someone] is talking rot, and that, in my view, is the main, if not the sole, purpose of education.

- John Alexander Smith, 1914 ([ref](#))



LEARNING OBJECTIVES

Following successful completion of this course, students should be able to:

- Analyze (with a focus on spreadsheet skills), interpret, and present policy-relevant data and statistics.
- Ask critical questions and be an informed consumer of statistical analysis as support for policy making and the administration of public programs and services.
- Map the appropriate application of statistical analysis to specific policy and administration problems.
- Critically evaluate the public benefit of investments in data collection and analysis, and the value of publicly releasing public sector information (i.e., open data), and
- Articulate public service responsibilities surrounding the collection, interpretation, communication, and presentation of public sector data and statistics.

ONLINE CLASSES

In light of the rapid reconfiguration of UofR classes in light of the COVID-19 pandemic, JSGS 807 has been transformed into an online format. This entails a significant change in the course design and delivery, and will require patience and extra effort from all of us as we adapt to this new setting.

Draft fall 2020 version:

- The course will be entirely online this fall.
- The class time (Tuesday evenings) is purely a legacy issue, as we will have no synchronous "class time".
- The class will consist of:
 - short video "lectures",
 - video tutorials,
 - readings,
 - exercises (much of it spreadsheet based) to apply the issues discussed in the videos, readings, and class notes.
 - A major assignment that involves analyzing a data set and preparing an infographic based on your analysis.
- Two additions being developed:
 - each student will meet by video with the course instructor 3 times throughout the term (for approximately 20 minutes) to discuss the course topics. These meetings will form part of your evaluation;
 - study groups will be formed where you can meet over video with some of your colleagues to discuss the course, assignments, etc. These are not graded, but are organized for your benefit.



COURSE AT-A-GLANCE

Session	Date	Topics
0	Sept 1	Preview of the syllabus and tech overview
1	Sept 8	Course Overview: Frameworks and Concepts
2	Sept 15	Foundations: Descriptive Statistics
3	Sept 22	Foundations: Inferential Statistics
4	Oct 6	Data Collection by Listening to People: Polling, Surveys, and Social Listening
5	Oct 20	Data Collection by Other Means: Administrative Data, Big Data, and Open Data
6	Oct 27	Models, Simulations, and Scenario Tools
7	Nov 3	Indices and Indicators
8	Nov 17	Probabilities and Predictive Analytics
9	Nov 24	In Graphic Detail: Information Visualization & Infographics
10	Dec 1	Can Government Play Moneyball, and Ethical Statistics

DUE DATE SCHEDULE

Item	Value	Due
Web Safari 1 #jsgs807fall2020ws1	1%	1pm Sept 8
Web Safari 1 #jsgs807fall2020ws2	1%	1pm Sept 15
Web Safari 1 #jsgs807fall2020ws3	1%	1pm Sept 22
Web Safari 1 #jsgs807fall2020ws4	1%	1pm Oct 6
Web Safari 1 #jsgs807fall2020ws5	1%	1pm Oct 20
Web Safari 1 #jsgs807fall2020ws6	1%	1pm Oct 27
Web Safari 1 #jsgs807fall2020ws7	1%	1pm Nov 3
Web Safari 1 #jsgs807fall2020ws8	1%	1pm Nov 17
Web Safari 1 #jsgs807fall2020ws9	1%	1pm Nov 24
Web Safari 1 #jsgs807fall2020ws10	1%	1pm Dec 1
Session Lab 1	5%	1pm Sept 15
Session Lab 2	5%	1pm Sept 22
Session Lab 3	5%	1pm Oct 6
Session Lab 4	5%	1pm Oct 20
Session Lab 5	5%	1pm Oct 27
Session Lab 6	5%	1pm Nov 3
Session Lab 7	5%	1pm Nov 17
Session Lab 8	5%	1pm Nov 24
Session Lab 9	5%	1pm Dec 1
Session Lab 10	5%	1pm Dec 8
Statistical Analysis Assignment Part 1	20%	1pm Oct 13
Statistical Analysis Assignment Part 2	20%	1 pm Dec 15



COURSE SESSIONS

Session 1 - Course Overview: Frameworks and Concepts

In this first full session, we will discuss the framework for the course stressing the need for policy analysts and public sector managers to use data and statistics accurately and ethically. And since much of the work done in this course (and a lot of statistical analysis done in the public sector) uses spreadsheets, we'll get started with some basic spreadsheet software exercises and techniques.

Session 2 - Foundations: Descriptive Statistics

Descriptive statistics involves the analysis of data that *describe* or summarize a sample or population. Descriptive statistics can be presented using tables, graphs and charts, and a narrative discussion of the results. Descriptive statistics will be the most frequent type of statistical analysis that you will produce or consume as a public servant.

Session 3 - Foundations: Inferential Statistics

Descriptive statistics are calculated on a population where you have data on every entity in a group. Often, however, you do not have access to the whole population you're interested in, but only a limited "sample" which is then used to make *inferences* about the whole population.

Session 4 - Data Collection by Listening to People: Polling, Surveys, and Social Listening

In these two related sessions, we'll consider how to source the data necessary for undertaking statistical analysis. In this first session, we'll focus on data that can be collected by asking people what they think, observing what they do, accessing information that they've recorded, or by listening to what they're saying in public.

Session 5 - Data Collection by Other Means: Administrative Data, Big Data, and Open Data

Governments collect, generate, and continually compile vast amounts of data from collecting vital statistics, to administering the tax system, recording government operations activity, managing public infrastructure and natural resources, surveying and recording public and private lands, processing regulatory requirements, managing social service delivery, and the census and survey work conducted by public statistics agencies. Now, with the emergence of new technological mechanisms for collecting "big data", and with vast sources of open data available to us, a whole new range of data sources are beginning to influence government policymaking and public administration. "Open Data" refers to initiatives by governments to make publicly available the vast stores of data they hold in filing cabinets and isolated computer hard drives. In this second part of our look at "data collection", we'll look at the vast range of data and the risks of the problem of too much data.

Session 6 - Models, Simulations, and Scenario Tools

Models are used in the decision-making process to provide insight into the consequences of different courses of action. Spreadsheets are a simple tool for the development of business decision models, allowing us to ask "what-if?" questions in the face of different quantitative inputs and probabilities.



Simulations model how markets, societies, organizations, and natural systems dynamically interact and react to changing assumptions, allowing users to consider different possible courses of action. And scenario tools represent a computerized version of scenario planning, using intuitive, game-like user interfaces that allow non-experts to explore the possible consequences of different policy choices. This session will explore how to build and work with spreadsheet models, simulations, and scenario tools.

Session 7 - Indices and Indicators

Index numbers (or indices) are among the most commonly used techniques in the area of economic statistics, used to combine large amounts of data about a given variable into a single number. Changes in this index number can then be compared over time. Some common statistics measured using index numbers include inflation, stock market prices, and volumes of production. And an indicator is a summary statistic or combination of statistics that explains system conditions from a high-level view. A good indicator highlights a problem early and identifies what needs to be done to solve it.

Session 8 - Probabilities and Predictive Analytics

Probability is a way of expressing how likely it is that a future event will happen. For public sector managers, understanding the probability of future events can help with planning, regulatory compliance, and policy analysis. But it is equally important to understand that a high probability of something happening also means the event might not happen. Predictive analytics, as a “big data” form or forecasting, can serve as an input into framing a policy problem before it appears, indicating where a need is being unmet or where an emerging problem might be countered early.

Session 9 - In Graphic Detail: Information Visualization & Infographics

Visualization is a category of techniques used for data analysis and the communication of quantitative findings using informative graphics. Good visualization should help the viewer understand the underlying data, appreciate the information contained in the analysis, and grasp the complexity of the system. Infographics are visual representations of information and data, combining visual summaries and explanatory text, for presenting information quickly to a generalist audience and in an appealing way.

Session 10 - Can Government Play Moneyball, and Ethical Statistics

In 2003, Michael Lewis published the book *Moneyball: The Art of Winning an Unfair Game* (later to become the movie *Moneyball*), that told the story of how the Oakland A’s became a high performing Major League Baseball team despite having about a third of the financial resources as the league’s wealthiest teams. The central point of *Moneyball* was that deep statistical analysis, and not intuition nor superficial statistics, could be used to make better decisions. Here, we take a step back from the techniques of data analysis to ask if we should increase the resources devoted to statistically evaluating the effectiveness of government programs? Or do decision makers care about statistical analysis? Effective data analysis in support of public policy making is about more than just technical competency. It also requires ethical integrity in analysis and presentation. We conclude with a consideration of the requirements of public servants to use data and statistical analysis ethically.



EVALUATION

Web Safari (10 @ 1%)	10%
Session Labs (10 @ 5%)	50%
Statistical Analysis Assignment Part 1	20%
Statistical Analysis Assignment Part 2	20%

Web Safari (10%): We are going to collectively build a collection of news stories, articles, videos, blog posts, and other web-accessible resources related to the course themes and content. Contribute 10 unique tweets related to the 10 session themes (1% each; for duplicate entries, first in gets the credit. If you find someone submitted your awesome find before you, you can occasionally post a duplicate entry if you add to the reason why it's an interesting article). Also post to the class notes following the table format in the session notes. Due at 9 am on the day of each session to allow for review before that day's session. No late submissions accepted. Only one submission per session note will count.

Session Labs (10 * 5%): Ten labs completed individually demonstrating competency in the concepts discussed in class. Worth 5% each. Due at 1 pm the day of the next session. An answer key (if applicable) will be posted after the lab is due and the lab questions will be discussed in the session on the day they are due (for that reason, late submissions cannot be accepted). Three of the session labs will involve a one-on-one video conversation with the course instructor, and your demonstrated understanding of the course material in that conversation will form part of the basis for your grade on that session lab.

Statistical Analysis Assignment Part 1 (20%): Due October 13 2020 at 1 pm posted to your blog and tweeted. Late submissions = 0.

The University of Regina annually publishes salaries of employees with a total salary of \$100,000 or greater ([as explained here](#)). This data for 2018 and 2019 is available here (to be activated in August 2020).

- (a) **4%** Open and share your own copy of the data file.
 - (i) In Google Sheets, you will need to be signed in to your Google account to select File-->Make a Copy.
 - (ii) Click on the filename "JSGS 807 Assignment 1" at the top-left corner of the screen and rename the file "JSGS 807 Assignment 1 *Your Name*".
 - (iii) Be sure to use appropriate spreadsheet conventions such as row headings, tabs to separate data tables from calculations, cell referencing, etc.
 - (iv) Make sure your copy is shared so that "Anyone with the link can comment" and then click on "Copy Link"
 - Submit the link to your spreadsheets [here](#). (You will continue to work on this spreadsheet until you complete the rest of the steps; it will not be looked at or graded until after the deadline).
- (b) Create a new blog post.



- (i) Give your post an appropriate title. Assume you are writing for an external audience (perhaps writing as a public interest data journalist).
 - (ii) **1%** Describe in a paragraph what the post is about (assume a reader who is unfamiliar with this issue). Include in your post a link to your Google spreadsheet file (the same link you submitted above in step (a.iv)).
- (c) Check the data and the spreadsheet structure.
- Note: the “Total” columns are the sum of the “Salary”, “Administrative / Research Stipend”, and “Market Supplement” columns (these last two are explained in the UofR [Collective Agreement](#) Appendix A and Appendix B, respectively).
 - Note: where salary information exists in one year but not the other, this means the person either was hired or left the University; or their salary was / went below the \$100,000 threshold;
 - Note: there are no formulas in what you downloaded; replicate the total columns, but with formulas to check for errors in the original data.
 - Are there any arithmetic errors in the file you downloaded?
- i) **4%** In your blog post, explain the spreadsheet structure (what the columns are) and use the information in the [Collective Agreement](#) to explain how salaries, stipends, and market supplements work. Also explain how you checked for errors. Whether you found errors or not, identify any concerns you may have about the accuracy of the data. Comment on the number of people not in the 2019 data who were in the 2018 data, and vice versa, and what their salary data says about departures and hires at the University over the two years (accounting for the possibility that their absence or presence in either year may be due to the \$100,000 threshold).
- (d) Perform some basic arithmetic and counting calculations on the data you’ve been given, both within each year and between the two years.
- (i) **3%** In your blog post, write about what you found in your analysis. Include summary statistics useful for understanding the issue of salary expenditures at the University. Identify any interesting anomalies and anything else you think your reader might want to know.
 - (ii) **1%** One measure of the percentage increase in total salaries from 2018 to 2019 is around 9.4%. But another way of expressing the percentage increase in total salaries from 2018 to 2019 is approximately 4.9%. Both of these are correct using the same data provided. How is that possible? And which is the better measure? Address this question in your blog post.
 - (iii) **1%** Is it possible to say that the average salary at the University decreased between 2018 and 2019? What calculation would you report in order to say that? Address this question in your blog post.
- (e) Add columns to the spreadsheet for data that you would like to have for a better understanding about salaries at the University (you don’t need to add the data - just insert the column headings for the information you would like to have if you were asked to analyze this data set in more depth).
- (i) **3%** Describe these “wish list” variables in your blog post and explain why having this data would be interesting for a deeper understanding of the issue of salaries at the University.



- Explain in your blog post whether you think the University should release the data attached to these variables along with the basic salary data, and why that data should be released.
- (f) In your blog post, address these questions **(2%)**:
- (i) Scassa (see p. 402 of Scassa, T. (2014). Privacy and open government. *Future Internet*, 6(2), 397-413. [link](#)) argues that "the transparency and accountability objectives of open government push towards greater disclosure of information in the hands of government, and it is here that there is the greatest tension between privacy and open government".
- When the University of Regina publishes salary information online, do you think this is an appropriate balancing of the publics' right to know and the protection of personal privacy?
 - Is the public interest being served by having this information posted online?
 - Does your opinion change because the information was published online, as opposed to previously when it was available only in the library as a paper copy?
 - The [website](#) where this is posted says "This listing may exclude certain salaries of \$100,000 or greater where such disclosure could threaten the safety of an individual." Is this a reasonable balancing of the privacy / openness issues? Are there any drawbacks to such a policy where an individual can ask that their salary information be protected?
- (g) **1%** After the submission deadline of October 13 at 1 pm, tweet your blog link tagging anyone who you think might be interested and use #JSGS807 and #opendata.

Statistical Analysis Assignment Part 2 (20%): Due December 15 at 1 pm. Late submissions = 0.

- Create an infographic about expenditures, services, and activities at the UofR based on your data analysis in assignment part 1 and other information of your choosing from the [UofR Office of Institutional Research](#).
- Assume you are creating this infographic for the Minister of Education supplemental to a briefing note in order to support decision making (you do **not** need to provide a briefing note in addition to the infographic).
- You can use a visualization product like Flourish (supplemented with [another data visualization software service](#) if you want), but focus on the one page infographic (use something simple like PowerPoint or other tool to create the final infographic). Marks will be allocated based on:
 - The articulation of the policy problem and identification of the data sources used to investigate the problem **(2%)**,
 - Your use of the software to create data visualizations derived from online data sources **(8%)**,
 - The creativity and depth of your analysis **(5%)**,
 - The clarity and "readability" of your infographic **(5%)**.
- Submit a pdf of your infographic [here](#).
 - You are welcome to put your infographic into a blog post and tweet the link to your blog (with the hashtags #JSGS807 and #infoviz) but this is not a requirement.


STUDENT RESOURCES

UNIVERSITY OF REGINA	UNIVERSITY OF SASKATCHEWAN
<p>Students with Special Needs Any student with a disability, injury or illness who feels they may need academic accommodation should discuss this with the course instructor after contacting the Centre for Student Accessibility at 306-585-4631</p>	<p>Students with Special Needs Students in this course who, because of a disability, may have a need for accommodations are encouraged to discuss this need with the instructor and to contact Access and Equity Services at (306) 966-7273</p>
<p>Students Experiencing Stress Students in this course who are experiencing stress can seek assistance from Counselling Services, or call (306) 585-4491. Additional information is available here.</p>	<p>Students Experiencing Stress Students in this course who are experiencing stress can seek assistance from the Student Wellness Centre or call (306) 966-4920.</p>
<p>Other Supports for Students</p> <ul style="list-style-type: none"> • Sexual Violence Prevention and Response • UofR Student Success Centre • ta-tawâw Student Centre • UR International • JSGS (UofR) Student Handbook • Being a Not-Rich Student 	<p>Other Supports for Students</p> <ul style="list-style-type: none"> • UofS Sexual Assault Information • UofS Student Learning Services • UofS Aboriginal Students' Centre • UofS International Student Centre • JSGS (UofS) Student Handbook • Being a Not-Rich Student
<p>Academic Integrity and Conduct Ensuring that you understand and follow the principles of academic integrity and conduct as set out by the University of Regina is vital to your success in graduate school. Ensuring that your work is your own and reflects both your own ideas and those of others incorporated in your work is important: ensuring that you acknowledge the ideas, words, and phrases of others that you use is a vital part of the scholarly endeavour. If you have any questions at all about academic integrity in general or about specific issues, contact your course instructor to discuss your questions.</p>	<p>Academic Integrity and Conduct Ensuring that you understand and follow the Regulations on Student Academic Misconduct is vital to your success in graduate school. Ensuring that your work is your own and reflects both your own ideas and those of others incorporated in your work is important: ensuring that you acknowledge the ideas, words, and phrases of others that you use is a vital part of the scholarly endeavour. If you have any questions at all about academic integrity in general or about specific issues, contact any faculty member and we can discuss your questions.</p>



See this note about grading: <https://jlphd.wordpress.com/2018/02/04/a-note-about-grading/>

Course Materials: All readings are accessible through the UofR / UofS library systems,¹ or free online. Since you already pay tuition, and part of your tuition goes to funding the materials available through the library, it does not seem fair to ask you to pay extra for a textbook. I've also not found a decent textbook for this course. However ...

Some free online textbooks you may find useful:

- Bounegru, L., Chambers, L., Gray, J. (2018). [The Data Journalism Handbook](#). 2nd edition. EJC.
- Black, Ken. (2013). [Business Statistics: For Contemporary Decision Making, 8th Edition](#). John Wiley & Sons
- Mahbobi, M., Tiemann, T. (2010). [Introductory Business Statistics with Interactive Spreadsheets](#). BCcampus.
- Diez, D.M., Barr, C.D., Cetinkaya-Rundel, M. (2015). [OpenIntro Statistics](#). Open Intro.
- Illowsky, B., Dean, S. et al. (2016). [Introductory Statistics](#). OpenStax.
- Shafer, D., Zhang, Z. (2012). [Introductory Statistics](#). Saylor Foundation.

Technology Requirements:

- To actively participate in live video conferencing, a web camera and microphone are required.
- Note that some governments (particularly the Government of Saskatchewan) have restricted access to Zoom meetings using corporate equipment. If you are using a computer provided to you by an organization that restricts access to Zoom, you will likely not be able to use that computer to engage in any Zoom sessions.
- If you do not have access to a computer that can access Zoom sessions, alternative options are available.
- To participate asynchronously, a standard Internet-connected device and web browser will be sufficient.
- All software used in this course is available for use for free from a standard current computer configuration (e.g., laptop computer) with access to the Internet.
- While Excel is the standard spreadsheet software in most organizations, we use [Google Sheets](#) when working with spreadsheets in this class as the operation of Sheets is similar to Excel. This also avoids problems with file transfers. While you will not be learning the particular features of Excel, learning how spreadsheets can be used in the sourcing, organizing, analysis, and presentation of data is transferable between programs.
- Some assignments may require the downloading and installing of free programs, which requires a computer that you have administrator privileges on (n.b., most workplace computers will not let you install software). There is no penalty for not being able to install this software, and alternative assignments will be available.
- Students who are unable to participate due to a technology barrier should contact the instructor as soon as possible to discuss alternative arrangements.

Non-registered Students: This course was developed and deployed for the benefit of graduate students registered in the [Johnson Shoyama Graduate School of Public Policy \(JSGS\)](#). It is made freely available as a learning resource for anyone interested in the course content. However, access to the JSGS student community, the resources of either university, or the course instructor is only available to formally registered students.



[Syllabus - Statistics for Public Sector Managers](#) by Justin Longo is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#). Permissions beyond the scope of this license may be available by contacting the instructor at <https://www.digitalgovernancelab.org/>

¹ Links are active for on-campus use at the University of Regina. If you are a UofR student connecting from off-campus, [use these instructions](#). If you are a UofS student connecting from off-campus, [use these instructions](#).