

1. Name (first and last)

Text Response

Daniel Liden

Statistic

Total Responses	Value
1	

2. Email

Text Response

djliden91@gmail.com

Statistic

Total Responses	Value
1	

3. Contact Phone

Text Response

6162956108

Statistic

Total Responses	Value
1	

4. In Fall 2018 you will consider yourself to be a:

#	Answer	Bar	Response	%
1	VT Freshmen Undergraduate Student		0	0%
2	VT Sophomore Undergraduate Student		0	0%
3	VT Junior Undergraduate Student		0	0%
4	VT Senior Undergraduate Student		0	0%
5	Between Undergrad and Graduate school		0	0%
6	Graduate school		1	100%
	Total		1	

Statistic	Value
Min Value	6
Max Value	6
Mean	6.00
Variance	0.00
Standard Deviation	0.00
Total Responses	1

5. What degree(s) and major(s) are you pursuing along with institution?

Text Response
Statistics MS University of Minnesota twin cities

Statistic	Value
Total Responses	1

6. List any Minors you are pursuing.

Text Response

N/A

Statistic

Total Responses

Value

1

7. List any honors and/or awards received:

Text Response

Undergraduate -- graduated with general honors; on Dean's List all four years

Statistic

Total Responses

Value

1

8. Currently we have 1 program accepting applications which includes:

#	Answer			Total Responses
15	Data Science for the Public Good		0	0
	Total		0	-

Statistic

Min Value

Max Value

Mean

Variance

Standard Deviation

Total Responses

Data Science for the Public Good

9. Essay (up to 500 words): "What do you want to get out of this experience?"

Text Response

In the spring of 2011, I was taking a sociology course taught by James Evans at the University of Chicago. He canceled a week and a half of classes that semester because, we were later told, he was in Washington, D.C., arguing for the science budget. At the time, I had not seen many academics actively engaged in policy work or discussions. Since then, I have been on the lookout for cases where academic work and public good intersect. That has informed much of my subsequent thought about my professional and academic activities since then. In that vein, there are three main things I want to get out of this program: new skills and experiences, community and peer engagement, and broader perspective on applying statistical work to social good. I decided to study statistics in large part because of the potential to apply it to social issues. The past projects from the DSPG program reflect this motivation well. For example, one of the projects presented initial work on investigating the impact of refugees in Virginia: an extremely important project, especially given the amount of misinformation and hearsay about the impact of refugees in this country. Two other projects discussed the impact of open source software. Much of my early education in statistics was made possible by access to open-source software, so exploring the impact and importance of access to such software strikes me as an excellent project whose results could ultimately address inequities in educational opportunities. I am very excited for the opportunity to work on similar projects with the possibility of direct and consequential impact on communities. I am eager to meet others who want to put their data science skills to use for public good. Most of my colleagues and classmates thus far have been primarily interested in using their skills to make money and advance their careers, and I am thrilled at the prospect of meeting a self-selecting group of people who are seriously interested in applying their skillsets to the public good rather than personal enrichment. Finally, I am interested in gaining a broader perspective on future career and educational opportunities focusing primarily on working toward the public good. This program isn't just a bullet point on a resume for me – this is the type of work to which I want to devote my career. The pathway to such a career isn't as clear as, for example, the path to a career in business. I am hoping the program mentors and peers, as well as government and industry partners, can offer some guidance and perspective about what such a career looks like, how to get there, and where to find other organizations whose missions primarily align with the public good. I am also excited for the possibility to share my own skills and experiences with others in the program. I believe my background has given me a unique and valuable perspective on data analysis and on the communication of statistical results.

Statistic

Total Responses

Value

1

10. Essay (up to 500 words): "Please describe any previous research experience and/or work experience you may have."

Text Response

Current Research My current research project focuses on tests of heterogeneity in risk difference in clinical trials. Suppose we are interested in mortality as a trial outcome, and we find that mortality is lower in the treatment group than in the control group. We may also be interested in whether age modifies the treatment effect: do older patients respond to the treatment as well as younger patients? Typically, one would test this using a Poisson regression model with an identity link function including an interaction term between age and treatment group. Unfortunately, in many cases, such models fail to converge. We can address this by discretizing the patient covariate of interest (e.g., age) and generating bootstrap p-values using Cochran's Q test statistic. This approach has its own limitations: the power and type-I error of the test depends greatly on how the covariate of interest is discretized. For example, we will obtain different results if we compare ages 30-50 to ages 51-80 than if we compare ages 30-40 to ages 41-80. One of the objectives of the project, then, is a simulation study showing how choice of "cutpoint" -- i.e., how the covariate is discretized -- impacts the test results. The final objective of the project is to issue guidance on optimal selection of cutpoints or, if possible, to develop a test that does not require discretization.

Work Experience For the past six years, I have worked for a pharmaceutical data analytics company called Context Matters. In the beginning, I helped to develop and maintain a relational database of drug reimbursement data collected from countries with national healthcare systems. In such healthcare systems, decisions on whether to fund drugs are made on the basis of efficacy and/or cost-effectiveness data. In the four years, my role has primarily involved statistical/analytical consulting. I meet with clients to discuss their analytical needs and work with a team to complete analysis projects using the data in our database. Many of my research projects for Context Matters have been published as abstracts in the journals Value in Health and the International Journal of Technology Assessment in Health Care. I gave a short talk at the International Society for Pharmaceutical Outcomes Research conference in Montreal in 2014. Most recently, two of my research projects were presented as talks at the Health Technology Assessment International conference in Rome. These most recent presentations focused on the reimbursement of oncology drugs in the UK. The 2014 project was an analysis of how provincial drug reimbursement decisions reflected national funding recommendations in Canada. Other Research In the summer of 2010, I worked in a pancreatic cancer lab at the Van Andel Research Institute. I primarily worked on a bioinformatics project using publicly-available microarray data to investigate sugar-binding proteins. The goal of this analysis was to obtain fine-specificity data from glycan-array data. The project was successful and we were able to identify new binding patterns using a novel approach dubbed "outlier-motif analysis." The results of the analysis were published in Glycobiology in 2012.

Statistic

Total Responses

Value

1

11. Essay (up to 500 words): "Please describe your background (courses taken, research projects, etc.) in statistics and mathematics."

Text Response

Statistics I am currently finishing my final semester of my statistics MS program. I have completed a year-long graduate-level mathematical statistics sequence in which I worked through Casella and Berger's Statistical Inference from start to finish. This sequence covered standard topics such as basic probability, useful distributions, joint and marginal distributions, random samples, data reduction, point estimation, hypothesis testing, asymptotic statistics, ANOVA, regression, and much more. I also completed a year-long graduate-level applied statistics sequence. The first semester focused on advanced regression and focused on the theory and application of linear regression models, GLMs, information criteria, ridge and lasso regression methods, cross-validation, and various machine learning techniques. The course concluded with a project wherein I used a gradient tree-boosted model to predict the frequency and magnitude of insurance claims. The second semester of this sequence covered experimental design, random- and mixed-effects models, time-series data analysis, and correlated data analysis. In the second year of the program, I took a statistical computing course that covered Monte-Carlo methods and simulation studies using R. It also covered a substantial amount of material on solving optimization problems in statistics. I also took a categorical data analysis course that covered the theory and application of GLMs in considerable detail. Both of these courses expanded my exposure to multivariate analysis through multivariate normal and multinomial models. In my final semester, I will take graduate-level courses on Bayesian data analysis and on statistical learning and data mining. The first course is intended as an introduction to Bayesian data analysis and a comparison of Bayesian to frequentist methods. The second will work through much of the material in Hastie, Tibshirani, and Friedman's The Elements of Statistical Learning, and will cover topics such as high-dimensional learning, graphical models and network analysis, recommender systems, big data, and unstructured data. Mathematics As an undergraduate, I took a year of calculus. In preparation for graduate school, I worked through two of Gilbert Strang's books, Calculus and Introduction to Linear Algebra. In graduate school, I took an applied linear algebra course and honors introduction to analysis. I also independently worked through parts of Tom Apostol's Calculus and Mathematical Analysis books. Research Please see the above question about research experience for details of my statistical research project about tests of heterogeneity for risk difference.

Statistic

Total Responses

Value

1

12. Essay (up to 500 words): "Please describe your background (courses taken, research projects, etc.) in social and behavioral sciences."

Text Response

My undergraduate major was the History, Philosophy, and Social Studies of Science and Medicine. I completed the following courses: Medicine since the Renaissance, Modern Science (Primarily focused on sociology of science), A course on the scientific revolution, a seminar covering classics such as Kuhn's Structure of Scientific Revolutions and Latour's Science in Action, a graduate course on the History and Philosophy of Biology, a graduate seminar titled Sciences of Mind and the Moving Image, a seminar on Science and Religion, a graduate course on Darwin's Origin of Species and Descent of Man, a seminar titled The Senses in Historical Context, and a graduate seminar Resuscitation in America from 1800 to the present. Through this program, I learned to think of modern science not as the inevitable result of an inexorable march of progress, but as a highly contingent and disunited subject with individual fields characterized as much by tacitly-learned and accepted work practices as by underlying shared principles. This field of study also sparked my interest in statistics: if the modern sciences can be said to have a shared language, that language is statistics. This became clear to me as I visited works such as Stephen Stigler's The History of Statistics, Ian Hacking's The Emergence of Probability, and Theodore Porter's Trust in Numbers. In addition to learning valuable research techniques in history, sociology, and anthropology, this program deeply informed the ways in which I think about statistics and data analysis. Learning to write about science has made me very conscious of how statistical results are communicated. How statistical results are used is at least as important as how the analyses are conducted – it is just as easy, intentionally or otherwise, to mislead with statistics as it is to bring facts to light. Responsible use of statistics is, therefore, something I think a lot about and take very seriously. Research I wrote my undergraduate thesis on classification manuals and their effects on knowledge generation in scientific work practice. I argued that "Practice based on a formal information infrastructure reifies the classifications comprising it, resulting in the appearance of fixity and the generation of 'the natural.'" That is, a classification guide is not just a dusty reference manual, but a prescriptive device that both guides and is reinforced by research practice. I included a case study about Rickettsial agents, which inhabited a vague taxonomic position between bacteria and viruses during the 1950s-60s. In spite of this, their position taxonomic guides at that time were expressed as fixed and certain, without reference to the surrounding controversy. From 2013 to 2016, I worked with Dr. Daniel Brauner at the University of Chicago on a project about the history of resuscitation and of the concept of cardiac arrest. The project started from the 16th-17th centuries when the role of the heart was still poorly, and continued to the present when some attempt at resuscitation is attempted on almost all patients who die in a hospital without a "Do Not Resuscitate" order.

Statistic

Total Responses

Value

1

13. Essay (up to 500 words): "Please describe your background in programming."

Text Response

Most of my programming experience to date involves the R Statistical Programming Language. Prior to graduate school, I took a number of online courses to learn R. The most notable were offered on Coursera by Johns Hopkins University, and covered topics such as reproducible research, exploratory data analysis, statistical inference, and general R programming. Through edX, I took a few courses offered by Harvard University on statistics (particularly biostatistics) and R. During that time, I also took an online Coursera course by the University of Toronto on Python programming. My MS program has used R almost exclusively. In my applied statistics sequence, I used R for linear models, GLMs, and a variety of machine learning techniques. I also used it for designing experiments and investigating the results of designed experiments and for developing random- and mixed-effects models. The course on statistical programming was the most focused and in-depth R course I took. Whenever possible, we avoided built-in functions in favor of learning to write our own code to accomplish various statistical tasks. For example, in learning Monte-Carlo methods, we only used the random uniform generating function to generate random draws from other distributions using inversion methods and rejection algorithms. A major focus of this course was learning how to develop algorithms into usable code. We were often given optimization algorithms with relevant mathematical notation and asked to write functions implementing the algorithms. As a result of this course, I am very comfortable using R as a programming language, not just as a glorified statistics calculator. In the statistical programming course and in my own research (see prior responses about research), I have done a considerable amount of work with simulation studies. This has proven to be a very useful toolset for exploring the behavior of statistical tests and comparing the power and error of different statistical procedures. It has also forced me to be mindful of efficiency in my programming practice. In my research, for instance, generating even a single test statistic took the better part of a minute, so a full simulation study was, at first, prohibitively computationally expensive. To address this, I learned to use some of the parallel computing packages in R. I have re-written many functions to take advantage of parallel computing and to avoid costly procedures such as loops. Because of this, I have gained an appreciation for the "nuts and bolts" of the R language, and can comfortably implement my own functions and procedures when the built-in options are not sufficient. I have used Linux regularly since high school and am very comfortable in a command line environment. For specific projects, I have been able to quickly learn the basics of other programming languages including Python and SAS. I have also spent some time learning C and FORTRAN. I also used Ruby in my work to develop a program to implement my company's pricing model. I am comfortable enough with general programming practices to pick up new languages as needed for a given application.

Statistic

Total Responses

Value

1

14. Essay (up to 500 words): "Please provide information about other significant courses you have taken within your field of study."

Text Response

Most of the important courses I've taken were covered above. I have additionally completed two graduate-level courses in epidemiology (one at the University of Chicago and one at the University of Minnesota). These provided useful real-world examples of how statistical procedures are used and how data are collected. They also provided a useful opportunity to think about the ethics involved in experimental design decisions and data collection and use. I also completed a course on genomics, in which we used Perl to analyze genomic data. I am passionate about self-study. As mentioned above, I have completed numerous online courses on R, statistics, programming, and Linux system administration. I have independently worked through several programming, statistics, and mathematics books. I am currently working through a book on linear programming (Bertsimas's Introduction to Linear Optimization), and will move on to Bertsekas's Nonlinear Programming next. Self-study has always been important to me. Formal coursework raises as many questions as it answers, and I am always on the lookout for materials to expand on and fill in the gaps in the materials covered in my coursework. I bring this same motivation to my professional work, and regularly consult the academic literature and textbooks concerning new projects and new subject areas I take on at work.

Statistic

Total Responses

Value

1

15. Please list the name and contact information for 2 references (teachers, mentors, or employers) that we will contact for a letter of reference/brief survey. Please make sure you list the correct email and they know we will be contacting them soon (within the next week). Only 2 references will be contacted; do not list more than 2.

Text Response

Professor Birgit Grund birgit@umn.edu ; Professor Adam Rothman arothman@umn.edu

Statistic

Total Responses

Value

1