

Victoria Stodden
Associate Professor
Department of Industrial and Systems Engineering
University of Southern California
http://stodden.net

Education History

Ph.D.	Sept 2006	Statistics, Stanford University
M.L.S.	Dec 2007	Stanford Law School
M.S.	June 2000	Statistics, Stanford University
M.S.	July 1996	Economics, University of British Columbia, Canada
B.Soc.Sci.	Dec 1994	Economics (<i>magna cum laude</i>), University of Ottawa, Canada

Employment History

<i>Jan 21 -</i>	Associate Professor, Department of Industrial and Systems Engineering, Viterbi School of Engineering, University of Southern California
<i>Aug 14 - Jan 21</i>	Associate Professor, School of Information Sciences, University of Illinois Urbana-Champaign, with courtesy appointments in the College of Law and the Departments of Statistics and Computer Science Faculty Affiliate, National Center for Supercomputing Applications
<i>2015 -</i>	Faculty Affiliate, Meta-Research Innovation Center (METRICS), Stanford University
<i>2014 -</i>	Affiliate Scholar, The Center for Internet and Society, Stanford Law School
<i>Jul 10 - Jul 14</i>	Assistant Professor, Department of Statistics, Columbia University, with Affiliate appointment in the Institute for Data Sciences and Engineering
<i>Aug 09 - Jul 10</i>	Kauffman Fellow in Law and Innovation, Yale Law School
<i>Jan 09 - Jul 09</i>	Postdoctoral Researcher, MIT Sloan School of Management
<i>Jan 08 - Dec 09</i>	Fellow, Berkman Center for Internet & Society, Harvard Law School

Selected Presentations, Honors, Recognitions, and Outstanding Achievements

<i>Aug 2022</i>	Committee of Visitors (COV) for the Office of Advanced Cyberinfrastructure (OAC), Directorate for Computer and Information Sciences and Engineering, National Science Foundation.
<i>Mar 2021</i>	Building a Community Roadmap to Robust Science in High-Throughput Applications, SIAM Conference on Computational Science and Engineering (CSE21). “ <i>Advancing Computational Scientific Discovery by Enabling Reproducibility and Transparency: Policies and Practice.</i> ”
<i>Dec 2020</i>	Invited Address , Scholarship for a Post-Pandemic World: A Conversation; Council of Graduate Schools 60th Annual Meeting. “ <i>Training and Scholarly Impact in a Digital World.</i> ”
<i>Sep 2020</i>	Epstein Institute Seminar, University of Southern California. “ <i>Two Projects for Advancing Scientific Reliability in Complex Computational and Human Systems.</i> ”
<i>July 2020</i>	Keynote , Applied Human Factors and Ergonomics Conference: The Human-Side of Service Engineering. “ <i>Toward a Computable Scholarly Record: Meta Science and Engineering Reproducibility in the Era of AI</i> ”

- June 2020* **Invited Address**, IEEE CS Ad Hoc Committee on Open Science and Reproducibility. “*Reproducibility and Replicability in Science*”
- 2016-* The National Academy of Engineering: Advisory Group for the Online Resource Center for Ethics Education in Engineering and Science; and Editorial Boards on Social and Behavioral Sciences and Research Ethics.
- 2017 - 2019* The National Academies of Science, Engineering, and Medicine committee: “Reproducibility and Replicability in Science”
- 2016 - 2019* The National Academies of Science, Engineering, and Medicine committee: “Roundtable on Data Science Post-Secondary Education”
- Nov 2019* Northwestern Computer Science **Distinguished Lecture** Series, Northwestern University. “*The Lifecycle of Data Science: A Framework for Advancing Computational and Data-enabled Research*”
- Nov 2019* Center for Data and Computing **Distinguished Speaker** Series, University of Chicago. “*Reproducibility is Not a Crisis. Now What? Next Steps for Advancing Computational and Data-enabled Science.*”
- Sept 2019* **Keynote**, Parallel Computing 2019: Symposium Tools and Infrastructure for Reproducibility in Data-Intensive Applications, Prague, Czech Republic. “*Advancing Reproducibility and Transparency in Data Inference Applications*”
- July 2019* **Invited Seminar**, Space Telescope Science Institute, Baltimore, MD. “*Reproducibility in Scientific Inference, Data Dissemination, and Computational Environments*”
- 2014 - 2018* The National Academies of Science, Engineering, and Medicine committee: “Responsible Science: Ensuring the Integrity of the Research Process”
- Nov 2018* **Keynote**, Computational Reproducibility at Exascale 2018 (CRE2018), Supercomputing18, Dallas, TX. “*Reproducibility in Computational and Data-enabled Science*”
- Nov 2018* **Plenary**, Ethics, law, and transparency. “Institute for the Secure Sharing of Online Data” (ISSOD) Workshop, Boston, MA. “*A Future of Research Transparency: Enabling Reproducibility in Repository Design*”
- June 2018* **Keynote**, The 27th International Symposium on High-Performance Parallel and Distributed Computing, Arizona State University. “*Reproducibility in Computational and Data-enabled Science*”
- June 2018* **Keynote**, IEEE Data Science Workshop, EPFL, Switzerland. “*Reproducibility and Generalizability in Data-enabled Discovery*”
- Apr 2018* **Keynote**, Northwestern Computational Research Day, Evanston, IL. “*Reproducibility in Computational Research: Code, Data, Statistics, and Implementation*”

- Mar 2018 **Invited Presentation**, Sandia National Laboratories, Oak Ridge National Laboratory, Swiss Institute of Technology (SOS) SOS-22 Workshop: HPC and Data Science, Waikoloa, Hawaii. “*Reproducibility at Exascale*”
- Aug 2017 European Alpbach Forum, Technology Symposium, Austria. “*Knowledge and Understanding in the Age of Data*”
- Sep 2017 **Keynote**, Research Data Management Implementations (RDMI) Workshop, Arlington, VA. “*Research Data Management Implementations: Towards the Reproducibility of Science*”
- Mar 2017 **The Judith Resnik Year of Women in ECE Seminar**, Carnegie Mellon University, Pittsburgh, PA. “*Reproducibility in Computationally-enabled Research*”
- Mar 2017 National Academy of Sciences, Arthur M. Sackler Colloquia, Reproducibility of Research: Issues and Proposed Remedies, Washington D.C. “*Reproducibility in Computationally-enabled Research*”
- Nov 2016 Seance de reflexion of the Swiss National Research Council on “2050: A Science Odyssey,” Interlaken, Switzerland. “*Science: Set the Default to Open*”
- July 2016 **Invited Talk**, SIAM Annual Meeting, Boston MA. “*Implementing Reproducibility in Computational Science*”
- June 2016 **Keynote**, Maria de Maeztu Annual Event: Data-driven Knowledge Extraction Workshop, Universitat Pompeu Fabra, Barcelona, Spain. “*Reproducibility in Computational Research*”
- Apr 2016 **Keynote**, Coalition for Networked Information Annual Meeting, San Antonio, TX. “*Defining the Scholarly Record for Computational Research*”
- Feb 2016 **Keynote**, AAAS Reproducibility Workshop: Modeling and Code, Washington, D.C. “*Software in Science*”
- Nov 2015 **Plenary**, SuperComputing15, Austin, TX. “*Reproducibility in High Performance Computing*”
- Oct 2015 **Keynote**, Open Access Week, Virginia Tech, Blacksburg, VA. “*Scholarly Communication in the Era of Big Data and Big Computation*”
- Feb 2015 **Invited Talk**, AAAS Annual Meeting, San Jose, CA. “*Integrity, Reproducibility, and the Changing Technological Environment for Research*”
- Dec 2014 **Keynote**, Berkeley Initiative for Transparency in the Social Sciences Conference, University of California, Berkeley. “*Framing Transparency in Research: Issues and Opportunities*”
- Nov 2014 **Keynote**, OpenCon 2014, American University Washington College of Law, Washington, D.C. “*Open Data and Reproducibility in Research*”

- Mar 2014 **Keynote**, Computational and Simulation Sciences and eResearch, Annual Conference, Melbourne, Australia. *“Open Data and Reproducibility in Research”*
- July 2013 **Plenary Speaker**, Open Repositories 2013, Charlottetown, Prince Edward Island, Canada. *“Re-use and Reproducibility: Opportunities and Challenges”*
- Mar 2013 Testified before The Congressional House Committee on Science, Space and Technology hearing on Scientific Integrity & Transparency, Washington, D.C.
- June 2012 **Keynote**, PDE Software Frameworks, Muenster, Germany. *“Reproducible Results: Challenges for Computational Science and the Scientific Method”*
- June 2012 **Keynote**, 74th EAGE Conference & Exhibition: Open-source E+P Software - Six Years Later, Copenhagen, Denmark. *“The Central Role of Geophysics in the Reproducible Research Movement”*
- Feb 2012 **Dean’s Lecture**, UC Berkeley School of Information, Berkeley, CA. *“The Credibility Crisis in Computational Science: An Information Issue”*
- Dec 2011 **Plenary Keynote**, Cyberinfrastructure Days, University of Michigan. *“The Credibility Crisis in Computational Science: A Call to Action”*
- Nov 2011 National Science Foundation, Advisory Committee on Cyberinfrastructure, Washington D.C. *“Report on Journal Policy and Reproducible Computational Research”*
- Oct 2011 **Keynote**, Open Science Summit, Computer History Museum, Mountain View, CA. *“Transparency in Scientific Discovery: Innovation and Knowledge Dissemination”*
- Mar 2011 National Science Board Expert Panel Discussion on Data Policy, Washington, D.C. *“Scientific Reproducibility: First Steps and Guiding Questions”*
- June 2010 **Keynote**, ICML Workshop on Machine Learning Open Source Software, Haifa, Israel. *“Reproducible Research in Computational Science: Problems and Solutions For Data and Code Sharing”*
- May 2010 **Dean’s Lecture**, UC Berkeley School of Information. *“The Digitization of Science and the Degradation of the Scientific Method”*
- Apr 2010 The New Biology: Pathways to Convergence in the Life Sciences, MIT/Kauffman Seminar for Senior Congressional and Executive Branch Staff, MIT. *“Innovation and Openness in Science and the Exceptional Role of the Biological Sciences”*
- Apr 2010 **Invited Chaired Session** on Reproducibility: New England Statistics Symposium, Harvard University. *“Scientific Integrity and Reproducibility: Data and Code Sharing”*
- Apr 2010 **Invited Lecture**, UC Berkeley School of Information. *“Open Licensing and Scientific Reproducibility”*
- Feb 2009 **Neyman Seminar**, Department of Statistics, UC Berkeley. *“The Reproducible Research Standard: Legal Barriers to Practicing the Scientific Method in Computational Research”*

- May 2009* **Keynote**, Scientific Software Days, Texas Advanced Computing Center, University of Texas at Austin. “*The Impact of Computational Science on the Scientific Method*”
- Aug 2008* *Kaltura Access to Knowledge Award* for outstanding scholarship in support of Open Access, for “*Enabling Reproducible Research: Open Licensing For Scientific Innovation*”
- June 2003* Stanford Centennial Teaching Award, Stanford University
- June 2003* Statistics Departmental Teaching Award, Stanford University
- June 2002* Statistics Departmental Teaching Award, Stanford University
- 1993 - 94* Merit Scholarship, Association of Professors of the University of Ottawa
- 1992 - 93* Merit Scholarship, University of Ottawa
- 1991 - 94* Dean’s Honor List for Outstanding Academic Performance, University of Ottawa

Research Grants

- Oct 20 - Sep 22* “Collaborative Research: PPOSS: Planning: Performance Scalability, Trust, and Reproducibility: A Community Roadmap to Robust Science in High-throughput Applications,” PI: Victoria Stodden \$30,000. Part of collaborative award lead by UTK (Entire award \$150,000). National Science Foundation #2028881
- Mar 16 - Feb 22* Supplement to “CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale,” PI: Bertram Ludascher (Illinois - lead institution), co-PIs: Victoria Stodden (Illinois), Niall Gaffney (UT Austin), Matthew Turk (Illinois), Kyle Chard (UChicago). Supplement award: \$304,774 with Illinois share \$54,774 (subawards to University of Chicago (\$150,000) and USC (\$100,000)). National Science Foundation #1541450
- Sep 19 - Aug 22* “EAGER: Reproducibility and Cyberinfrastructure for Computational and Data-Enabled Science,” PI: Victoria Stodden. co-PI: Michela Taufer (UTK), \$300,000 (\$149,997 to UTK). National Science Foundation #1941443
- Aug 18 - July 22* “EAGER: Preserve/Destroy Decisions for Simulation Data in Computational Physics and Beyond,” PI: Victoria Stodden. co-PI: Darko Marinov (Illinois), \$300,000. National Science Foundation #1839010
- Jun 18 - Aug 18* Discovery Partners Institute, PI: Victoria Stodden, \$3,000. University of Illinois Economic Development and Innovation (VPEDI) International Travel Grant
- Sep 16 - Aug 19* “EAGER: Collaborative Proposal: Supporting Public Access to Supplemental Scholarly Products Generated from Grant Funded Research,” PI: Victoria Stodden. Collaborative proposal with Illinois share: \$59,991. National Science Foundation #1649555

- Mar 16 - Feb 21* “CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale,” PI: Bertram Ludaescher (Illinois - lead institution), co-PIs: Victoria Stodden (Illinois), Niall Gaffney (UT Austin), Matthew Turk (Illinois), Kyle Chard (UChicago). Co-operative proposal \$5,887,240. National Science Foundation #1541450
- Jul 19 - Feb 21* “CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale,” PI: Bertram Ludaescher (Illinois - lead institution), co-PIs: Victoria Stodden, Niall Gaffney (UT Austin), Matthew Turk (Illinois), Kyle Chard (UChicago), Jarek Nabryzski (Notre Dame), Matthew B. Jones (UCSB). Supplement \$293,559. National Science Foundation #1541450
- Jul 12 - Nov 14* “Facilitating Transparency in Scientific Publishing,” PI: Victoria Stodden. \$420,640. Alfred P. Sloan Foundation
- Sep 11 - Dec 13* “EAGER: Policy Design for Reproducibility and Data Sharing in Computational Science,” PI: Victoria Stodden, \$168,800. National Science Foundation #1153384
- Jul 11* “Community Forum on Reproducible Research Policies,” PI: Victoria Stodden. \$3,800. Alfred P. Sloan Foundation

Publications

DOCTORAL THESIS

Model Selection When the Number of Variables Exceeds the Number of Observations [1]
Stanford University Department of Statistics, 2006.

With ever more data being collected, stored, and analyzed, investigators are increasingly finding themselves in situations where the number of predictors in their model far outweighs the number of observations. As is well-known, standard regression modeling techniques break down in this setting. Typically in these cases researchers use a model selection procedure (Forward Stepwise, Lasso, or LARS for example). My thesis investigated how these model selection techniques perform when the number of predictors is much greater than the number of observations, compared to the “optimal” model i.e. the one chosen by All-Subsets Regression. Under appropriate conditions, Basis Pursuit (a Lasso-like method from the field of Signal Processing) finds the optimal model in this under-determined setting. Building on this result, my thesis discovered conditions under which Forward Stepwise, Lasso, and LARS recover the optimal model. A Matlab toolbox for collaborative reproducible research, SparseLab, is introduced, as part of providing the research community with open source tools for model selection. Along with other reproducible results in the field of sparse regression, this toolbox was made available at <http://sparselab.stanford.edu>, creating an active repository for reproducible research in the field.

REFEREED JOURNAL ARTICLES

- I. Ur Rahman, I. Drori, V. Stodden, D. Donoho, P. Schroeder, “*Multiscale Representations of Manifold-valued Data*,” SIAM J. Multiscale Modeling and Simulation, 4, 2005, p. 1201-1232. DOI:10.1137/050622729 [2]
- E. Hurowitz, I. Drori, V. Stodden, D. Donoho, P. Brown, “*Virtual Northern Analysis of the Human Genome*,” PLOS ONE, May 23, 2(5), 2007. DOI:10.1371/journal.pone.0000460 [3]
- D. Donoho, A. Maleki, I. Rahman, M. Shahram, and V. Stodden, “*Reproducible Research in Computational Harmonic Analysis*,” IEEE Computing in Science and Engineering, 11(1), January 2009, p. 8-18. DOI:10.1109/MCSE.2009.15 [4]
- V. Stodden, “*The Legal Framework for Reproducible Research in the Sciences: Licensing and Copyright*,” IEEE Computing in Science and Engineering, 11(1), January 2009, p. 35-40. DOI:10.1109/MCSE.2009.19 [5]
- V. Stodden, “*Enabling Reproducible Research: Open Licensing For Scientific Innovation*,” International Journal of Communications, Law and Policy, Issue 13, Winter 2008-09, p. 1-25. **Winner of the Access to Knowledge Kaltura prize.** [6]
- V. Stodden with the Toronto International Data Release Workshop Authors, “*Prepublication Data Sharing*,” Nature, 461(10), September 2009, p. 168-70. DOI:10.1038/461168a [7]
- V. Stodden, “*Open Science: Policy Implications for the Evolving Phenomenon of User-Led Scientific Innovation*,” Journal of Science Communication, 9(1), March 2010. DOI:10.22323/2.09010205 [8]
- V. Stodden with Yale Roundtable Participants, “*Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science*,” IEEE Computing in Science and Engineering, vol. 12, no. 5, p. 8-13, Sep./Oct. 2010 [9]
- R. LeVeque, I. Mitchell, V. Stodden, “*Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture*,” IEEE Computing in Science and Engineering, Vol. 14, Issue 4, 2012, p. 13-17. DOI:10.1109/MCSE.2012.38 [10]
- V. Stodden, P. Guo, Z. Ma, “*Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals*,” PLOS ONE 8(6), 2013. DOI:10.1371/journal.pone.0067111 [11]
Chosen for inclusion in the PLOS Open Data Collection.
- V. Stodden, S. Miguez, “*Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research*,” Journal of Open Research Software, 2(1):e21, 2014, p. 1-6. DOI:10.5334/jors.ay [12]
- V. Stodden, “*The Reproducible Research Movement in Statistics*,” Statistical Journal of the International Association of Official Statistics, 30(2), 2014, p. 91-93. DOI:10.3233/SJI-140818 [13]

- V. Stodden, J. Seiler, S. Miguez, “*ResearchCompendia: CyberInfrastructure for Reproducibility and Collaboration in Computational Science*,” IEEE Computing in Science and Engineering, 17(1), Jan/Feb 2015, p. 12-19. DOI:10.1109/MCSE.2015.18 [14]
- V. Stodden, “*Reproducing Statistical Results*,” Annual Review of Statistics and Its Application, Vol. 2, 2015, p. 1-19. DOI:10.1146/annurev-statistics-010814-020127
Chosen by Annual Reviews for Open Access in support of the 2019 National Academies of Science, Engineering and Medicine report “Reproducibility and Replicability in Science.” [15]
- B. Alberts, R. J. Cicerone, S. E. Fienberg, A. Kamb, M. McNutt, R. M. Nerem, R. Schekman, R. Shiffrin, V. Stodden, S. Suresh, M. T. Zuber, B. Kline Pope, K. Hall Jamieson, “*Self-correction in Science at Work*,” Science, Vol. 348 (6242), 2015, p. 1420-1422. DOI:10.1126/science.aab3847 [16]
- V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. A. Ioannidis, M. Taufer, “*The ‘Reproducibility Enhancement Principles’ for Computational Methods*,” Science, 354 (6317), Dec 2016. DOI:10.1126/science.aah6168 [17]
- D. Greenbaum, J. Rozowsky, V. Stodden, M. Gerstein, “*Structuring Supplemental Materials in Support of Reproducibility*,” Genome Biology, 2017. DOI:10.1186/s13059-017-1205-3 [18]
- Jiminez et al. “*Four Simple Recommendations to Encourage Best Practices in Research Software*,” F1000Research 2017, 6 (876), June 2017. DOI:10.12688/f1000research.11407.1 [19]
- Kafkafi et al. “*Reproducibility and Replicability of Rodent Phenotyping in Pre-clinical Studies*,” Neuroscience & Biobehavioral Reviews, Jan 2018. DOI:10.1016/j.neubiorev.2018.01.003 [20]
- V. Stodden, J. Seiler, Z. Ma, “*Journal Policy for Computational Reproducibility*,” Proceedings of the National Academy of Sciences, March 2018. DOI:10.1073/pnas.1708290115 [21]
- F. Berman, S. Davidson, D. Estrin, B. Halipern, M. Franklin, M. Martonosi, P. Raghavan, R. Rutenbar, V. Stodden, A. Szalay, “*Realizing the Potential of Data Science*,” Communications of the ACM, March 2018. DOI:10.1145/3188721 [22]
- B. Ludäscher, K. Chard, N. Gaffney, M. B. Jones, J. Nabrzyski, V. Stodden, M. Turk, K. Turner, “*Computing Environments for Reproducibility: Capturing the ‘Whole Tale’*,” Future Generation Computer Systems, May 2019. DOI:10.1016/j.future.2017.12.029 [23]
- H. Monajemi, R. Murri, E. Jonas, P. Liang, V. Stodden, and D. Donoho, “*Ambitious Data Science Can Be Painless*,” Harvard Data Science Review, June 2019. DOI:10.1162/99608f92.02ffc552 [24]
- J. Jeschke, K. Börner, V. Stodden, and K. Tockner, “*Open Access Journals Need to Become First Choice in Invasion Ecology and Beyond*,” NeoBiota, 52, Nov 2019. DOI: 10.3897/neo-biota.52.39542 [25]

- V. Stodden, “*The Data Science Life Cycle: A Disciplined Approach to Advancing Data Science*,” [26]
Communications of the ACM, 63(7), July 2020. DOI:10.1145/3360646
- V. Stodden, M. Taufer, “*Approaching a Vision of Reproducibility in Cyberinfrastructure: Building on Community Efforts*,” [27]
Supercomputing Frontiers and Innovations, 7(1), 2020. DOI:10.14529/jsfi200106
- C. Willis and V. Stodden, “*Trust but Verify: How to Leverage Policies, Workflows, and Infrastructure to Ensure Computational Reproducibility in Publication*,” [28]
Harvard Data Science Review, 2(4). 2020. DOI:10.1162/99608f92.25982df
- H. Fineberg, V. Stodden, and X.L. Meng, “*Highlights of the US National Academies Report on ‘Reproducibility and Replicability in Science’*,” [29]
Harvard Data Science Review, 2(4), 2020. DOI:10.1162/99608f92.cb310198
- M. Krafczyk, A. Shi, A. Bhaskar, D. Marinov, V. Stodden, “*Three Empirical Principles for Computational Reproducibility and their Implementation: The Reproduction Package*,” [30]
Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences, March 29, 2021. DOI:10.1098/rsta.2020.0069
- M. Schweinsberg, et al., “*Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis*,” [31]
Organizational Behavior and Human Decision Processes, 165, 2021. DOI:10.1016/j.obhdp.2021.02.003.

EDITED WORKS

- V. Stodden, F. Leisch, and R. Peng, eds., **Implementing Reproducible Research, (A Volume in the R Series)**, [32]
Taylor & Francis, April 2014.
- J. Lane, V. Stodden, S. Bender, H. Nissenbaum, eds., **Privacy, Big Data, and the Public Good: Frameworks for Engagement**, [33]
Cambridge University Press, June 2014.
- V. Stodden, “*Reproducible Research: Tools and Strategies for Scientific Computing*,” [34]
Guest Editor’s Introduction, IEEE Computing in Science and Engineering, Vol. 4, Issue 4, p. 11-12, 2012. DOI: 10.1109/MCSE.2012.82. (Seven contributed articles).
- D. Allison, R. Shiffrin, V. Stodden, “*Reproducibility of Research: Issues and Proposed Remedies*,” [35]
Guest Editor’s Introduction, Proceedings of the National Academy of Sciences, Vol. 115, No. 11, p. 2561-2562, March 2018. DOI: 10.1073/pnas.1802324115. (Twelve contributed articles).
Co-organizers’ edited volume from the National Academies Arthur M. Sackler Colloquium on Improving the Reproducibility of Scientific Research.
- V. Stodden, “*Reproducibility and Replicability in Science*,” [36]
Guest Editor’s Introduction, Harvard Data Science Review. 2020. (Five contributed articles).

CHAPTERS IN BOOKS

- V. Carey and V. Stodden, “*Reproducible Research Concepts and Tools for Cancer Bioinformatics*,” [37]
in **Biomedical Informatics for Cancer Research**, M. F. Ochs, J. T. Casagrande, R. V. Davuluri, eds., Springer, 2010. p. 149-175.
- V. Stodden, “*Innovation and Growth through Open Access to Scientific Research: Three Ideas for High-Impact Rule Changes*,” in **Rules for Growth: Promoting Innovation and Growth Through Legal Reform**, edited by The Kauffman Task Force on Law, Innovation, and Growth. February, 2011. p. 409-432. [38]
- V. Stodden, “*What Computational Scientists Need to Know About Intellectual Property Law: A Primer*,” in **Opening Science: The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing**, S. Bartling and S. Friesike eds., Springer, 2013. p. 225-235. [39]
- V. Stodden, “*Policy and Intellectual Property Rights in Computational Science*,” in **Implementing Reproducible Research**, V. Stodden, F. Leisch, R., Peng, eds., Taylor & Francis, 2014. p. 325-342. [40]
- C. Hurlin, C. Perignon, V. Stodden, “*RunMyCode.org: A Research-Reproducibility Tool for Computational Sciences*,” in **Implementing Reproducible Research**, V. Stodden, F. Leisch, R., Peng, eds., Taylor & Francis, 2014. p. 367-382. [41]
- V. Stodden, “*Enabling Reproducibility in Big Data Research: New Approaches to Intellectual Property and Privacy Law for Scientific Integrity*,” in **Privacy, Big Data, and the Public Good: Frameworks for Engagement**, J. Lane, V. Stodden, H. Nissenbaum, and S. Bender, eds., Cambridge University Press, 2014. p. 112-132. [42]
- D. Bailey, J. Borwein, V. Stodden, “*Facilitating Reproducibility in Scientific Computing: Principles and Practice*,” in **Reproducibility: Principles, Problems, Practices, Prospects**, H. Atmanspacher, S. Maasen, eds., Wiley, 2015. p. 205-232. [43]
- D. Donoho, V. Stodden, “*Reproducible Research in Computational Mathematics*,” invited to **The Princeton Companion to Applied Mathematics**, edited by N. J. Higham; M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa & J. Tanner, associate editors, 2015, p. 916-925. [44]

REFEREED CONFERENCE ARTICLES

- D. Donoho and V. Stodden, “*When Does Non-Negative Matrix Factorization Give a Correct Decomposition Into Parts?*,” Proceedings NIPS 2003. [45]

- D. Donoho, V. Stodden, “*Breakdown Point of Model Selection When the Number of Variables Exceeds the Number of Observations*,” Proc. IEEE International Joint Conference on Neural Networks, Vancouver, BC., 2006. [46]
- E. Hurowitz, I. Drori, V. Stodden, “*Fast l_1 Minimization for Genomewide Analysis of mRNA Lengths*,” IEEE International Workshop on Genomic Signal Processing and Statistics, 2006. [47]
- D. Donoho and V. Stodden, “*Breakdown Point of Model Selection When the Number of Variables Exceeds the Number of Observations*,” IEEE World Congress on Computational Intelligence 2006. [48]
- V. Stodden and P. Meier, “*A Global Empirical Evaluation of New Communication Technology Use and Democratic Tendency*,” Nominated for best paper, 3rd IEEE/ACM International Conference on Information and Communication Technologies and Development, Doha, Qatar, April 2009. [49]
- V. Stodden, “*Data Sharing in Social Science Repositories: Facilitating Reproducible Computational Research*,” NIPS workshop: Computational Science and the Wisdom of Crowds, Dec 2010. [50]
- V. Stodden, C. Hurlin, C. Perignon, “*RunMyCode.org: A Novel Dissemination and Collaboration Platform for Executing Published Computational Results*,” IEEE International Conference on eScience, *Workshop on Analyzing and Improving Collaborative eScience with Social Networks*, 2012. [51]
- V. Stodden, H. Reich, “*Software Patents as a Barrier to Scientific Transparency: An Unexpected Consequence of Bayh-Dole*,” Conference on Legal Empirical Studies, Stanford, Nov. 2012. [52]
- V. Stodden, S. Miguez, “*Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research*,” Workshop on Sustainable Software for Science: Practice and Experiences, 2013. [53]
- V. Stodden, S. Miguez, “*Provisioning Reproducible Computational Science Information*,” reproducibility@XSEDE: An XSEDE14 Workshop, 2014. [54]
- B. Ludäscher, K. Chard, N. Gaffney, M. B. Jones, J. Nabrzyski, V. Stodden, M. Turk, “*Capturing the ‘Whole Tale’ of Computational Research: Reproducibility in Computing Environments*,” Science Gateways 2016. [55]
- H. Monajemi, D. L. Donoho, V. Stodden, “*Making Massive Computational Experiments Painless*,” IEEE BigData 2016, Open Science in Big Data (OSBD 2016), Dec 5, 2016. [56]
- V. Stodden, X. Wu, “*Defining the AIM: An Abstraction for Improving Machine Learning Prediction*,” American Statistical Association Symposium on Data Science and Statistics, May 2018. [57]
- V. Stodden, M. S. Krafczyk, A. Bhaskar, “*Enabling the Verification of Computational Results: An Empirical Evaluation of Computational Reproducibility*,” First International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS18), June 2018. DOI: 10.1145/3214239.3214242 [58]

- V. Stodden, X. Wu, V. Sochat, “*AIM: An Abstraction for Improving Machine Learning Prediction*,” [59]
IEEE Data Science Workshop Proceedings, June 2018. DOI: 10.1109/DSW.2018.8439914
- V. Stodden, M. S. Krafczyk, “*Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context*,” The 1st Workshop on Reproducible, Customizable and Portable Workflows for HPC (ResCuE-HPC), November 2018. [60]
- B. Mecum, S. Wyngaard, C. Willis, M. Turk, T. Thelen, I. Taylor, V. Stodden, D. Perez, [61]
J. Nabrzyski, B. Ludaescher, S. Kulasekaran, K. Kowalik, M. B. Jones, M. Hategan,
N. Gaffney, K. Chard, A. Brinckman, “*Science, Containerized: Integrating Provenance and Compute Environments with the Whole Tale*,” AGU Fall Meeting Abstracts, 2018.
- K. Chard, N. Gaffney, M. B. Jones, B. Ludäscher, J. Nabrzyski, V. Stodden, M. Turk, C. Willis, [62]
“*Implementing Computational Reproducibility in the Whole Tale Environment*,” Second International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS19), June 2019. DOI: 10.1145/3322790.3330594
- V. Welch, E. Deelman, V. Stodden, M. Taufer, “*Initial Thoughts on Cybersecurity and Reproducibility*,” [63]
Second International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS19), June 2019. DOI: 10.1145/3322790.3330593
- M. S. Krafczyk, A. Shi, A. Bhaskar, D. Marinov and V. Stodden, “*Scientific Tests and Continuous Integration Strategies*,” [64]
Second International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS19), June 2019. DOI: 10.1145/3322790.3330595
- V. Stodden, V. Ferrini, M. Gabanyi, K. Lehnert, J. Morton, H. Berman, “*Open Access to Research Artifacts: Implementing the Next Generation Data Management Plan*,” [65]
Association for Information Science and Technology (ASIST19), 2019. DOI: 10.1002/pra2.51
- K. Chard, N. Gaffney, M. B. Jones, K. Kowalik, B. Ludäscher, J. Nabrzyski, V. Stodden, I. Taylor, [66]
T. Thelen, M. J. Turk, C. Willis, “*Application of BagIt-Serialized Research Object Bundles for Packaging and Re-execution of Computational Analyses*,” IEEE 15th International Conference on e-Science (e-Science), San Diego, 2019. DOI: 10.1109/eScience.2019.00068
- W. Lam, S. Winter, A. Astorga, V. Stodden and D. Marinov, “*Understanding Reproducibility and Characteristics of Flaky Tests Through Test Reruns in Java Projects*,” [67]
2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE), 2020. DOI:10.1109/ISSRE5003.2020.00045
- V. Stodden, “*Beyond Open Data: A Model for Linking Digital Artifacts to Enable Reproducibility of Scientific Claims*,” [68]
Third International Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS20), June 2020. DOI:10.1145/3391800.3398172
- P. Zhang, Y. Jiang, A. Wei, V. Stodden, D. Marinov and A. Shi, “*Domain-Specific Fixes for Flaky Tests with Wrong Assumptions on Underdetermined Specifications*,” [69]
IEEE/ACM 43rd International Conference on Software Engineering (ICSE), 2021. DOI:10.1109/ICSE43902.2021.00018

TECHNICAL REPORTS

- D. Donoho, V. Stodden, Y. Tsaig, “*SparseLab Architecture*,” Stanford Department of Statistics Technical Report, SparseLab 2.0, March 2007. [70]
- D. Donoho, V. Stodden, Y. Tsaig, “*About SparseLab*,” Stanford Department of Statistics Technical Report, SparseLab 2.0, March 2007. [71]
- V. Stodden, “*The Scientific Method in Practice: Reproducibility in the Computational Sciences*,” MIT Sloan Research Paper No. 4773-10. 2010. DOI:10.2139/ssrn.1550193 [72]
- M. A. Heroux, L. Barba, M. Parashar, V. Stodden, M. Taufer, “*Toward a Compatible Reproducibility Taxonomy for Computational and Computing Sciences*,” Sandia National Lab (SNL-NM), Albuquerque, NM (United States), 2018. [73]

OTHER SCHOLARLY PUBLICATIONS

- V. Stodden, “*Cogitamus, Ergo Sum? The ‘Difference Between Knowing the Name of Something and Knowing Something’*” reply to The Edge Annual Question 2009: *How Has the Internet Changed the Way You Think?* Jan 2010. [74]
- V. Stodden, “Remarks,” presented before The National Academies of Sciences, Engineering, and Medicine Committee on The Impact of Copyright Policy on Innovation in the Digital Era, Washington DC, Oct 15, 2010. [75]
- V. Stodden with Task Force co-authors, “*Cyber Science and Engineering: A Report of the NSF Advisory Committee on Cyberinfrastructure*,” Task Force on Grand Challenges, Nov 2010. [76]
- V. Stodden, “*White Paper for Expert Panel Discussion on Data Policies*,” for a Workshop of the National Science Board Expert Panel on Data Policies, March 27-29, 2011. [77]
- V. Stodden, “*Trust your Science? Open Your Data and Code*,” Amstat News, July 1, 2011. [78]
- V. Stodden with participants, “*Changing the Conduct of Science: Summary Report of the Workshop Held on November 12, 2010*,” at the National Science Foundation Workshop Changing the Conduct of Science in the Information Age, June, 2011. [79]
- V. Stodden and S. Arbesman, “*Scientists, Share Secrets or Lose Funding*,” Bloomberg View, Jan 10, 2012. [80]
- V. Stodden, “*Phase Transitions And ‘Scale Transitions:’ Conceptualizing Unexpected Changes Due To Scale*” reply to The Edge Annual Question 2010: *What Scientific Concept Would Improve Everybody’s Cognitive Toolkit?* Jan 2011. Published in *This Will Make You Smarter: New Scientific Concepts to Improve Your Thinking*, Harper Perennial, February 14, 2012. [81]

- V. Stodden, “*Fact, Fiction, and Our Probabilistic World*” reply to The Edge Annual Question 2011: *What is your favorite deep, elegant, or beautiful explanation?*, Jan 2012. Published in *This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works*, Harper Perennial, January 22, 2013. [82]
- V. Stodden, “*Where did you get that fact?*” reply to The Edge Annual Question 2012: *What should we be worried about?*, Jan 2013. [83]
- V. Stodden, D. Bailey, J. Borwein, R. LeVeque, W. Rider, and W. Stein “*Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics*,” ICERM Workshop Report, 2013. [84]
- V. Stodden, “*Testimony Before the House Committee on Science, Space and Technology Subcommittee on Research*, Hearing on Scientific Integrity & Transparency, March 5, 2013. [85]
- D. Bailey, J. Borwein, V. Stodden, “*Set the Default to ‘Open’*,” Notices of the AMS, June 2013. [86]
- V. Stodden, D. Bailey, J. Borwein, “*‘Setting the Default to Reproducible’ in Computational Science Research*,” SIAM News, June 2013. [87]
- V. Stodden, “*Resolving Irreproducibility in Computational and Empirical Research*,” invited IMS Bulletin, December 2013. [88]
- J. Lane, V. Stodden, “*What, Me Worry? What to do about Privacy, Big Data, and Statistical Research*,” Amstat News, December 2013. [89]
- V. Stodden, “*Reproducibility*” reply to The Edge Annual Question 2014: *What scientific idea is ready for retirement?*, Jan 2014. [90]
- D. James, N. Wilkins-Diehr, V. Stodden, D. Colbry, C. Rosales, “*Standing Together for Reproducibility in Large-Scale Computing: Report on reproducibility@XSEDE, An XSEDE14 Workshop*,” December 2014. [91]
- V. Stodden with Committee Members, “*Realizing the Potential of Data Science*,” Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group, Dec 2016. [92]
- V. Stodden, “*Epsilon*” reply to The Edge Annual Question 2017: *What scientific term ought to be more widely known?*, Jan 2017. [93]
- V. Stodden, “*How do I know the right level of abstraction at which to explain a phenomenon?*” reply to The Edge Annual Question 2018: *What is the Last Question?*, Jan 2018. [94]
- V. Stodden with Committee Members, “*Fostering Integrity in Research*,” National Academies of Sciences, Engineering, and Medicine, April 2017. [95]
- V. Stodden with Committee Members, “*Reproducibility and Replication in Science*,” National [96]

Academies of Sciences, Engineering, and Medicine, April 2019.

V. Stodden with Committee Members, “*Data Science at the National Science Foundation*,” a report [97] of the Statistics Subcommittee (chaired by Iain Johnstone and Fred Roberts) of the Advisory Committee for the National Science Foundation Directorate for Mathematical and Physical Sciences, forthcoming.

CREATIVE WORKS (SOFTWARE)

- Sep 05 - present* Developed and maintaining SparseLab webpage for the collaborative matlab toolbox distribution from my dissertation (over 7000 downloads in 2008).
<http://sparselab.stanford.edu>
- Mar 12 - present* RunMyCode.org: Collaboration to develop open availability of code and data with published computational results.
- Apr 13 - present* ResearchCompendia.org: Sloan Foundation funded project to develop an open-source platform to support online research “compendia” that make available data and code alongside published results, and verify findings in the cloud.
- Sep 16 - present* ezdmp.org: NSF funded collaborative project to develop an open-source next generation data management plan tool.
- Oct 17 - present* WholeTale.org: NSF funded collaborative project to develop an open-source platform that supports modern research tools such as the Jupyter notebook, and creates online research “compendia” that make available data and code as a re-executable package called a “Tale.”

Mentoring

POSTDOCTORAL SCHOLARS

- Aug 2016 - Aug 2019* Matthew Krafczyk (National Center for Supercomputing Applications, University of Illinois Urbana-Champaign. Now at C3.ai Digital Transformation Institute, Illinois)
- Aug 2013 - Aug 2015* Jennifer Seiler (Department of Statistics, Columbia University. Now Computational Astrophysicist at Giant Army, Washington D.C.)

DOCTORAL STUDENTS

- August 2021 -* Adhithya Bhaskar (Department of Industrial and Systems Engineering, School of Engineering, University of Southern California); Advisor

<i>Graduated August 2020</i>	Craig Willis (School of Information Sciences, University of Illinois Urbana-Champaign); Advisor and Director of Research
<i>August 2019 - January 2020</i>	Rutvik Chaudhury (Department of Computer Science, School of Engineering, University of Illinois Urbana-Champaign); Advisor
<i>Expected graduation May 2021</i>	Kelechi Ikegwu (Informatics); Doctoral Committee Member
<i>Graduated May 2018</i>	Q. Chelsea Song (Psychology); Doctoral Committee Member; Now Assistant Professor, Department of Psychology, Purdue University

MASTER'S AND UNDERGRADUATE STUDENTS

<i>Aug 2019 - July 2020</i>	Peilun Zhang (Master's Thesis Primary Co-Advisor, Department of Computer Science, University of Illinois Urbana-Champaign)
<i>June 2020 - Aug 2020</i>	Mingzhe Wang (Undergraduate Summer Intern, Department of Computer Science, University of Illinois Urbana-Champaign)
<i>June 2019 - Aug 2019</i>	Yang Yu (Master's, Department of Statistics Summer Intern, University of Illinois Urbana-Champaign)
<i>June 2017 - Aug 2017</i>	Jennifer Beamer (Master's Summer Intern, Library Sciences, University of Hawai'i at Manoa)
<i>Aug 2017 - May 2018</i>	Xiaomian Wu (Master's, Department of Statistics, University of Illinois Urbana-Champaign, co-authored [54])
<i>Jan 2017 - May 2018</i>	Yantong Zheng (Undergraduate and Master's, Department of Statistics, University of Illinois Urbana-Champaign)
<i>Jan 2017 - May 2018</i>	Daniel Lee (Undergraduate, Department of Statistics, University of Illinois Urbana-Champaign)
<i>Aug 2016 - Dec 2016</i>	Kefu Zhu, Undergraduate, James Scholar Honors Project Supervision (Department of Statistics, University of Illinois Urbana-Champaign)
<i>Aug 2016 - Dec 2016</i>	Saumil Padhya, Undergraduate, James Scholar Honors Project Supervision (Department of Statistics, University of Illinois Urbana-Champaign)
<i>Aug 2015 - May 2016</i>	William Pooler (Master's, School of Information Sciences, University of Illinois Urbana-Champaign)
<i>Aug 2015 - May 2016</i>	April Tang (Undergraduate, Department of Statistics, University of Illinois Urbana-Champaign)
<i>June 2014 - Aug 2014</i>	Pakwesi Taylor (Undergraduate, Department of Statistics, Columbia University)
<i>Aug 2013 - Dec 2013</i>	Christine Byun (Undergraduate, Independent Study, Department of Statistics, Columbia University)
<i>June 2012 - Aug 2012</i>	Xiaokun Ma (Undergraduate, Department of Statistics, Columbia University, co-authored [11] and [21])
<i>June 2012 - Aug 2012</i>	Isabel Reich (Undergraduate, Department of Statistics, Columbia University, co-authored [49])
<i>June 2011 - Aug 2011</i>	Peixuan Guo (Undergraduate, Department of Statistics, Columbia University, co-authored [11])
<i>June 2011 - Aug 2011</i>	Michael Fusella (Undergraduate, Department of Statistics, Columbia University)
<i>June 2011 - Aug 2011</i>	Matthew Lewis (Undergraduate, Department of Statistics, Columbia University)

Teaching

COURSES DEVELOPED

Jan 2022 **Predictive Analytics**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Jan 2022 **Engineering Statistics**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Aug 2021 **Discrete Systems Simulation**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Aug 2019 **Concepts of Machine Learning**, Undergraduate Course, School of Information Sciences, University of Illinois Urbana-Champaign
Aug 2019 **Data Management, Curation & Reproducibility**, Undergraduate Course, School of Information Sciences, University of Illinois Urbana-Champaign
Aug 2019 **Legal Aspects of Information Systems**, Undergraduate Course, School of Information Sciences, University of Illinois Urbana-Champaign
Aug 2018 **Methods for Data Science**, Master's Course, School of Information Sciences, University of Illinois Urbana-Champaign
Aug 2015 **Introduction to Data Science**, Master's Course, School of Information Sciences, Departments of Computer Science and Statistics, University of Illinois Urbana-Champaign
Aug 2015 **Intellectual Property for Scholarship**, Master's Course, School of Information Sciences, University of Illinois Urbana-Champaign
Aug 2014 **Data Policy Seminar**, Master's Course, School of Information Sciences, University of Illinois Urbana-Champaign
Jan 2012 **Replicating Computational Results**, Doctoral Level Course, Columbia University IGERT "From Data to Knowledge" crossover course, jointly offered between statistics and EECS
Jan 2011 **Capstone Course in Data Science**, Master's Course, Department of Statistics, University of California at Berkeley
Aug 2010 **Statistical Computing in SAS**, Master's Course, Department of Statistics, Columbia University
Aug 2010 **Introduction to Data Science**, Master's Course, Department of Statistics, Columbia University
June 2003 **Statistical Computing in SAS**, Master's Course, Department of Statistics, Stanford University

COURSES TAUGHT

Jan 2022 **Predictive Analytics**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Jan 2022 **Engineering Statistics**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Aug 2020-21 **Discrete Systems Simulation**, Undergraduate Course, Department of Industrial and Systems Engineering, University of Southern California
Fall 2019 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.
Spring 2019 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.
Fall 2018 **Introduction to Data Science**, Master's Course, University of Illinois Urbana-Champaign.
 Introduction to Data Science, Master's Course, University of Illinois Urbana-Champaign.
Spring 2018 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.

Methods for Data Science, Master's Course, University of Illinois Urbana-Champaign.
Journal Club on Data Science and Reproducibility, University of Illinois Urbana-Champaign.

Fall 2017 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.

Data, Statistics, and Information, Master's Course, University of Illinois Urbana-Champaign.

Spring 2017 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.

Data, Statistics, and Information, Master's Course, University of Illinois Urbana-Champaign.

Fall 2016 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.

Data, Statistics, and Information, Master's Course, University of Illinois Urbana-Champaign.

Spring 2016 **Introduction to Data Science**, Master's Course, (crosslist with CS and Stats), University of Illinois Urbana-Champaign.

Data, Statistics, and Information, Master's Course, University of Illinois Urbana-Champaign.

Fall 2015 **Introduction to Data Science**, Master's Course, University of Illinois Urbana-Champaign.

Spring 2015 **Data Policy Seminar**, Master's Course, University of Illinois Urbana-Champaign.

(co-taught) **Socio-technical Data Analysis**, Master's Course, University of Illinois Urbana-Champaign.

Fall 2014 **Applied Data Mining**, Undergraduate Course, Columbia University.

Summer 2014 **Introduction to Data Science**, Undergraduate Course, Columbia University.

Summer 2014 **Introduction to Probability**, Undergraduate, Columbia University.

Spring 2014 **Capstone Masters course in Data Science**, Master's Course, University of California at Berkeley.

Fall 2013 **Linear Regression Models**, Undergraduate Course, Columbia University.

Summer 2013 **Concepts in Computing with Data**, Undergraduate Course, University of California at Berkeley.

Spring 2012 **Replicating Computational Results**, Columbia University IGERT "From Data to Knowledge" crossover course, jointly offered between statistics and EECS.

Fall 2012 **Statistical Computing in SAS**, Undergraduate Course, Columbia University.

Fall 2011 **Linear Regression Models**, Undergraduate Course, Columbia University.

Fall 2011 **Statistical Computing in SAS**, Undergraduate Course, Columbia University.

Summer 2011 **Linear Regression Models**, Undergraduate Course, Columbia University.

Spring 2007 **Quantitative Methods: Statistical Inference**, Stanford Law School.

Fall 2006 **Empirical Research Seminar** (co-taught), Stanford Law School.

Service

SCHOOL AND DEPARTMENT SERVICE

Aug 2021 - Aug 2022 Chair, Department of Industrial and Systems Engineering Hiring Committee, University of Southern California.

Aug 2021 - Aug 2022 Chair, Department of Industrial and Systems Engineering Diversity Committee, University of Southern California.
Aug 2021 - Aug 2022 Department of Industrial and Systems Engineering PhD Admissions Committee, University of Southern California.
Summer 2021 Chair, Department of Industrial and Systems Engineering Hiring Committee, University of Southern California.
Aug 2020 - Aug 2022 Department of Industrial and Systems Engineering PhD Curriculum Committee, University of Southern California.
Aug 2019 - Aug 2020 School of Information Sciences Diversity Committee, University of Illinois Urbana-Champaign.
Aug 2019 - Aug 2020 School of Information Sciences Master's of Information Management Committee, University of Illinois Urbana-Champaign.
Aug 2016 - Aug 2019 School of Information Sciences Curriculum Committee, University of Illinois Urbana-Champaign.
Aug 2014 - Aug 2015 School of Information Doctoral Students Committee, University of Illinois Urbana-Champaign.

UNIVERSITY SERVICE

Aug 2019 - Aug 2022 Budget Committee, University of Illinois Urbana-Champaign.
Aug 2018 - Aug 2020 Discovery Partners Institute (DPI) Culture & Society (C&S) Inaugural Working Subcommittee, University of Illinois Urbana-Champaign.
Oct 2017 - Aug 2020 Scientific Advisory Board for the University of Illinois News Bureau, University of Illinois Urbana-Champaign.
Aug 2015 - Aug 2018 IT Committee, University of Illinois Urbana-Champaign.
Aug 2016 - Aug 2019 Senate IT Committee, Columbia University.

EXTERNAL ACTIVITIES

Aug 2022 Committee of Visitors (COV) for the Office of Advanced Cyberinfrastructure (OAC), Directorate for Computer and Information Sciences and Engineering, National Science Foundation.
2016 - National Academy of Engineering (NAE), Advisory Group for the Online Resource Center for Ethics Education in Engineering and Science; Editorial Boards for Research Ethics and Social and Behavioral Sciences.
2020 - 21 IEEE CS Ad Hoc Committee on Open Science and Reproducibility, reporting to the Board of Governors.
2020 - Chair, ACM Emerging Interest Group on *Reproducibility and Independent Verification*.
2020 - Advisory Committee for American Educational Research Association (AERA) & Council of Graduate Schools (CGS) "Examining Impact and Fostering Academic Support for Open Science Products."

- 2019 - 2022 University of California Irvine, Advisory Committee for the Machine Learning Repository under the NSF Award “CCRI: ENS: Machine Learning Democratization via a Linked, Annotated Repository of Datasets.”
- 2019 - Associate Editor, *Harvard Data Science Review*.
- 2019 - National Information Standards Organization (NISO) Working Group Member: Taxonomy, Definitions, and Recognition Badging Scheme Working Group.
- 2018 - Associate Editor for Reproducibility, *IEEE Transactions on Parallel and Distributed Systems*.
- 2017 - 19 ACM Ethics & Plagiarism Committee.
- 2017 - Advisor to the Curating for Reproducibility (CURE) Consortium, Yale University.
- 2016 - ACM Task Force on Data, Software, and Reproducibility in Publication.
- 2016 - Transparency and Openness Promotion (TOP) Guidelines Coordinating Committee, Center for Open Science.
- 2016 - Advisory Board Member, Research Ideas and Outcomes. *The Open Science Journal*.
- 2016 - 19 American Statistical Association, Committee on Professional Ethics.
- 2016 - 18 Associate Editor for Reproducibility, *Journal of the American Statistical Association, Applications and Case Studies*.
- 2015 - Social Science Research Council, Digital Culture Advisory Board.
- 2015 - Associate Editor, Institute of Mathematical Statistics, *Annals of Applied Statistics*.
- 2015 - 18 Advisory Group on Reproducibility to the Supercomputing Conference, ACM, and IEEE.
- 2015 - Project TIER (Teaching Integrity in Empirical Research), Haverford University, Advisory Board. Alfred P. Sloan Foundation funded.
- 2014 - The International Mathematical Union, Committee on Electronic Information and Communication.
- 2013 - Board of Advisors, American Statistical Association, *Statistical Analysis and Data Mining*.
- 2013 - 15 American Statistical Association, Committee on Privacy and Confidentiality.
- 2013 - 14 American Statistical Association, Presidential Strategic Initiative, Developing a Prototype Statistics Portal.

- 2013 - 14* American Statistical Association, co-chair of the Committee on Data Sharing and Reproducibility.
- 2014 - 17* Member, *NSF Advisory Committee for the Computing and Information Science and Engineering (CISE) Directorate*.
- Jul 13 - 15* Co-chair of the *NSF Advisory Committee for Advanced CyberInfrastructure*, Office of Advanced Cyberinfrastructure, Computing and Information Science and Engineering (CISE) Directorate (ACCI).
- Jul 11 - 15* Member, *NSF Advisory Committee for Advanced CyberInfrastructure*, Office of Advanced Cyberinfrastructure, Computing and Information Science and Engineering (CISE) Directorate (ACCI).



ISE 529: PREDICTIVE ANALYTICS

Spring 2022

Mondays and Wednesdays 3:30-4:50PM

Victoria Stodden (Professor)

stodden@usc.edu

Office hours: Mondays and Wednesdays 5-6PM on zoom and by appointment
Weekdays drop in hours 6-7PM (about 90% of the time)

Adhithya Bhaskar (Teaching Assistant)

adhithya@usc.edu

Office hours: Tuesday and Wednesdays 10-11AM on zoom

COURSE DESCRIPTION AND OBJECTIVES

Prerequisite(s): An undergraduate course on Statistics.

Recommended Preparation: ISE 225 (Engineering Statistics I) or equivalent, working knowledge of a programming language.

Analytics for supervised and unsupervised statistical learning. Generalized linear models, discriminant analysis, support vector machines. Nonparametric classification, trees, ensemble methods, k-nearest neighbors. Principal components, clustering.

This course focuses on building models for prediction, classification, and clustering. For the first two cases the objective is to predict a numeric value or a category. For clustering the objective is to group observations in clusters that share some common attributes.

The standard multiple linear regression model is the foundational prediction model. This model is extended to shrinkage models (ridge and lasso regression) for improved accuracy and dimension reduction. Overfitting, bias, cross validation, and AIC are reviewed to help evaluate the performance of these models.

The course also focuses on regression models for a categorical response. Trees and ensembled trees (random forests, bagging, and boosting), discriminant analysis, and support vector machines. For these models the prediction is a category.

Clustering focuses on discovering relationships between variables when a response is not available. Principal components analysis and clustering (K-means clustering and hierarchical clustering) are reviewed.

Learning Objectives and Outcomes

- To understand the difference between supervised and unsupervised learning methods.
- To learn common data aggregation operations.
- To build regression models for prediction and classification.
- To understand key concepts for predictive analytics (overfitting, shrinkage, regularization, R^2 , adjusted R^2 , VIF, mean square prediction error, cross-validation).
- To compare the performance of different prediction and classification models.
- To build models to classify observations into two or more classes.
- Using principal components for clustering and prediction.
- To be able to explain the difference between different clustering methods.

COURSE MATERIALS

All course materials will be available on Blackboard. A reminder that you are responsible for taking notes in class and attending class.

Primary resources:

- (ISLR) An Introduction to Statistical Learning with Applications in R, Second Edition, by James, Witten, Hastie, and Tibshirani (2021). Available in blackboard and can be downloaded for free here: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
- (ESL) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, by Hastie, Tibshirani, and Friedman. Second Edition. Available in blackboard and can be downloaded for free here: <https://hastie.su.domains/Papers/ESLII.pdf>

Supplemental resource: (OIS) OpenIntro Statistics, Fourth Edition, by David Diez, Mine Çetinkaya-Rundel, Christopher D Barr (2019). Available in blackboard and can be downloaded for free here: <https://leanpub.com/os>

Supplemental R resources (free download):

- P. Kuhnert & B. Venables, An Introduction to R: Software for Statistical Modeling & Computing http://cran.r-project.org/doc/contrib/Kuhnert+Venables-R_Course_Notes.zip
- J.H. Maindonald, Using R for Data Analysis and Graphics <http://cran.r-project.org/doc/contrib/usingR.pdf>
- W.J. Owen, The R Guide <http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>

- W.N. Venables & D. M. Smith, An Introduction to R
<http://cran.r-project.org/doc/manuals/R-intro.pdf>

GRADING

30% Homework
 30% Midterm Exam (March 9)
 40% Final Exam (May 6)

Homework will be assigned and due weekly. Your two lowest homework scores will be dropped. No late assignments will be accepted. Class participation will allow me to round up (e.g. from A- to A) if your final score is close. You must be here for the final exam - no travel exceptions will be given. Accommodations are expected to go through the Office of Student Accessibility Services (OSAS). Please let me know of any accommodations at the start of the semester.

All grades changes must be done by Prof. Stodden. Submit your work to her with a separate written description of why you think the grade you received is incorrect.

HOMEWORKS

All homeworks must be turned in in blackboard. For computational questions expect us to rerun your computations to produce your tests: after HW2 you will turn in an .Rmd file and the corresponding pdf containing a narrative answering the homework questions, along with code and results. You must write your own code.

SCHEDULE AND READINGS

Subject to Change

Week of	Topic	Reading
January 10	What is Statistical Learning? R Review	ISLR ch 2
January 17	Simple Linear Regression	ISLR 3.1
January 24	Multiple Linear Regression	ISLR 3.2
January 31	Logistic Regression, Linear Discriminant Analysis	ISLR 4.1-4.5

February 7	Generalized Linear Models. Resampling: Cross-validation	ISLR 4.6 ch 5
February 14	Resampling cont: The Bootstrap	ISLR ch 5
February 21	Linear Model Selection: Subset Selection	ISLR 6.1
February 28	Linear Model Selection: Regularization and Shrinkage Methods	ISLR 6.2
March 7	Review for exam (March 7) Midterm March 9 (possibly take home)	
March 14	SPRING RECESS	
March 21	Dimensionality Reduction and Principal Components Analysis	ISLR 6.3, 6.4
March 28	Polynomial Regression and Basis Functions	ISLR 7.1-7.3
April 4	Regression Splines and Smoothing Splines	ISLR 7.4-7.5
April 11	Decisions Trees and Bagging, Random Forests, Boosting, and Bayesian Additive Regression Trees	ISLR 8.1, 8.2
April 18	Support Vector Machines	ISLR ch 9
April 25	Deep Learning. Possibility of a take home final exam.	ISLR ch 10
May 6 2-4pm	Final Exam (university)	

	scheduled)	
--	------------	--

ACADEMIC INTEGRITY

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” <https://policy.usc.edu/scampus-part-b/>.

Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct>.

COVID-19 Statement:

In keeping with University policy, all students are required to engage in appropriate behavior to protect the health and safety of our community. If you feel ill or are unable to come to class or complete class assignments due to issues related to COVID-19, including but not limited to: testing positive yourself, feeling ill, caring for a family member with COVID-19, or having unexpected child-care obligations, contact the instructor immediately.

Intellectual Property Policies:

Any misuse, inappropriate dissemination, or attempted sale of class recordings and handouts, as well the appropriation of intellectual property is not acceptable. It is the student’s responsibility to appropriately use and handle these recordings under existing campus policies regarding class notes (<https://policy.usc.edu/scampus-part-c/>).

Students are not permitted to create their own class recordings without the instructor’s permission. Violations of these policies will be met with the appropriate disciplinary sanction.

Netiquette Policies:

- Be respectful and considerate towards each other. Pay attention to the cultural and background differences.
- While in Zoom sessions please show your complete first name and last name, as in the roster.
- While in Zoom sessions please have your VIDEO ON. It assists in building a sense of community. However, if Internet signal issues preclude you from having a reasonable audio, you

may wish to “split” your signal by connecting with both your phone (for audio) and computer (for video)

- While in Zoom sessions be respectful to the LEARNING ENVIRONMENT such as: be attentive, have proper background, follow business casual norms, and be in a stationary location (not driving). Please reduce distractions for others.

Support Systems:

The Office of Student Accessibility Services (OSAS)

Provides certification for students with disabilities and helps arrange relevant accommodations.

<http://dsp.usc.edu>

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

<https://engemannshc.usc.edu/counseling/>

National Suicide Prevention Lifeline - 1-800-273-8255

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

<https://engemannshc.usc.edu/rsvp/>

Sexual Assault Resource Center

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

Office of Equity and Diversity (OED)/Title IX Compliance – (213) 740-5086

Works with faculty, staff, visitors, applicants, and students around issues of protected class.

<https://equity.usc.edu/>

Bias Assessment Response and Support

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response.

<https://studentaffairs.usc.edu/bias-assessment-response-support/>

Student Support and Advocacy – (213) 821-4710

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

USC Emergency Information

Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, <http://emergency.usc.edu>

USC Department of Public Safety – 213-740-4321 (UPC) and 323-4

ISE 435 Discrete Systems Simulation

Fall 2021

Lecture: MW 3:30-4:50pm (LVL 16 and Zoom)

Labs: Th 4-4:50pm (SAL 126) and 5-5:50pm (SAL 127)

Website: blackboard.usc.edu

Syllabus version: Oct 1, 2021

Instructor: Victoria Stodden

Office: Online (see zoom link in blackboard)

Office Hours: Mondays and Wednesdays 5-6pm, Fridays 10-11am

Phone: 213-740-4893

E-mail: stodden@usc.edu

TA: Adhithya Bhaskar

Office: Online (see zoom link in blackboard)

Office Hours: Mondays noon-1pm, Thursdays 6-7pm

E-mail: adhithya@usc.edu

TA: Ryan Schimpf

Office: Online (see zoom link in blackboard)

Office Hours: Thursdays 2-3pm, Fridays noon-1pm

E-mail: schimpf@usc.edu

Objective: In this course, you will learn how to create computer models of discrete event systems, and how to use these models to make decisions about the design/improvement of the actual physical systems that the models represent. You will learn how to evaluate a system and identify the input and output variables. You will learn how to evaluate field data to obtain input information and how to evaluate output predictions from the simulation model to select effective operating policies.

Text: There is no assigned textbook for this class. We will make use of several resources that will be made available during the course.

A supplemental textbook, Kelton, Sadowski, and Zupick, Simulation with Arena, 6th edition, is available on reserve at the Science and Engineering Library.

Software: We will use Excel, Matlab, and Arena for this course. We will use the Simulink toolbox in Matlab. No previous experience with Matlab or Arena is required.

Grading Policies:

Points Breakdown -	
Quizzes	8
Homework	25
Midterm Exam	25
Final Exam	<u>42</u>
Total	100

Course grades will be determined by the distribution of point totals for the class. “Natural groupings” will be used to assign letter grades. The highest scoring group will receive A’s, the next group is the B’s, and so on. A single point will not be the difference between any two letter grades. A “gap” must exist to create a grade boundary.

Homework assignments will be due in blackboard by the start of class on Mondays. Late homeworks will not be accepted, however your *two* lowest homeworks will be dropped.

The MIDTERM EXAM will be on **Wednesday, October 13**.

The FINAL EXAM is scheduled for **Monday, December 13, from 2-4pm**. This date is set by the university. It will be comprehensive with an emphasis on the material after the midterm.

The exams will cover the material presented up to and including the preceding homework assignment. Points will be assigned to each section of the exam. Partial credit will be awarded according to work shown. No re-takes will be allowed. No make-up exam will be given.

Course Schedule:

(subject to change)

Week	Topic	References
1	Discrete Event Systems Queuing Systems Terms & Definitions	KSZ: 1.1-1.4, KP: 2.1-2.2, Handout
2	Simulation Overview / Simulations by Hand	KSZ: 2.1, 2.2, 2.6-2.8 KP: 3.3-3.4
3	Introduction to Matlab and Simulink QUIZ 1	Handout
4	Modeling Basic Operations	Handout
5	Input Analysis	KP: 4.1-4.3

6	Modeling Detailed Operations QUIZ 2	KP: 4.1-4.3
7	Output Analysis for Terminating Simulations	KP: 6.1-6.5
8	Output Analysis for Steady-State Simulations MIDTERM	KP: 6.1-6.5
9	Inventory Systems	Handout
10	Random Numbers and Random Variates Nonstationary Poisson Process QUIZ 3	KSZ: 12.1-12.3
11	Batch Means Method	KP: 6.10
12	Simulation Experiments QUIZ 4	Handout
13	Conducting Simulation Studies	KSZ: 13.1-13.9
14	Comparisons with Arena	KSZ: 13.1-13.9
15	Review / Crumple Zone / Special Topics	

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Part B, Section 11, “Behavior Violating University Standards” <https://policy.usc.edu/scampus-part-b/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct>.

COVID-19 Statement:

In keeping with University policy, all students are required to engage in appropriate behavior to protect the health and safety of our community. If you feel ill or are unable to come to class or complete class assignments due to issues related to COVID-19, including but not limited to: testing positive yourself, feeling ill, caring for a family member with COVID-19, or having unexpected child-care obligations, contact the instructor immediately.

Intellectual Property Policies:

This is a clarification that any misuse, inappropriate dissemination, or attempted sale of class recordings and handouts, as well the appropriation of intellectual property is not acceptable. It is student’s responsibilities towards the appropriate use and handling of these recordings under existing campus policies regarding class notes (<https://policy.usc.edu/scampus-part-c/>). Students are not permitted to create their own class recordings without the instructor’s permission. Violations of these policies will be met with the appropriate disciplinary sanction.

Netiquette Policies:

- Be respectful and considerate towards each other. Pay attention to cultural and background differences.
- While in Zoom sessions please show your complete first name and last name, as in the roster.
- While in Zoom sessions please have your VIDEO ON. It assists in building a sense of community. However, if Internet signal issues preclude you from having a reasonable audio, you may wish to “split” your signal by connecting with both your phone (for audio) and computer (for video).
- While in Zoom sessions be respectful to the learning environment such as: be attentive, have proper background, follow business casual norms, and be in a stationary location (not driving). Please reduce distractions for others.

Support Systems:*The Office of Disability Services and Programs*

Provides certification for students with disabilities and helps arrange relevant accommodations.

<http://dsp.usc.edu>

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. <https://engemannshc.usc.edu/counseling/>

National Suicide Prevention Lifeline - 1-800-273-8255

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

<https://engemannshc.usc.edu/rsvp/>

Sexual Assault Resource Center

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

Office of Equity and Diversity (OED)/Title IX Compliance – (213) 740-5086

Works with faculty, staff, visitors, applicants, and students around issues of protected class.

<https://equity.usc.edu/>

Bias Assessment Response and Support

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. <https://studentaffairs.usc.edu/bias-assessment-response-support/>

Student Support and Advocacy – (213) 821-4710

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

Diversity at USC

Information on events, programs and training, the Diversity Task Force (including representatives for each school), chronology, participation, and various resources for students. <https://diversity.usc.edu/>

USC Emergency Information

Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, <http://emergency.usc.edu>

USC Department of Public Safety – 213-740-4321 (UPC) and 323-442-1000 (HSC) for 24-hour emergency assistance or to report a crime.

Provides overall safety to USC community. <http://dps.usc.edu>



ISE 225: ENGINEERING STATISTICS

Spring 2022

Mondays and Wednesdays 12:00-1:50PM

Victoria Stodden (Professor)

stodden@usc.edu

Office hours: Mondays and Wednesdays 2-3PM on zoom and by appointment
Weekdays drop in hours 6-7PM (about 90% of the time)

Zhengqi Wu (Teaching Assistant)

zhengqi@usc.edu

Office hours: Wednesdays and Thursdays 4-6PM on zoom

Mariah Burdette (Teaching Assistant)

mlburdet@usc.edu

Office hours: Fridays 1-2PM on zoom

COURSE DESCRIPTION AND OBJECTIVES

Prerequisite: ISE 220 Probability Concepts in Engineering

Objective: In this course, you will develop the skills necessary to gather data that is representative of the phenomenon that you intend to study, to describe the uncertainty in scientific results, and to draw valid conclusions in the face of uncertainty. Key topics include: Descriptive statistics; Sampling distributions; Confidence intervals; Hypothesis tests; ANOVA; Linear regression; Use of statistical software; Professional Ethics.

COURSE MATERIALS

All course materials will be available on Blackboard. A reminder that you are responsible for taking notes in class and attending class.

Our primary resource is (OIS) OpenIntro Statistics, Fourth Edition, by David Diez, Mine Çetinkaya-Rundel, Christopher D Barr (2019). Available in blackboard and can be downloaded for free here: <https://leanpub.com/os>

Supplemental resource: (ISLR) An Introduction to Statistical Learning with Applications in R, Second Edition, by James, Witten, Hastie, and Tibshirani (2021).

Available in blackboard and can be downloaded for free here:

https://hastie.su.domains/ISLR2/ISLRv2_website.pdf

Supplemental R resources (free download):

- P. Kuhnert & B. Venables, An Introduction to R: Software for Statistical Modeling & Computing http://cran.r-project.org/doc/contrib/Kuhnert+Venables-R_Course_Notes.zip
- J.H. Maindonald, Using R for Data Analysis and Graphics <http://cran.r-project.org/doc/contrib/usingR.pdf>
- W.J. Owen, The R Guide <http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf>
- W.N. Venables & D. M. Smith, An Introduction to R <http://cran.r-project.org/doc/manuals/R-intro.pdf>

GRADING

30% Homework

30% Midterm Exam (March 9)

40% Final Exam (May 6)

Homework will be assigned and due weekly. Your two lowest homework scores will be dropped. No late assignments will be accepted. Class participation will allow me to round up (e.g. from A- to A) if your final score is close. You must be here for the final exam - no travel exceptions will be given. Accommodations are expected to go through the Office of Student Accessibility Services (OSAS). Please let me know of any accommodations at the start of the semester.

All grades changes must be done by Prof. Stodden. Submit your work to her with a separate written description of why you think the grade you received is incorrect.

HOMEWORKS

All homeworks must be turned in in blackboard by the deadline. For computational questions expect us to rerun your computations to produce your tests: after HW2 you will turn in an .Rmd file and the corresponding pdf containing a narrative answering the homework questions, along with code and results. You must write your own code.

SCHEDULE AND READINGS

Subject to Change

Week of	Topic	Reading
---------	-------	---------

January 10	Intro to R / Probability Review	OIS ch. 1-4, ISLR 2.3
January 17	Sampling and Foundations for inference: Point estimates and sampling variability	OIS 1.3, 5.1
January 24	Foundations for inference: Confidence intervals for a proportion; Hypothesis testing for a proportion	OIS 5.2, 5.3
January 31	Inference for categorical data: Inference for a single proportion; Difference of two proportions	OIS 6.1, 6.2
February 7	Inference for categorical data: Testing for goodness of fit using chi-square; Testing for independence in two-way tables	OIS 6.3, 6.4
February 14	Inference for numerical data: One-sample means with the t-distribution; Paired data	OIS 7.1, 7.2
February 21	Difference of two means; Power calculations for a difference of means	OIS 7.3, 7.4
February 28	Comparing many means with ANOVA	OIS 7.5
March 7	Review for exam (March 7) Midterm March 9 (possibly take home)	

March 14	SPRING RECESS	
March 21	Linear regression: Fitting a line, residuals, and correlation; Least squares regression	OIS 8.1, 8.2
March 28	Linear Regression Diagnostics: Types of outliers in linear regression; Inference for linear regression	OIS 8.3, 8.4; ISLR 3.1
April 4	Multiple regression; Model selection	OIS 9.1, 9.2; ISLR 3.2
April 11	Multiple regression diagnostics: Checking model conditions using graphs; Multiple regression case study: Mario Kart	OIS 9.3, 9.4
April 18	Logistic Regression; Bias-Variance Tradeoff	OIS 9.5, ISLR 2.2.2
April 25	Review for exam and crumple zone. Possibility of a take home final exam.	
May 6 11am-1pm	Final Exam (university scheduled)	

ACADEMIC INTEGRITY

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” <https://policy.usc.edu/scampus-part-b/>.

Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific

misconduct, <http://policy.usc.edu/scientific-misconduct>.

COVID-19 Statement:

In keeping with University policy, all students are required to engage in appropriate behavior to protect the health and safety of our community. If you feel ill or are unable to come to class or complete class assignments due to issues related to COVID-19, including but not limited to: testing positive yourself, feeling ill, caring for a family member with COVID-19, or having unexpected child-care obligations, contact the instructor immediately.

Intellectual Property Policies:

Any misuse, inappropriate dissemination, or attempted sale of class recordings and handouts, as well the appropriation of intellectual property is not acceptable. It is the student's responsibility to appropriately use and handle these recordings under existing campus policies regarding class notes (<https://policy.usc.edu/scampus-part-c/>).

Students are not permitted to create their own class recordings without the instructor's permission. Violations of these policies will be met with the appropriate disciplinary sanction.

Netiquette Policies:

- Be respectful and considerate towards each other. Pay attention to the cultural and background differences.
- While in Zoom sessions please show your complete first name and last name, as in the roster.
- While in Zoom sessions please have your VIDEO ON. It assists in building a sense of community. However, if Internet signal issues preclude you from having a reasonable audio, you may wish to "split" your signal by connecting with both your phone (for audio) and computer (for video)
- While in Zoom sessions be respectful to the LEARNING ENVIRONMENT such as: be attentive, have proper background, follow business casual norms, and be in a stationary location (not driving). Please reduce distractions for others.

Support Systems:

The Office of Student Accessibility Services (OSAS)
Provides certification for students with disabilities and helps arrange relevant accommodations.
<http://dsp.usc.edu>

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.
<https://engemannshc.usc.edu/counseling/>

National Suicide Prevention Lifeline - 1-800-273-8255
Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.
<https://engemannshc.usc.edu/rsvp/>

Sexual Assault Resource Center
For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

Office of Equity and Diversity (OED)/Title IX Compliance – (213) 740-5086
Works with faculty, staff, visitors, applicants, and students around issues of protected class.
<https://equity.usc.edu/>

Bias Assessment Response and Support
Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response.
<https://studentaffairs.usc.edu/bias-assessment-response-support/>

Student Support and Advocacy – (213) 821-4710
Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

USC Emergency Information
Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, <http://emergency.usc.edu>
USC Department of Public Safety – 213-740-4321 (UPC) and 323-4