Module 5: Sampling

**Lesson 5: Sampling Ethics and Equity**

**Estimated Time**: 4.5 hours (plus final assessment)

**Concepts:** data privacy; de-identification; respondent burden; external validity; race and equity

**Lesson Description**: This lesson examines working with survey data and sample estimates in and ethical and equitable manner. We conclude with the final assessment.

**Instructor Preparation**: Look over slides, read core texts, review literature on Canadian and US census and race

|  |  |
| --- | --- |
| **Materials and Resources** | **Learning Goals** |
| 1. Slides 2. Core texts    1. Salganik, M. (2019). Understanding and managing informational risk. In *Bit by bit: Social research in the Digital age* (pp. 307–314). Chapter, Princeton University Press.    2. Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., OBrien, D.R., Steinke, T., & Vadhan, S. (2018). [Differential privacy: A primer for a non-technical audience](https://salil.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience). *Vanderbilt Journal of Entertainment & Technology Law,* 21(1) 209-275.    3. Peter Bacchetti, Leslie E. Wolf, Mark R. Segal, Charles E. McCulloch, Ethics and Sample Size, *American Journal of Epidemiology*, Volume 161, Issue 2, 15 January 2005, Pages 105–110, <https://doi.org/10.1093/aje/kwi014>    4. Patrick Dattalo, Ethical Dilemmas in Sampling, *Journal of Social Work Values and Ethics*, Volume 7, Issue 1, 2010, <https://jswve.org/download/2010-1/2dattalo-Ethical-dilemmas-in-sampling.pdf>    5. Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Developmental review : DR*, *33*(4), 357–370. <https://doi.org/10.1016/j.dr.2013.08.003>    6. James, A. (2008). Making Sense of Race and Racial Classification. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, white methods : Racism and methodology* (pp. 31-45). Rowman & Littlefield Publishers.    7. Harnois, C. (2018). Analyzing race and ethnicity with the GSS. In *Analyzing inequalities: An introduction to race, class, gender, and sexuality using the general social survey* (pp. 65-96). SAGE Publications, Inc., <https://dx.doi.org/>    8. Ashok, S. (2016, August 27). The rise of the American ‘Others’. The Atlantic. <https://www.theatlantic.com/politics/archive/2016/08/the-rise-of-the-others/497690/>    9. Krause, H. (2021, August 27). We need to fill in the blanks in our social identity data. *We All Count*. <https://weallcount.com/2021/08/27/we-need-to-fill-in-the-blanks-in-our-social-identity-data/>    10. Krause, H. (2020, December 4). No one is an asterisk. *We All Count*. <https://weallcount.com/2020/12/04/no-one-is-an-asterisk/>    11. Krause, H. (2020, June 26). Proxy variables part 2: race. *We All Count*. https://weallcount.com/2020/06/26/proxy-variables-part-2-race/ | * Understand how and why differential privacy techniques are implemented for large data sets * Identify ethical and unethical approaches to working with estimates derived from samples * Identify differing approaches to asking about racial characteristics in surveys and working with resulting data |

|  |  |  |
| --- | --- | --- |
| **Time** | **Lesson Content** | **Instructor Notes** |
| 15 minutes  60 minutes | **14-Differential privacy**  **Introduction**   1. Review previous day’s material 2. Have students skim the intro to [Differential privacy: A primer for a non-technical audience](https://salil.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience)   **Lesson**   1. [14-Differential privacy-slides](https://github.com/UofT-DSI/sampling/blob/main/lectures/14-Differential%20privacy-slides.pptx)    1. **Slide 8**: video by Simply Explained – https://www.youtube.com/watch?v=gI0wk1CXlsQ&t=1s |  |
| 20 minutes  30 minutes  75 minutes | **15-Ethics**  **Introduction**   1. Distribute and have students skim over texts for this section    1. Peter Bacchetti, Leslie E. Wolf, Mark R. Segal, Charles E. McCulloch, Ethics and Sample Size, *American Journal of Epidemiology*, Volume 161, Issue 2, 15 January 2005, Pages 105–110, <https://doi.org/10.1093/aje/kwi014>    2. Patrick Dattalo, Ethical Dilemmas in Sampling, *Journal of Social Work Values and Ethics*, Volume 7, Issue 1, 2010, <https://www.jswve.org/wp-content/uploads/2010/01/10-007-108-JSWVE-2010.pdf>    3. Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Developmental review : DR*, *33*(4), 357–370. <https://doi.org/10.1016/j.dr.2013.08.003>   **Lesson**   1. [15-Ethics-slides](https://github.com/UofT-DSI/sampling/blob/main/lectures/15-Ethics-slides.pptx)   **Discussion/Activity**   1. Students should read the following articles (skim the original paper, particularly the methods section, and then read the commentary article):  * *Forecasting elections with non-representative polls*   + Wang, W. et al., Forecasting elections with non-representative polls, International Journal of Forecasting (2014),<https://doi.org/10.1016/j.ijforecast.2014.06.001>   + Commentary: A review of “Forecasting elections with non-representative polls, by Rohan Alexander:<https://rohanalexander.com/posts/2020-02-11-a-review-of-forecasting-elections-with-non-representative-polls/> * COVID-19 antibody seroprevalence in Santa Clara County   + Eran Bendavid, Bianca Mulaney, Neeraj Sood, Soleil Shah, Rebecca Bromley-Dulfano, Cara Lai, Zoe Weissberg, Rodrigo Saavedra-Walker, Jim Tedrow, Andrew Bogan, Thomas Kupiec, Daniel Eichner, Ribhav Gupta, John P A Ioannidis, Jay Bhattacharya, COVID-19 antibody seroprevalence in Santa Clara County, California, *International Journal of Epidemiology*, Volume 50, Issue 2, April 2021, Pages 410–419, <https://doi.org/10.1093/ije/dyab010>   + Commentary: How (Not) to Do an Antibody Survey for SARS-CoV-2, by Catherine Offord:<https://www.the-scientist.com/news-opinion/how-not-to-do-an-antibody-survey-for-sars-cov-2-67488> * Discussion surrounding ethical considerations involved in both studies. Compare and contrast the approaches taken within each study. Consider:   + Sample size   + Sample recruitment   + Respondent burden and potential impacts   + Net burden   + Study value   + Internal validity   + External validity   + Selection bias   + Representation   + Analysis methods   + Potential adverse study outcomes | Potential discussion questions:   * What is the target populations? In what ways were the samples representative or non-representative of the target population? * What was the sampling approach used? How did this approach impact who was able to participate in the survey? * What factors impact the net respondent burden for the survey? * Is the study internally valid? * Is the study externally valid? * Did the researchers note areas of underrepresentation? What analysis methods were used to account for underrepresentation of groups? * What are the potential negative outcomes if the reported results of the study are inaccurate? |
| 30 minutes  40 minutes | **16-Inequity**  **Introduction**   1. Ensure students have access to texts. Give a brief overview of material and time for students to skim/review each text.    1. James, A. (2008). Making Sense of Race and Racial Classification. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, white methods : Racism and methodology* (pp. 31-45). Rowman & Littlefield Publishers.    2. Harnois, C. (2018). Analyzing race and ethnicity with the GSS. In *Analyzing inequalities: An introduction to race, class, gender, and sexuality using the general social survey* (pp. 65-96). SAGE Publications, Inc., <https://dx.doi.org/>    3. Ashok, S. (2016, August 27). The rise of the American ‘Others’. The Atlantic. <https://www.theatlantic.com/politics/archive/2016/08/the-rise-of-the-others/497690/>    4. Krause, H. (2021, August 27). We need to fill in the blanks in our social identity data. *We All Count*. <https://weallcount.com/2021/08/27/we-need-to-fill-in-the-blanks-in-our-social-identity-data/>    5. Krause, H. (2020, December 4). No one is an asterisk. *We All Count*. <https://weallcount.com/2020/12/04/no-one-is-an-asterisk/>    6. Krause, H. (2020, June 26). Proxy variables part 2: race. *We All Count*. https://weallcount.com/2020/06/26/proxy-variables-part-2-race/   **Lesson**   1. [16-Inequity-slides](https://github.com/UofT-DSI/sampling/blob/main/lectures/16-Inequity-slides.pptx)    1. **Slide 4**: Discussion on race. Questions on slide. |  |