**Survey methodology for Carolyn Forestiere’s “Dataprac” Dataset**

Most of Dataprac’s questions are identical to items that appear on the World Values Survey (WVS), which has been administered in successive waves (i.e., the questions have been fielded repeatedly) in over 100 countries since 1981. Dataprac has responses that are more up to date than what was available when the edition of the textbook was written. Specially, data from Wave 7 of the WVS were released after this textbook was completed in 2022, and the US sample was administered in 2017: <https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>. The version of Dataprac at the textbook’s website does not include any variable labels, but your instructor has added those to the version you will be using for this class (and I provided datasets with value labels to the textbook author if she wants to post those for other classes using this book).

The Dataprac survey was fielded specifically to create a sample dataset for Forestiere’s textbook. Its methodology is unusual in a few ways, and the data have been “cleaned” to make the dataset more friendly to students learning how to use statistical analysis programs for the first time. The survey was administered to a US-only, online convenience sample and includes 1031 respondents. In terms of the larger political context, the survey was administered in the summer of 2019, more than a year out from the 2020 contested US presidential election. Unlike the WVS, this survey was not administered to a randomly drawn set of respondents where—at least in theory—every American would have had the same probability of being selected for the study. Instead, the Dataprac survey used an opt-in sample provided by the Qualtrics corporation. This platform, called “qBus,” is an on-line survey omnibus that gives researchers access to low-cost national convenience samples that are recruited through social media and other groups whose demographics and views may, in aggregate, differ from the US population as a whole.

Because the Dataprac survey relies on this type of a “convenience” sample, anyone using this dataset should be aware of the possibility that its respondents could systematically differ in their views relative to what one would have found in the US population as a whole at the same time this survey was fielded. Due to the high costs of fielding a survey using a stratified random sample, datasets like Dataprac are frequently used in political science research, especially for survey experiment studies. However, it is preferable to use a data derived from a high-quality, nationally representative stratified random sample if one is available. In this case, the 2022 wave of the WVS would provide superior data.

Despite the limitations of convenience samples, researchers who have compared survey results for self-recruited on-line panels similar to the one used to create Dataprac with the results of surveys using stratified random samples of the US population mostly find that that these type of surveys produce similar results to best-practice survey methods as long as the convenience sample respondents are matched with quotas that align with the demographic and political makeup of the US population as a whole. Both surveys like those administered by qBus and those using stratified random samples need to include a strategy of dealing with the fact that some types of people are more likely to volunteer to take a survey when asked. Also, the characteristics related to being willing to take a survey may well be connected to some attitudes assessed in that same survey (e.g., trust of others). For stratified random samples, data are typically analyzed with post-hoc survey weights that raise or decrease the influence of different types of individuals, depending on whether they are under- or over-represented in the response pool.

What kind of individuals make up a qBus sample?: Per Ian G. Anson (“Taking The Time? Explaining Effortful Participation Among Low-Cost Online Survey Participants.” *Research and Politics*, July-Sept. 2018: 1–8:

… (Pg. 2) qBus participants are recruited using ‘traditional, actively managed market research panels’ (Qualtrics, 2014: 3) in addition to social media recruitment methods. Much like Survey Sampling International (SSI), another firm that provides access to survey participants, the third-party panels used by Qualtrics have been certified for quality by Mktg Inc.’s Grand Mean Certification Program.

… Qualtrics selects adult panelists residing in the US to participate in surveys. They are sampled on the basis of demographic quotas; this demographic information is collected through initial screening surveys. Panelists are contacted by Qualtrics via email or social media accounts, inviting them to participate in a survey for research purposes only. The Qualtrics documentation states that respondents’ rewards for survey participation include “airline miles, gift cards, redeemable points, sweepstakes entrance and vouchers” (Qualtrics, 2014: 5).

… In addition, Qualtrics seeks to limit frequency of participation by ensuring that historical records are maintained for each panelist. This reduces the ability of qBus participants to act like “professional panelists,” as the number of invitations they receive is necessarily limited.

… Importantly, Qualtrics makes a concerted effort to monitor and safeguard against respondent inattentiveness and lack of effort… [And] qBus removes respondents who exhibit evidence of extreme “speeding” behavior (answering survey questions too quickly).

One other important difference between the Dataprac survey and most stratified random samples is that in order to be compensated, it looks like respondents had to complete each item in the survey (alternatively, it may be that the survey’s administrators have removed any respondent who had not answered items, which is not something we would see in most datasets). And in most cases, respondents don’t appear to have had the option of either not answering a particular item or responding “don’t know.” This approach has the advantage of producing a dataset with no missing responses; however, not allowing a respondent to refuse to answer any item may have induced inaccurate answers if respondents truly didn’t have an opinion for some questions (e.g., where they sit on a 1-10 political ideology scale) or to give socially desirable, but inaccurate response to items they might otherwise have skipped.