**12. A Project in Weighting a Personnel Survey**

In this project you will develop survey weights and deliver an analysis file for a survey of military personnel. Members of the military reserves were asked a variety of questions about job satisfaction. Some examples of the questions are:

* Suppose that you have to decide whether to continue to participate in the National Guard/Reserve. Assuming you could stay, how likely is it that you would choose to do so?
* Overall, how would you rate the current level of stress in your personal life?
* Taking all things into consideration, how satisfied are you, in general, with each of the following aspects of being in the National Guard/Reserve?

The type of work you do in your military job

Your total compensation (i.e., base pay, allowances, and bonuses)

The data file includes records for all persons who were in the initial sample—respondents, nonrespondents, and ineligibles. There are also several demographic variables from administrative record files for each sample person. The files to be used are listed at the end of this chapter.

The following tasks have yet to be completed and are assigned to your team. Each task should be documented in the project final report; be sure to justify the decisions your team has made.

1. Develop the design weights (inverse of the selection probabilities) for this single-stage stratified simple random sampling design and verify your calculations. The field STRATUM defines the sample design strata. Each record contains counts of the number of persons in the population (NSTRAT) and the sample (NSAMP) in the design stratum to which the record belongs. The field V\_STRAT identifies design strata that were collapsed together for variance estimation. Note that, if a population count is needed for a variance stratum, the values of NSTRAT will need to be summed for the design strata that are combined into a V\_STRAT.
2. Specify how you will classify the various response status codes (RESPSTAT) into the general categories—eligible respondent, eligible nonrespondent, ineligible, or unknown—described in Chapter 6. The variable values and value labels for RESPSTAT are provided in the section below, *Data Files and Other Information*.
3. Apply weight adjustments to the design weights and verify your calculations. You should include the adjustment methods we have discussed in class—unknown eligibility, nonresponse, and calibration. In the case of either unknown eligibility or nonresponse adjustments, compare weighting cell and propensity adjustments by actually carrying out your own implementation of each method. You may encounter some cases, in either the data file for respondents or the file for population counts, that have missing data for fields that you would like to use in weighting. If so, you need to explain how you handled those in the various steps used in weighting.
4. Prepare an analysis file (in SAS, Stata, or text format) containing the variables from the original data file (*SOFR.sas7bdat*), the base weights, the final analysis weights (you may choose only one set from task 3 above), and any adjustments applied to the design weights to create the final weights. Additionally, create any necessary indicators you would need to analyze the questionnaire responses and eliminate any unnecessary data records. All variables must have a descriptive label. For any newly created categorical variable, provide a description of the variable values in the report.
5. Using your final analysis weights, tabulate the proportions of personnel who are
   1. Dissatisfied or very dissatisfied with their total compensation (RA006A)
   2. Very unlikely or unlikely to stay in the Reserves (RA008)

Make these tabulations separately for each service and for enlisted personnel and officers. Include the point estimates of proportions and standard errors. Describe the method you use for standard error estimation and any limitations that the method may have.

1. Include a description for data users of which cases and weights should be used for various types of data analyses. Provide some brief examples of software code that would be used to estimate means or proportions associated with a typical questionnaire item. Examples should be given for at least two software packages. Your report should describe how the software must be used in order to account for weights and design features like strata.

**Contents of the Weighting Report**

Below is a list of topic areas that should be included in your weighting report. Questions and suggestions are included in each section to assist with the development of the text. The order of the sections in your report does not have to be the same as that given below. You should construct your report in a way that presents topics in an order that seems logical to your team.

The report should be written to a client whose staff includes managers and technical personnel. Managers will be more interested in understanding the broad outline of the steps used in weighting. Technical personnel will be interested in understanding the details of weight computation, including appropriate formulae, and in being able to appropriately analyze the data. You should consider how to structure your report to serve these audiences.

**Topic Areas for Weighting Report**

Title Page (project title, date of submission, and name of project contact person)

Introduction (overview of the document)

Study Weights

* Brief discussion of sampling design
* Methods to calculate design weights
* Types of weight adjustments and why they were used. Comparison of adjustments.
* Evaluation of weights and methods used to check or compare calculations

Analysis File

* Summary of analysis file contents (include PROC CONTENTS or the equivalent in an appendix)
* Variables of interest

References

Appendix

* PROC CONTENTS or codebook of data file

**Data Files and Other Information**

SOFR codebook.pdf—code values for each variable in the SOFR.sas7bdat data file.

***Need to add some codebook information for the variables on the file. Otherwise—mass confusion. Note that some have been imputed—CHANGE THE LABELS OF THOSE***

RCCPDS57 codebook.pdf—code values for each variable in the RCCPDS57.sas7bdat data file.

Annotated questionnaire.pdf—The survey questionnaire with annotations showing variable names and code values for all questions. Note that the data file for this project contains only a subset of the questions in the survey. Also, some questions have been recoded to have different names and fewer values in the data file than are on the questionnaire.

SOFR.sas7bdat—edited data file from the survey in SAS version 9 format. The same data is in the SAS transport file, SOFR.xpt.

RCCPDS57.sas7bdat—file of population counts. The same data are in the SAS transport file, RCCPDS57.xpt.

This file is the result of matching the sampling frame to the most current personnel file available as of the start of the data collection period. The personnel file consists of all persons on the payroll as of the date the file was constructed. Thus, these counts should cover only eligible cases. The labels for the field names contain the name of the variable in sofr.sas7bdat to which the counts correspond.

formats.sas7bcat—format library for both SAS data files

To access this library in a SAS program include the following type of libname statement:

LIBNAME library “C:\PracTools\”;

To be sure that SAS searches that library for formats use

options fmtsearch=(library)

The folder name PracTools should be changed to the location where you save the format file. This format library will give access to the variable and value labels for the fields in sofr.sas7bdat.

Variable Values and Value Labels for the RESPSTAT variable

|  |  |
| --- | --- |
| 1 = | Questionnaire Returned – Completed |
| 2 = | Questionnaire Returned – (Sufficient) Partial Complete |
| 3 = | Questionnaire Returned – (Insufficient) Partial Complete |
| 4 = | Questionnaire Returned – Ineligible |
| 5 = | Questionnaire Returned – Blank |
| 18 = | No Return – Deceased |
| 19 = | No Return – Incarcerated |
| 22 = | No Return – Separated/Retired |
| 23 = | No Return – Active Refusal |
| 25 = | No Return – Other |
| 26 = | No Return – Eligible based on administrative records |
| 27 = | Postal Non-delivery |
| 29 = | Not Locatable |
| 35 = | Ineligible – No Questionnaire Sent |