**Lab 9. Examining documentation on weighting and variance estimation**

**MSDS 6370**

**Submitted by Hari Narayan Sanadhya**

**Objective:**

* For the student to learn more about weighting of surveys with complex sample designs.
* For the student to become familiar with the documentation for the weighting of large surveys that provide identify-protected data files available to the public.

Look up the documentation from surveys that interests you.

Here are some examples:

1. National crime victimization survey (<http://www.bjs.gov/content/pub/pdf/ncvstd13.pdf>)
2. National household travel survey (<http://nhts.ornl.gov/2009/pub/UsersGuideV2.pdf>)
3. National health information survey (<ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2013/srvydesc.pdf>)
4. Current population survey (<https://www.census.gov/prod/2006pubs/tp-66.pdf>)

Answer as many of the following questions as you can from the documentation.

1. Are the weights provided by the data producer for your survey a composite of several weight adjustments? If yes, what are they?

**National crime victimization survey** – A single combined weight is provided for each unit i.e. personal and household victimization datasets. The weights provided are adjusted to account for non-responsiveness, design error and sampling (under coverage) error. Household, person, and victimization data from the NCVS sample are adjusted to give quarterly and annual estimates of crime experienced by the U.S. household population age 12 or older.2 Adjustments are first made to account for non-interview followed by adjustment to reduce bias caused due to under coverage. This is done by reducing the variance of the estimate by correcting for differences between the distribution of the sample by age, sex, and race and the distribution of the population by these characteristics.

**National household travel survey**: Individual and Composite weights are provided. Weights are provided for person and household datasets for all “useable” households which are defined as households in which retrieval interviews were completed with at least 50 percent of the adults in the household. Other weights included are initial household weights which are reciprocal of the known probability of selection, Travel Day Weight which simply equals the final useable household person weight multiplied by 365, Vehicle Weights which is the weight due to the useable household-level and a set of 100 replicate weights provided for each level (household, person, special child, vehicle) generated using jackknife procedure.

**National health information survey**: Combined weights are provided. The base weight is provided at both household and personal level which is the product of inverses of the probability of selection at each sampling stage. The base weights are then adjusted for not-response and ratio adjusted to create final sampling weights. The person level weights are then finally adjusted according to the quarterly poststratification by age/sex/race/ethnicity classes based on population estimates produced by US census bureau.

**Current population survey**: Combined weights are provided. The survey by design is self-weighting i.e. sampling fraction is same in all strata which means by default, the overall state sampling interval is fixed. Statistical weight for each person in the sample is then computed by performing:

• Adjustment for nonresponse.

• adjustment to reduce variances due to PSU sampling error

• National and state coverage adjustments

• adjustment to reduce variances by controlling CPS estimates of the population to independent estimates of the current population.

• Composite estimation using estimates from previous months to reduce the variances.

• Seasonally-adjusted estimates for key labor force statistics.

In addition to these, additional weighting steps were produced to get to the final estimate including:

• household and family weights.

• outgoing rotation group’s weights.

• veterans’ weights.

• longitudinal weights.

2. Are there more than a single set of weights, depending on the units of analysis?

**National crime victimization survey** – A single combined weight (obtained after all the required adjustments) is provided for each unit i.e. personal and household victimization datasets.

**National household travel survey** – yes more than a single set of weights are provided which are to be used depending on the unit of analysis. Weights provided include Household Weights, Vehicle Weights, Travel day Weights, Person weights and replication weights for household, person and travel days.

**National health information survey** – Two sets of weights (Weight - Final Annual and Weight - Interim Annual) are provided for each unit i.e. for each set of analysis. This includes Person or Person-Level Analyses, Sample Child or Sample Adult Analyses, Household or Household-Level Analyses, Family or Family level analysis

**Current population survey** – Separate sets of weights for household and family weights are provided which are the weights assigned from the householder or reference person after all adjustments have been made and should be used when tabulating estimates of families-households.

3. If there is a nonresponse adjustment, what kinds of characteristics are used to form weight adjustment cells?

**National crime victimization survey** – there are three kinds of nonresponse – household nonresponse, person nonresponse and item nonresponse of which adjustments are made only for household and person nonresponse but not for item nonresponse. For item nonresponse data imputation is performed.

Variables used for household nonresponse are:

• type of living quarters (i.e., a housing unit or not a housing unit)

• CBSA/Metropolitan Statistical Area (MSA) status (part of a principal city within a CBSA/MSA, not part of a principal city but still within a CBSA/MSA, or outside of a CBSA/MSA)

• urban status

• race of the reference person or head of household (white only, all other races and combinations of races).

Variables used for person nonresponse adjustment are region, age, sex, and race.

**National household travel survey** – household interview (screener nonresponse) is performed on the base weight. The variable used for nonresponse adjustment calculation is residency rates computed for two situations for the residential numbers in the sample

1. only ring-no-answers on repeated calls (with no evidence that the number is non-working or business),
2. repeated calls only reach an answering machine.

Also characteristics of the telephone exchanges, such as percentage of listed households by race or by tenure, were examined to determine the difference in response rates.

**National health information survey** – nonresponse adjustments are made on all the set of weights provided in the study. Response rates are provided for total household response rate, conditional response rate for the Family component, conditional response rate for the Family component conditional response rate for the sample Child component and conditional response rate for the sample audit component. But the document does not provide the details regarding the characteristics used to determine these rates.

**Current population survey** – There are 2 types of nonresponses – item nonresponse and household nonresponse and the adjustment in weight is made for household nonresponse using units demographic and socioeconomic characteristics. Item nonresponses are adjusted not in weights but by imputation.

4. Is there any evidence that the survey makes a poststratification adjustment; i.e., weights to some known standard of demographic characteristics? If yes, what standard is used and what variables?

**National crime victimization survey** – yes, there is evidence showing that the final household and person-weights were post stratified adjusted. The standard used were the current forward projections obtained from Census 2000 and control totals were adjusted for both household and person data.

**National household travel survey** – The document provides evidence for post stratification adjustment performed. The document tells that poststratification were preformed to reduce sampling error and bias but does not provide details about the same except for replicate weights which were post stratification adjusted to external control totals (no details here) resulting in decrease in estimated variance obtained.

**National health information survey** – Yes post stratification adjustment were made on final personal level weights. Standard used is US Census Bureau and the variables used were age, sex, race and ethnicity.

**Current population survey** – not able to find any evidence in support of post stratification adjustment.

5. Are replicate weights provided by the data producer? If so, what do they say about the variance estimation method they are implementing?

**National crime victimization survey** – Replicate weights are provided each of the four types of weights (i.e., household, person, incident, and victimization). These weights are obtained by multiplying unbiased weights (base weight \* special weight) with replicate factors and then adjusting it through the non-interview adjustment, the first-stage ratio adjustment and the second-stage ratio adjustments just as the full sample is weighted. The document although specifies about how to calculate the variance using replicate weights but the study is using GVF (Generalized Variance Function) parameters for variance estimation. The reason of doing this is GVFs are easier to use than replicate weights, and they stabilize variance estimates and are computationally more efficient than variances estimated from replicate weights. The methodology involved is to firstly use the final weights to estimate totals of the 43 major crime categories. The totals are estimated for all of the persons and households, and are estimated by a variety of key domains. Overall, hundreds of estimates are calculated. The replicate weights are then used to calculate direct variance estimates for all totals.

**National household travel survey** – a set of 100 replicate weights are provided which are obtained using JACKKNIFE procedure. The unbiased estimator of variance can be obtained by sum of squared differences between these replicate estimates and the full sample weight estimate.

**National health information survey** – Replicate weights are not provided

**Current population survey** – Yes the replicate weights are provided which are obtained by multiplying unbiased weights (base weight \* special weight) with replicate factors and then adjusting it through the non-interview adjustment, the first-stage ratio adjustment, national and state coverage adjustments, the second-stage ratio adjustments, and compositing just as the full sample is weighted. The approach to obtain the variance is provided which is estimated variance for the characteristic of interest is a sum of squared differences between each replicate estimate and full sample estimate, mathematically stated as:

