**W2 Specific Steps**

**Overview**

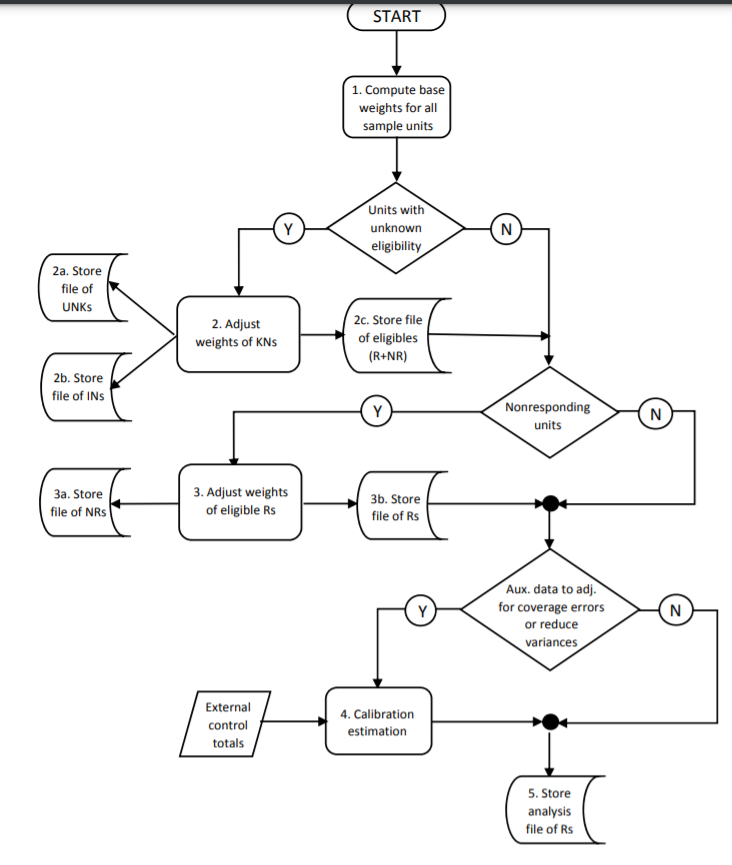
Probability Samples: Four Steps in weighting

• compute base weights 🡺 Inverse of selection probabilities; we keep track of those

• adjust base weights to account for units with unknown eligibility (if any)

• adjust for nonresponse

• calibrate to pop control totals



Non-probability samples

• No base weights in probability sampling sense

• Identify ineligible units

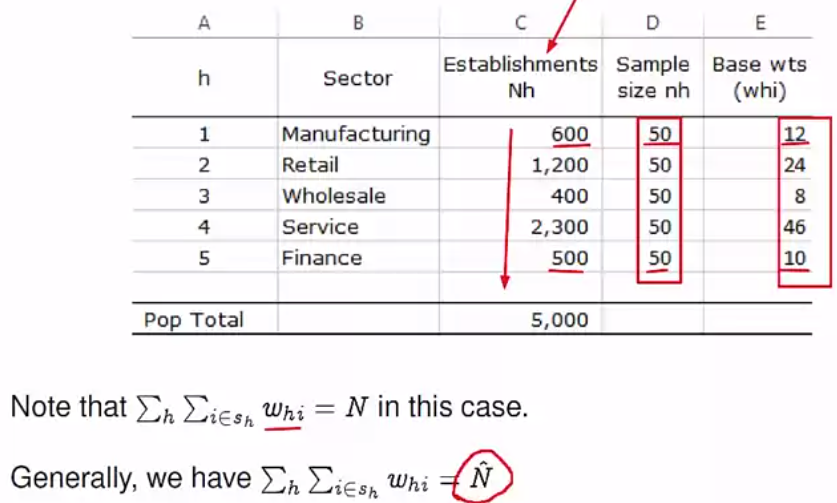
• No nonresponse in probability sampling sense

• Maybe compute a “pseudo-inclusion” probability and use inverse as a “base weight”

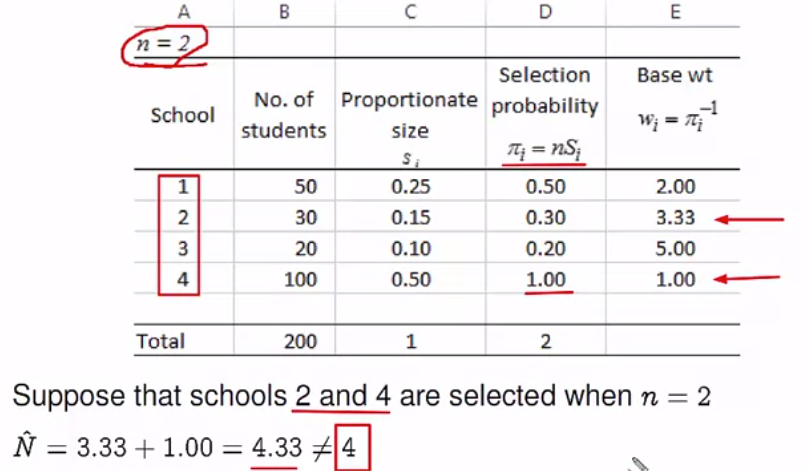
• Calibrate to population control totals

**Base Weights**

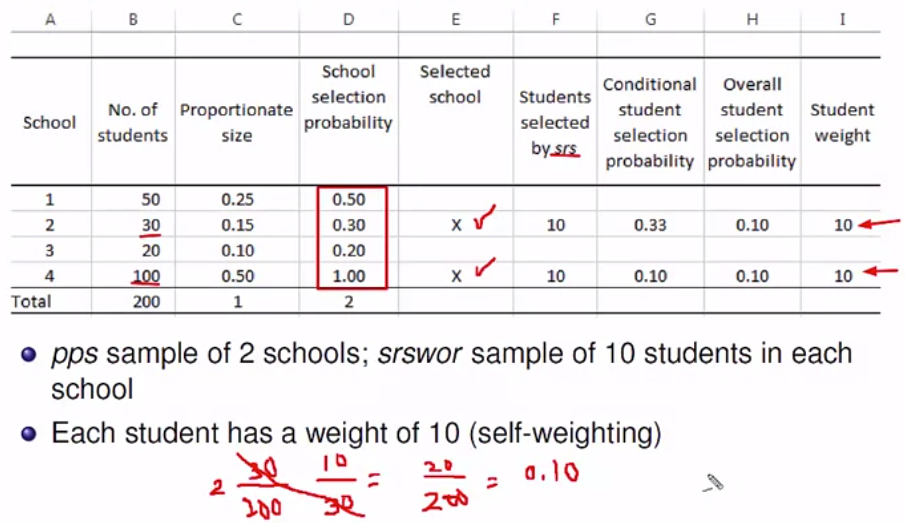
Example: Stratified simple random sample



Example: Probability Proportional to size sample



Example: Two stage sample



🡺 Self weighting sample

If you think there is no reason to think certain student is worth more

• In a stratified simple random sample, every sample unit in a given stratum **cannot** have a different base weight.

**Nonresponse Adjustments**

• Missing completely at random (MCAR)

• Missing at random (MAR)

• Nonignorable NR (NINR)

Whether a unit responds or not is treated a random event when response is categorized as one of these

Responding could be considered deterministic, i.e. a unit is guaranteed to respond or not. But, random or “stochastic” response is the formulation behind the NR adjustments used.

Missing Data Mechanisms-defined

• MCAR

- every unit has same probability of responding 🡺 responding is just an extra stage of Bernoulli sampling

- no weight adjustment needed for means; one overall adjustment needed for totals

• MAR

- probability of responding depends on covariates

- adjustment possible if covariates known for both Rs and NRs

- e.g. Suppose that it is known that response rate in a survey of schools depends on a measure of the socioeconomic level of the neighborhood where the school is located and that this level is known for every school in the frame. This is an example of MAR

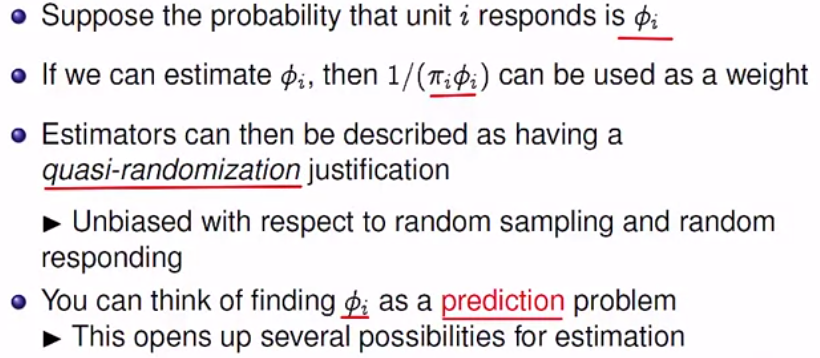
• NINR

- probability of responding depends on analytic variables (y’s) and possibly covariates

- adjustment difficult or impossible

- A survey of schools will be done to assess the extent to which computers are used in teaching mathematics. Suppose that the response rate in the survey depends on both a measure of the socioeconomic level of the neighborhood where the school is located and on whether a school has computers available for students. This is an example of NINR.

Missing data mechanisms-defined



**Response Propensities**

General Procedure

• Estimate response probabilities (propensities) for each R and NR

- regress binary response variable on covariates

- logistic regression typically used

• Form groups (cells) for NR adjustment

- Sort Rs and NRs from low to high by estimated propensity

- Divide file into groups

- 5 groups is popular but more can be used, especially if sample is large

• Use of NR adjustment within each group

- A single adjustment smooths out effects of any extreme propensities produced by the binary regression

- options within a cell are unweighted RR, weighted RR, average propensity, median propensity

- options will be similar if range of propensities in a cell is not large