II. Likelihood in statistical inference and survey sampling

- Problems with design-based inference
- Likelihood principle, conditionality principle and sufficiency principle
- Likelihood and likelihood principle in survey sampling
- **Underpins model-based inference to sample survey**

Traditional approach

Design-based inference

- Population (Target population): The universe of all units of interest for a certain study: U = {1,2, …, N}
  - All units can be identified and labeled
  - Variable of interest y with population values $\mathbf{y} = (y_1, y_2, ..., y_N)$
  - Typical problem: Estimate total t or population mean t/N
- Sample: A subset s of the population, to be observed
- Sampling design $p(s)$ of all possible subsets;
  - The probability distribution of the stochastic sample

Problems with design-based inference

- Generally: Design-based inference is with respect to *hypothetical* replications of sampling for a *fixed* population vector **y**
- The basis of inference is the known sampling distribution *p(s)* **---** also referred to as **frequentistic** approach
- However, sampling variance may fail to reflect – in an intuitive manner – the information in a *given sample*

Problem with design-based variance measure
Illustration 1

a) $N+1$ possible samples: $\{1\}$, $\{2\}$,…,$\{N\}$, $\{1,2,\ldots N\}$

b) Sampling design: $p(\{i\}) = 1/2N$, for $i = 1,..,N$;
$p(\{1,2,\ldots N\}) = 1/2$

c) Use $\bar{y}_s$ as estimator for population mean μ

$$\text{Unbiased}: E(\bar{y}_s) = \sum_s p(s)\bar{y}_s = \sum_{i=1}^N \frac{1}{2N} y_i + \frac{1}{2}\mu = \mu$$

Design-variance:

$$Var(\bar{y}_s) = E(\bar{y}_s - \mu)^2 = \sum_{i=1}^N (y_i - \mu)^2 \cdot \frac{1}{2N} = \frac{1}{2} \cdot \tilde{\sigma}^2$$

d) Assume we select the "sample" $\{1,2,\ldots,N\}$. Then the "actual" precision (of census) is not $\tilde{\sigma}^2/2$

Problem with design-based variance measure Illustration 2

a) Expert 1: SRS and estimate $\bar{y}_s$

Precision is measured by $(1 - f)\dfrac{\sigma^2}{n}$

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2, \; f = n/N$$

b) Expert 2: SRS with replacement and estimate $\bar{y}_s$

measures precision by $\tilde{\sigma}^2 / n$

In case both experts select the <u>same</u> sample (with no duplicates), compute the <u>same</u> estimate, but give <u>different</u> measures of precision…

The likelihood principle, LP

Model: $X \sim f_\theta(x), \theta \in \Omega; \theta$ are the unknown parameters in the model

- The likelihood function, with *data x*: $l_x(\theta) = f_x(\theta)$

Measures the likelihood of different θ values given the data x

l is conceptually different to f ---

x is fixed and θ varies in l; θ is fixed x varies in f

- Likelihood Axiom: The likelihood function contains all information about the unknown parameters
- LP: Two proportional likelihood functions for θ, from the same or different experiments, should give identically the same statistical inference

- Maximum likelihood estimation satisfies LP, using the curvature of the likelihood as a measure of precision (Fisher)
- LP is controversial, but hard to argue against because of the fundamental result by Birnbaum, 1962 (Theorem)
- LP follows from sufficiency (SP) and conditionality principles (CP) that are found to be intuitive:
    - SP: Statistical inference should be based on sufficient statistics
    - CP: If you have 2 possible experiments and choose one at random, the inference should depend only on the chosen experiment [NB. There exits other formulations.]

Illustration of CP

- A choice is to be made between a census og taking a sample of size 1. Each with probability ½.

- Census is chosen

- Unconditional approach:

$$\pi_i = P(census) + P(\text{sample of size 1 and } i \text{ is selected})$$
$$= 1/2 + P(\text{sample of size 1})P(i \text{ is selected} \mid \text{sample of size 1})$$
$$= 1/2 + \frac{1}{2} \cdot \frac{1}{N} \approx \frac{1}{2}.$$

The Horvitz-Thompson estimator:

$$\hat{t}_{HT} \approx 2\sum_U y_i = 2t\,!$$

Conditional approach: $p_i = 1$ and HT estimate is t

LP, SP and CP

Model: $X \sim f_\theta(x), \theta \in \Omega; \theta$ are the unknown parameters in the model

Experiment is a triple $E = \{X, \theta, \{f_\theta\}, \theta \in \Omega\}$

$I(E,x)$: Inference about θ in the experiment E with observation x

Likelihood principle:

Let $E_1 = \{X_1, \theta, \{f_\theta^1\}\}$ and $E_2 = \{X_2, \theta, \{f_\theta^2\}\}$. Assume

$l_{1,x_1}(\theta) = c l_{2,x_2}(\theta)$, c independent of θ. $(f_\theta^1(x_1) = c f_\theta^2(x_2))$

Then: $I(E_1, x_1) = I(E_2, x_2)$

This includes the case where $E_1 = E_2$ and x_1 and x_2 are two different observations from the *same* experiment

Sufficiency principle: Let T be a sufficient statistics for θ in the experiment E. Assume $T(x_1) = T(x_2)$. Then $I(E, x_1) = \mathrm{I}(E, x_2)$.

Conditionality principle:

Let $E_1 = \{X_1, \theta, \{f_\theta^1\}\}$ and $E_2 = \{X_2, \theta, \{f_\theta^2\}\}$.

Consider the mixture experiment E^* where E_1 is chosen with probability 1/2 and x_1 is observedand E_2 is chosen with probability 1/2 and x_2 is observed. The observation in E^* is then the value of $X^* = (J, X_J)$, $J = 1,2$.

$$E^* = \{X^*, \theta, \{f_\theta^*\}\} \text{ where } f_\theta^*(j, x_j) = \tfrac{1}{2} f_\theta^j(x_j)$$

$$\mathrm{CP}: I(E^*(j, x_j)) = I(E_j, x_j)$$

$$\text{Theorem: CP and SP} \Leftrightarrow \text{LP}$$

Consequences for statistical analysis

- Statistical analysis given the observed data: the sampling distribution *p(s)* is irrelevant if it does not depend on θ.
- Standard inference procedures such as confidence intervals and P-values may be inflict with the LP
- History and dicussion after Birnbaum, 1962: An overview in ”*Breakthroughs in Statistics,1890-1989, Springer 1991*”

Illustration- Bernoulli trials

$X_1, ..., X_i, ..$

$X_i = 1$ (success) with probability θ

Two experiments to gain information about θ:

$E_1 : n = 12$ observations and observe $Y_1 = \sum_{i=1}^{12} X_i$

E_2 : Continue trials until we get 3 failures (0's) and observe Y_2 = number of successes

Suppose the results are $y_1 = y_2 = 9$

The likelihood functions:

$$l_9^{(1)}(\theta) = \binom{12}{9}\theta^9(1-\theta)^3 \qquad \text{binomial}$$

$$l_9^{(2)}(\theta) = \binom{11}{9}\theta^9(1-\theta)^3 \qquad \text{negative binomial}$$

Proportional likelihoods: $l_9^{(2)}(\theta) = (1/4)l_9^{(1)}(\theta)$

LP: Inference about q should be identical in the two cases

Frequentistic procedures give different results:

F.ex. test $H_0 : \theta = 1/2$ against $H_1 : \theta > 1/2$

$(E_1, 9) : \text{P-value} = 0.0730 \qquad (E_2, 9) : \text{P-value} = 0.0327$

because different sample spaces: (0,1,..,12) and (0,1,...)

Likelihood function in design-based inference

- Unknown parameter: $\mathbf{y} = (y_1, y_2 ..., y_N)$
- Data: $x = \{(i, y_{obs,i}) : i \in s\}$
- Likelihood of parameter $\mathbf{y} = (y_1, y_2 ..., y_N)$
- Compatible parameters with the observed sample:

$$\Omega_x = \{\mathbf{y} : y_i = y_{obs,i} \text{ for } i \in s\}$$

- Sampling design: $p(s)$
- Likelihood function: $l_x(\mathbf{y}) = \begin{cases} p(s) \text{ if } \mathbf{y} \in \Omega_x \\ 0 \text{ otherwise} \end{cases}$
- All compatible **y** equally likely !!

- Likehood principle, LP : The likelihood function contains all information about the unknown parameters
- According to LP:
  - The design-model is such that the data contains no information about the unobserved part of $\mathbf{y}$, $\mathbf{y}_{unobs}$
  - One has to assume in advance that there is a relation between the data and $\mathbf{y}_{unobs}$:
    - As a consequence of LP: Necessary to assume a model
  - The sampling design is irrelevant for statistical inference, because two sampling designs leading to the same *s* will have proportional likelihoods

Let p_0 and p_1 be two sampling designs. Assume we get the same sample s in either case. Then the data x are the same and W_x is the same for both experiments.

The likelihood function for sampling design p_i , i = 0,1:

$$l_{i,x}(\mathbf{y}) = \begin{cases} p_i(s) \text{ if } \mathbf{y} \in \Omega_x \\ 0 \text{ otherwise} \end{cases}$$

$$\Rightarrow l_{1,x}(\mathbf{y}) / l_{0,x}(\mathbf{y}) = p_1(s) / p_0(s) \text{ if } \mathbf{y} \in \Omega_x$$

and then for *all* $\mathbf{y}$:

$$l_{1,x}(\mathbf{y}) = \frac{p_1(s)}{p_0(s)} l_{0,x}(\mathbf{y})$$

- Same inference under the two different designs. This is in direct opposition to usual design-based inference, where the only stochastic evaluation is with respect to the sampling design, for example the Horvitz-Thompson estimator.
- Concepts like design unbiasedness and design variance are irrelevant for inference under the LP.
- This *does not mean* the sampling design is not important.
  - After all the likelihood depends on the sample data!
  - Are there especially good or bad samples?
  - Sampling needs to be non-informative for inference to be valid under the (correct) population model [an example; more later in connection with treatment of missing data]

Model-based inference

- Assumes a model for the **y** vector
- Conditioning on the actual sample
- Use model as the basis of statistical inference
- **Problem:** dependence on model
  - Introduces a subjective element
  - impractical to model all variables in a survey
- Design approach is "objective" in a perfect world without nonsampling errors

III. Model-based and model-assisted approaches in survey sampling

- Model-based approach. Also called the prediction approach
  - Assumes a model for the **y** vector
  - Use modeling to construct estimator
  - Ex: ratio estimator
- Model-based inference given non-informative sampling
  - Inference is based on the assumed model
  - Treating the sample *s* as fixed, conditioning on the actual sample
- Best linear unbiased predictors
- Uncertainty assessment: mean squared error of prediction

Model-based approach

$y_1, y_2, \ldots, y_N$ are realized values of random variables $Y_1, Y_2, \ldots Y_N$

Two stochastic elements:

1) sample $s \sim p(\cdot)$ 2) $(Y_1, Y_2, \ldots Y_N) \sim f_\theta$

Treat the sample s as fixed

We can decompose the total t as follows:

$$t = \sum_{i=1}^{N} y_i = \sum_{i \in s} y_i + \sum_{i \notin s} y_i$$

Since $\sum_{i \in s} y_i$ is known, the problem is to estimate

$z = \sum_{i \notin s} y_i$, the realized value of $Z = \sum_{i \notin s} Y_i$

- The unobserved z is a realized value of the random variable Z, so the problem is actually to *predict* the value z of Z.

Can be done by predicting each unobserved y_i: $\hat{y}_i, i \notin s$

$$\text{Estimator}: \hat{t}_{pred} = \sum_{i \in s} y_i + \sum_{i \notin s} \hat{y}_i = \sum_{i \in s} y_i + \hat{z}$$

$\hat{z}$ is a predictor for z

- The prediction approach: Determine $\hat{y}_i$ by modeling

Remarks:

1. Any estimator can be expressed on the “prediction form:

$$\hat{t} = \sum_{i \in s} y_i + \hat{z}_{\hat{t}}$$

$$\text{letting } \hat{z}_{\hat{t}} = \hat{t} - \sum_{i \in s} y_i$$

2. Can then use this form to see if the estimator makes any sense

Ex 1. $\hat{t} = N\bar{y}_s = \sum_{i\in s} y_i + (N-n)\bar{y}_s = \sum_{i\in s} y_i + \sum_{i\notin s} \bar{y}_s$

Hence, $\hat{z} = \sum_{i\notin s} \bar{y}_s$ and $\hat{y}_i = \bar{y}_s$, for all $i \in s$

Ex.2 $\hat{t}_{HT} = \sum_{i\in s} y_i/\pi_i$ and $\pi_i = nx_i/t_x,\ t_x = \sum_{i=1}^{N} x_i$

Reasonable sampling design when *y* and *x* are positively correlated

$$\hat{t}_{HT} = \sum_{i\in s} \frac{t_x y_i}{nx_i} = \sum_{i\in s} y_i + \sum_{i\in s} y_i \left(\frac{t_x}{nx_i} - 1 \right)$$

$$= \sum_{i\in s} y_i + \underbrace{\frac{1}{n} \sum_{i\in s} \frac{y_i}{x_i} \left(\frac{(t_x - nx_i)}{t_x - n\bar{x}_s} \right)}_{\hat{\beta}_{HT}} \sum_{i\notin s} x_i = \sum_{i\in s} y_i + \hat{z}_{HT}$$

$$\hat{z}_{HT} = \sum_{i\notin s} \hat{\beta}_{HT} x_i = \sum_{i\notin s} \hat{y}_i$$

$\hat{\beta}_{HT}$ is a rather unusual regression coefficient

Three common models

I. A model for business surveys, the ratio model:

- assume the existence of a positive auxiliary variable *x* for all units in the population.

$$Y_i = \beta x_i + \varepsilon_i \quad \text{with } E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2 x_i \text{ and } Cov(\varepsilon_i, \varepsilon_j) = 0$$

$$\Leftrightarrow E(Y_i) = \beta x_i \, , Var(Y_i) = \sigma^2 x_i \text{ and } Cov(Y_i, Y_j) = 0$$

II. A model for social surveys, simple linear regression:

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \ , \ E(\varepsilon_i) = 0, \ Var(\varepsilon_i) = \sigma^2 \text{ and } Cov(\varepsilon_i, \varepsilon_j) = 0$$

- Ex: x_i is a measure of the "size" of unit i, and y_i tends to increase with increasing x_i – allow for negative x-values

III. Common mean model:

$$E(Y_i) = \beta \ , \ Var(Y_i) = \sigma^2 \text{ and the } Y_i's \text{ are uncorrelated}$$

Model-based estimators (predictors)

1. Predictor: $\hat{T} = \sum_{i \in s} Y_i + \hat{Z}$
2. Model parameters: θ
3. $\hat{T}$ is model-unbiased if $E_\theta(\hat{T} - T \mid s) = 0 \ \forall \theta, \ T = \sum_{i=1}^{N} Y_i$
4. Model variance of model-unbiased predictor is the variance of the *prediction error*, or *prediction variance* or *mean squared error of prediction*

$$Var_\theta(\hat{T} - T \mid s) = E_\theta((\hat{T} - T)^2 \mid s)$$

5. From now on, skip *s* in the notation: all expectations and variances are given the selected sample *s*, for example

$$E(\hat{T} - T) = E(\hat{T} - T \mid s)$$

$$Var(\hat{T} - T) = Var(\hat{T} - T \mid s)$$

Prediction variance as a variance measure for the actual observed sample

$N+1$ possible samples: $\{1\}$, $\{2\}, \ldots, \{N\}$, $\{1,2,\ldots N\}$

Use $\hat{T} = N\overline{Y}_s$ as the estimator for the population total T

Assume we select the "sample" $\{1,2,\ldots,N\}$.

Then $\hat{T} = N\overline{Y} = T$

Prediction variance: $Var(\hat{T} - T) = Var(0) = 0$

Linear predictor: $\hat{T} = \sum_{i \in s} a_i(s) Y_i$

6. Definition:

$\hat{T}_0$ is the best linear unbiased (BLU) predictor for T if

1) $\hat{T}_0$ is model-unbiased

2) $\hat{T}_0$ has uniformly minimum prediction variance among all model-unbiased linear predictors:

For any model-unbiased linear predictor $\hat{T}$

$$Var_\theta(\hat{T}_0 - T) \le Var_\theta(\hat{T} - T) \text{ for all } \theta$$

Ratio Model

$$Y_i = \beta x_i + \varepsilon_i \text{ , } E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2 v(x_i)$$

$Y_1, ..., Y_N$ are uncorrelated, $Cov(\varepsilon_i, \varepsilon_j) = 0$

Usually, $v(x) = x^g$, $0 \le g \le 2$

Suggested Predictor:

$$\hat{T}_{pred} = \sum_{i \in s} Y_i + \sum_{i \notin s} \hat{\beta}_{opt} x_i$$

where $\hat{\beta}_{opt}$ is the best linear unbiased estimator (BLUE) of β

$$\hat{\beta}_{opt} = \frac{\sum_{i \in s} x_i Y_i / v(x_i)}{\sum_{i \in s} x_i^2 / v(x_i)}$$

$$\hat{\beta} = \sum_{i \in s} c_i(s) Y_i$$

$$E(\hat{\beta}) = \beta \sum_{i \in s} c_i(s) x_i = \beta, \forall \beta \Leftrightarrow \sum_{i \in s} c_i(s) x_i = 1$$

$$Var(\hat{\beta}) = \sigma^2 \sum_{i \in s} c_i^2 v(x_i)$$

Minimize $\sum_{i \in s} c_i^2 v(x_i)$ subject to $\sum_{i \in s} c_i x_i = 1$

using Lagrange method

$$Q = \sum_{i \in s} c_i^2 v(x_i) + \lambda \left(\sum_{i \in s} c_i x_i - 1\right)$$

$$\partial Q / \partial c_i = 2 c_i v(x_i) + \lambda x_i = 0$$

$$\Leftrightarrow c_i = (-\lambda / 2) \frac{x_i}{v(x_i)}$$

Determine $(-\lambda/2)$ such that $\sum_{i\in s} c_i x_i = 1$:

$$-\lambda/2 \sum_{i\in s} x_i^2 / v(x_i) = 1$$

$$\Rightarrow (-\lambda/2) = 1 / \sum_{i\in s} x_i^2 / v(x_i)$$

and $c_{i,opt} = \dfrac{x_i / v(x_i)}{\sum_{j\in s} x_j^2 / v(x_j)}$

and $\hat{\beta}_{opt} = \sum_{i\in s} c_{i,opt} Y_i = \dfrac{\sum_{i\in s} x_i Y_i / v(x_i)}{\sum_{j\in s} x_j^2 / v(x_j)}$

This is the least squares estimate based on $Y_i / \sqrt{v(x_i)}$

$\hat{T}_{pred}$ is the best linear unbiased (BLU) predictor for T

Let $\hat{T}$ be a model-unbiased and linear predictor

Let $\hat{Z} = \hat{T} - \sum_{i \in s} Y_i$ and $\hat{\beta} = \hat{Z} / \sum_{i \notin s} x_i$.

$\Rightarrow \hat{T} = \sum_{i \in s} Y_i + \hat{\beta} \sum_{i \notin s} x_i$

$\hat{T}$ linear predictor $\Leftrightarrow \hat{\beta}$ is linear in $(Y_i, i \in s)$

and $\hat{T}$ model-unbiased $\Leftrightarrow E(\hat{\beta}) = \beta$

since $E(\hat{T}-T)=E(\hat{\beta}\sum_{i\notin s}x_i-\sum_{i\notin s}Y_i)$

$$=E[\hat{\beta}\sum_{i\notin s}x_i]-\sum_{i\notin s}\beta x_i=[E(\hat{\beta})-\beta]\sum_{i\notin s}x_i$$

such that $E(\hat{T}-T)=0\Leftrightarrow E(\hat{\beta})=\beta$

The prediction variance of model-unbiased predictor:

$$Var(\hat{T}-T)=Var(\hat{\beta}\sum_{i\notin s}x_i-\sum_{i\notin s}Y_i)$$

$$=Var(\hat{\beta}\sum_{i\notin s}x_i)+Var(\sum_{i\notin s}Y_i)$$

$$=(\sum_{i\notin s}x_i)^2Var(\hat{\beta})+\sigma^2\sum_{i\notin s}v(x_i)$$

To minimize the prediction variance is equivalent to minimizing $Var(\hat{\beta})$

Giving us $\hat{T}_{pred}$ as the BLU predictor

The prediction variance of the BLU predictor (**BLUP**):

$$Var(\hat{T}_{pred} - T) = (\sum_{i \notin s} x_i)^2 Var(\hat{\beta}_{opt}) + \sigma^2 \sum_{i \notin s} v(x_i)$$

$$= (\sum_{i \notin s} x_i)^2 \frac{\sigma^2}{\sum_{i \in s} x_i^2 / v(x_i)} + \sigma^2 \sum_{i \notin s} v(x_i)$$

$$= \sigma^2 \left(\frac{(\sum_{i \notin s} x_i)^2}{\sum_{i \in s} x_i^2 / v(x_i)} + \sum_{i \notin s} v(x_i) \right)$$

A variance estimate is obtained by using the model-unbiased estimator for s^2

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i \in s} \frac{1}{v(x_i)} (Y_i - \hat{\beta}_{opt} x_i)^2$$

$$\hat{V}(\hat{T}_{pred} - T) = \hat{\sigma}^2 \left(\frac{(\sum_{i \notin s} x_i)^2}{\sum_{i \in s} x_i^2 / v(x_i)} + \sum_{i \notin s} v(x_i) \right)$$

The central limit theorem applies such that for large *n*, *N*-*n* we have that

$(\hat{T}_{pred} - T) / \sqrt{\hat{V}(\hat{T}_{pred} - T)}$ is approximately $N(0,1)$

Approximate 95% *prediction interval* for *T*:

$$\hat{t}_{pred} \pm 1.96 \sqrt{\hat{V}(\hat{T}_{pred} - T)}$$

Three special cases: 1) $v(x) = x$, the ratio model, 2) $v(x) = x^2$ and 3) $x_i = 1$ for all i, the common mean model

1. $v(x) = x$

$$\hat{\beta}_{opt} = \frac{\sum_{i\in s} x_i Y_i / v(x_i)}{\sum_{i\in s} x_i^2 / v(x_i)} = \frac{\sum_{i\in s} Y_i}{\sum_{i\in s} x_i} = \hat{R}\text{, the usual sample ratio}$$

$$\hat{T}_{pred} = \sum_{i\in s} Y_i + \sum_{i\notin s} \hat{R} x_i$$

$$= \hat{R}\sum_{i\in s} x_i + \hat{R}\sum_{i\notin s} x_i = \hat{R}\cdot t_x$$

the usual ratio estimator

$$Var(\hat{T}_{pred} - T) = \sigma^2\left((\sum_{i\notin s} x_i)^2 / (\sum_{i\in s} x_i) + \sum_{i\notin s} x_i \right)$$

$$= N^2 \frac{1-f}{n} \cdot \frac{\bar{x}_r \bar{x}}{\bar{x}_s} \sigma^2,$$

$$f = n/N, \bar{x}_r = \sum_{i\notin s} x_i / (N-n) \quad \text{and } \bar{x} = \sum_{i=1}^{N} x_i / N$$

2. $\boldsymbol{v(x) = x^2}$

$$\hat{\beta}_{opt} = \frac{\sum_{i\in s} x_i Y_i / v(x_i)}{\sum_{i\in s} x_i^2 / v(x_i)} = \frac{\sum_{i\in s} Y_i / x_i}{n}, \text{ the sample mean of the ratios}$$

$$\hat{T}_{pred} = \sum_{i\in s} Y_i + \sum_{i\notin s} \hat{\beta}_{opt} x_i$$

$$= \sum_{i\in s} Y_i + \left(\frac{1}{n} \sum_{i\in s} \frac{Y_i}{x_i}\right) \sum_{i\notin s} x_i$$

$$Var(\hat{T}_{pred} - T) = \sigma^2 \left(\frac{(\sum_{i\notin s} x_i)^2}{\sum_{i\in s} x_i^2 / v(x_i)} + \sum_{i\notin s} v(x_i) \right)$$

$$= \sigma^2 \left(\frac{(\sum_{i\notin s} x_i)^2}{n} + \sum_{i\notin s} x_i^2 \right)$$

Resembles the H-T estimator when $\pi_i = nx_i / t_x$:

Let $R_i = Y_i / x_i$ and $\bar{R}_s = \sum_{i \in s} R_i / n$

$$\hat{T}_{HT} = \sum_{i \in s} \frac{t_x Y_i}{n x_i} = t_x \cdot \bar{R}_s$$

$$\hat{T}_{pred} = \sum_{i \in s} Y_i + \bar{R}_s \sum_{i \notin s} x_i = t_x \cdot \bar{R}_s + \sum_{i \in s} (Y_i - \bar{R}_s x_i)$$

When the sampling fraction *f* is small or when the x_i values vary little, these two estimators are approximately the same. In the latter case:

$$\bar{R}_s \approx \frac{1}{n \bar{x}_s} \sum_{i \in s} Y_i \text{ and } \sum_{i \in s} \bar{R}_s x_i \approx \sum_{i \in s} Y_i$$

3. $\mathbf{x_i = 1}$ *Model* :

$$Y_i = \beta + \varepsilon_i \ , E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2$$

$$Y_1, \ldots, Y_N \text{ are uncorrelated }, Cov(\varepsilon_i, \varepsilon_j) = 0$$

$$\hat{\beta}_{opt} = \frac{\sum_{i \in s} x_i Y_i / v(x_i)}{\sum_{i \in s} x_i^2 / v(x_i)} = \frac{1}{n} \sum_{i \in s} Y_i = \bar{Y}_s, \text{ the sample mean}$$

$$\hat{T}_{pred} = \sum_{i \in s} Y_i + \sum_{i \notin s} \bar{Y}_s = N \cdot \bar{Y}_s$$

$$Var(N \cdot \bar{Y}_s - T) = \sigma^2 \left(\frac{(\sum_{i \notin s} x_i)^2}{\sum_{i \in s} x_i^2 / v(x_i)} + \sum_{i \notin s} v(x_i) \right)$$

$$= \sigma^2 \left(\frac{(N-n)^2}{n} + (N-n) \right) = N^2 (1 - f) \frac{\sigma^2}{n}$$

This is also the usual, design-based variance formula under SRS

We see that the variance estimate is given by

$$N^2(1-f)\frac{\hat{\sigma}^2}{n}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y}_s)^2$$

the sample variance

Exactly the same as in the design-approach, but the interpretation is different!!

Simple Linear regression model

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i,\ E(\varepsilon_i) = 0,\ Var(\varepsilon_i) = \sigma^2$$

$Y_1, \ldots, Y_N$ are uncorrelated

BLUP:

$$\hat{T}_{pred} = \sum_{i \in s} Y_i + \sum_{i \notin s} (\hat{\beta}_1 + \hat{\beta}_2 x_i)$$

where

$\hat{\beta}_1$ and $\hat{\beta}_2$ are the LS estimators,

$$\hat{\beta}_2 = \frac{\sum_{i \in s} (x_i - \bar{x}_s)(Y_i - \bar{Y}_s)}{\sum_{i \in s} (x_i - \bar{x}_s)^2} = \frac{\sum_{i \in s} (x_i - \bar{x}_s) Y_i}{\sum_{i \in s} (x_i - \bar{x}_s)^2}$$

$$\hat{\beta}_1 = \bar{Y}_s - \hat{\beta}_2 \bar{x}_s$$

$$\hat{T}_{pred} = \sum_{i\in s} Y_i + \sum_{i\notin s} (\hat{\beta}_1 + \hat{\beta}_2 x_i)$$

$$= n\overline{Y}_s + (N-n)\overline{Y}_s + \hat{\beta}_2 (\sum_{i\notin s} x_i - (N-n)\bar{x}_s)$$

$$= N\overline{Y}_s + \hat{\beta}_2 (t_x - N\bar{x}_s)$$

$$\Rightarrow \hat{T}_{pred} = N[\overline{Y}_s + \hat{\beta}_2 (\bar{x} - \bar{x}_s)]$$

Clearly, $\hat{T}_{pred}$ is model - unbiased :

$$E(T) = \sum_{i=1}^{N} (\beta_1 + \beta_2 x_i) = N(\beta_1 + \beta_2 \bar{x})$$

and

$$E(\hat{T}_{pred}) = N\{\frac{1}{n} \sum_{i\in s} (\beta_1 + \beta_2 x_i) + \beta_2 (\bar{x} - \bar{x}_s)\} = N(\beta_1 + \beta_2 \bar{x})$$

We shall now show that this is BLUP.

Assume first that $\bar{x} \neq \bar{x}_s$. Let $\hat{T}$ be a linear,

model-unbiased predictor, and let $b = (\hat{T}/N - \bar{Y}_s)/(\bar{x} - \bar{x}_s)$.

$$\frac{1}{N}\hat{T} - \bar{Y}_s = b(\bar{x} - \bar{x}_s) \Rightarrow \hat{T} = N[\bar{Y}_s + b(\bar{x} - \bar{x}_s)]$$

Hence, **any** predictor can be expressed on this form and the predictor is linear if and only if *b* is linear in the Y_i's

Also, $\hat{T}$ is model-unbiased $\Leftrightarrow E(b) = \beta_2$:

$$E(\hat{T}) = E(T) = N(\beta_1 + \beta_2\bar{x})$$

$$\Leftrightarrow N[\beta_1 + \beta_2\bar{x}_s + (\bar{x} - \bar{x}_s)E(b)] = N(\beta_1 + \beta_2\bar{x})$$

$$\Leftrightarrow (\bar{x} - \bar{x}_s)E(b) = \beta_2\bar{x} - \beta_2\bar{x}_s = \beta_2(\bar{x} - \bar{x}_s).$$

Prediction variance:

$$Var(\hat{T}-T) = Var\Big((N-n)\overline{Y}_s + Nb(\bar{x}-\bar{x}_s)\Big) + (N-n)\sigma^2$$

$$b = \sum_{i\in s} c_i(s)Y_i, \text{ unbiased estimator of } \beta_2 :$$

$$E(b) = \beta_2 \Leftrightarrow \sum_{i\in s} c_i(\beta_1 + \beta_2 x_i) = \beta_2$$

$$\Leftrightarrow \beta_1 \sum_{i\in s} c_i + \beta_2 \sum_{i\in s} c_i x_i = \beta_2$$

$$E(b) = \beta_2 \Leftrightarrow (1) \sum_{i\in s} c_i = 0 \text{ and } (2) \sum_{i\in s} c_i x_i = 1$$

So we need to minimize the prediction variance with respect to the c_i's under (1) and (2) i.e. minimize

$$Var\left((N-n)\overline{Y}_s + Nb(\bar{x}-\bar{x}_s)\right) = Var\sum_{i\in s} Y_i\left(\frac{N-n}{n} + N(\bar{x}-\bar{x}_s)c_i\right)$$

$$=\sigma^2\sum_{i\in s}\left(\frac{N-n}{n} + N(\bar{x}-\bar{x}_s)c_i\right)^2$$

$$=\sigma^2[N^2(\bar{x}-\bar{x}_s)^2\sum_{i\in s}c_i^2 + 2\frac{N-n}{n}N(\bar{x}-\bar{x}_s)\sum_{i\in s}c_i + \frac{(N-n)^2}{n}]$$

Since $\sum_{i\in s} c_i = 0$,

it is enough to minimize $\sum_{i\in s} c_i^2$ under conditions (1) and (2)

$$Q = \sum_{i \in s} c_i^2 - 2\lambda_1 (\sum_{i \in s} c_i) - 2\lambda_2 (\sum_{i \in s} c_i x_i - 1)$$

$$\partial Q / \partial c_i = 2c_i - 2\lambda_1 - 2\lambda_2 x_i = 0 \Leftrightarrow c_i = \lambda_1 + \lambda_2 x_i$$

$$(1) \ \sum_{i \in s} c_i = 0 \Rightarrow \lambda_1 + \lambda_2 \bar{x}_s = 0$$

$$(2) \ \sum_{i \in s} c_i x_i = 1 \Rightarrow \lambda_1 n \bar{x}_s + \lambda_2 \sum_{i \in s} x_i^2 = 1$$

$$(1) \ \Rightarrow \lambda_1 = -\lambda_2 \bar{x}_s$$

$$\text{from } (2): \ \lambda_2 \sum_{i \in s} x_i^2 - \lambda_2 n \bar{x}_s^2 = 1$$

$$\lambda_2 = 1 / \sum_{i \in s} (x_i - \bar{x}_s)^2$$

$$c_i = \lambda_1 + \lambda_2 x_i = \lambda_2 (x_i - \bar{x}_s) = \frac{x_i - \bar{x}_s}{\sum_{j \in s} (x_j - \bar{x}_s)^2}$$

$$\text{and } b = \sum_{i \in s} Y_i \frac{x_i - \bar{x}_s}{\sum_{j \in s} (x_j - \bar{x}_s)^2} = \frac{\sum_{i \in s} (x_i - \bar{x}_s) Y_i}{\sum_{j \in s} (x_j - \bar{x}_s)^2} = \hat{\beta}_2$$

The prediction variance is given by

$$Var(\hat{T}_{pred} - T) = \frac{N^2}{n} \sigma^2 \left[(1 - \frac{n}{N}) + \frac{n(\bar{x} - \bar{x}_s)^2}{\sum_{i \in s} (x_i - \bar{x}_s)^2} \right]$$

and variance estimate is obtained by estimating σ^2 with

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i \in s} (Y_i - \bar{Y}_s - \hat{\beta}_2 (x_i - \bar{x}_s))^2$$

So far, $\bar{x} \neq \bar{x}_s$. What if $\bar{x} = \bar{x}_s$?

Then $\hat{T}_{pred} = N\bar{Y}_s$ and is the BLU predictor.

For any linear predictor, $\hat{T} = \sum_{i \in s} a_i Y_i$

$$Var(\hat{T} - T) = Var[\sum_{i \in s} (a_i - 1) Y_i] + (N - n)\sigma^2$$

$$= \sigma^2 [\sum_{i \in s} (a_i - 1)^2 + (N - n)]$$

Let $\hat{T}_a = a \sum_{i \in s} Y_i,\ a = \sum_{i \in s} a_i / n = \bar{a}$

$$Var(\hat{T}_a - T) = \sigma^2 [\sum_{i \in s} (\bar{a} - 1)^2 + (N - n)]$$

$$= \sigma^2 [n(\bar{a} - 1)^2 + (N - n)]$$

$$\sum_{i\in s}(a_i-1)^2 \ge n(\bar{a}-1)^2$$

$$\Rightarrow Var(\hat{T}-T) \ge Var(\hat{T}_a - T)$$

and $\hat{T}$ model-unbiased $\Rightarrow \sum_{i\in s} a_i = N$ and $\bar{a} = N/n$:

$$\hat{T}_a = N\bar{Y}_s = \hat{T}_{pred}.$$

Anticipated variance

A variance measure that tells us about the expected uncertainty over *repeated* surveys

1. Conditional on the sample s, with model-unbiased $\hat{T}$:

$Var(\hat{T} - T)$ measures the uncertainty for *this* particular sample s

2. The expected uncertainty for repeated surveys:

$E_p\{Var(\hat{T} - T)\}$, over the sampling distribution $p(\cdot)$

3. This is called the ***anticipated variance***.

4. It evaluates the uncertainty of the ***strategy*** consisting of a sampling design and an associated estimator

If $\hat{T}$ is not model-unbiased, we use

$$E_p\{E(\hat{T}-T)^2\}$$

as a criterion for uncertainty, the anticipated mean square error

Note: If $\hat{T}$ is design-unbiased then

$$E_p\{E(\hat{T}-T)^2\} = E\{E_p(\hat{T}-T)^2 \mid \mathbf{Y})\}$$

and

$$E_p(\hat{T}-T)^2 \mid \mathbf{Y} = \mathbf{y}) = E_p(\hat{t}-t)^2 = Var_p(\hat{t})$$

And the anticipated MSE becomes the expected design-variance, also called the anticipated design variance

$$E_p\{E(\hat{T}-T)^2\} = E\{Var_p(\hat{T})\}$$

Example: Simple linear regression and simple random sample

If sample mean $N \cdot \bar{Y}_s$ is used : It is not model - unbiased, but is design - unbiased :

$$E_p\{E(N \cdot \bar{Y}_s - T)^2\} = E\{Var_p(N \cdot \bar{Y}_s)\} = N^2 \frac{1-f}{n} E\{\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2\}$$

$$= N^2 \frac{1-f}{n}\{\sigma^2 + \frac{1}{N-1}\sum_{i=1}^{N}(\mu_i - \bar{\mu})^2\}$$

$$\mu_i = E(Y_i) = \beta_1 + \beta_2 x_i,\ \bar{\mu} = \beta_1 + \beta_2 \bar{x}$$

$$E\{Var_p(N \cdot \bar{Y}_s)\} = N^2 \frac{1-f}{n}\{\sigma^2 + \beta_2^2 S_x^2\}$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Let us now study the BLUP. (It can be shown that it is approximately design-unbiased)

$$Var(\hat{T}_{pred} - T) = \frac{N^2}{n}\sigma^2\left[(1-\frac{n}{N}) + \frac{n(\bar{x}-\bar{x}_s)^2}{\sum_{i\in s}(x_i-\bar{x}_s)^2}\right]$$

$$\Rightarrow E_p\{Var(\hat{T}_{pred} - T)\} = \frac{N^2}{n}\sigma^2\left[(1-\frac{n}{N}) + E_p\,\frac{n(\bar{x}-\bar{x}_s)^2}{\sum_{i\in s}(x_i-\bar{x}_s)^2}\right]$$

$$\approx \frac{N^2}{n}\sigma^2\left[(1-\frac{n}{N}) + \frac{E_p\{n(\bar{x}-\bar{x}_s)^2\}}{E_p\sum_{i\in s}(x_i-\bar{x}_s)^2}\right]$$

$$E_p n(\bar{x}_s - \bar{x})^2 = nVar_p(\bar{x}_s) = (1-f)S_x^2$$

$$E_p\sum_{i\in s}(x_i-\bar{x}_s)^2 = (n-1)S_x^2$$

$$E_p\{Var(\hat{T}_{pred} - T)\} \approx \frac{N^2}{n}\sigma^2\left[(1-f) + \frac{1-f}{n-1}\right]$$

$$= \frac{N^2}{n-1}(1-f)\sigma^2 \approx \frac{N^2}{n}(1-f)\sigma^2$$

compared to

$$E\{Var_p(N\cdot\bar{Y}_s)\} = N^2\,\frac{1-f}{n}\{\sigma^2 + \beta_2^2 S_x^2\}$$

$\hat{T}_{pred}$ eliminates the term $\beta_2^2 S_x^2$

and is much more efficient than $N\cdot\bar{Y}_s$

Remarks

- From a design-based perspective, the sample mean is unbiased, the linear regression estimator is not
- Considering only the *design-bias*, we might choose the sample mean based estimator
- The linear regression estimator could be selected over the sample mean estimator because the strategy of (SRS, BLUP) has smaller anticipated variance, i.e. than (SRS, HT), over all possible samples and populations (under the model)

Robust variance estimation

- The assumed model is really a prediction “tool”
- Often, the variance assumption is somewhat misspecified
  - like constant variance
  - variance proportional to size measure x_i
- Standard least squares variance estimates is sensitive to misspecification of variance assumption
- Concerned with robust variance estimators

Variance estimation for the ratio estimator

Working model:

$$Y_i = \beta x_i + \varepsilon_i \text{ , } E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2 x_i$$

$$Y_1, \ldots, Y_N \text{ are uncorrelated , } Cov(\varepsilon_i, \varepsilon_j) = 0$$

Under this working model, the unbiased estimator of the prediction variance of the ratio estimator is

$$\hat{V}_R(\hat{R} \cdot t_x - T) = N^2 \frac{1-f}{n} \cdot \frac{\bar{x}_r \bar{x}}{\bar{x}_s} \hat{\sigma}^2$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i \in s} \frac{1}{x_i} (Y_i - \hat{R} \cdot x_i)^2$$

$$\hat{R} = \bar{Y}_s / \bar{x}_s$$

This variance estimator is non-robust to misspecification of the variance model.

Suppose the true model has

$$E(Y_i) = \beta x_i \text{ and } Var(Y_i) = \sigma^2 v(x_i)$$

Ratio estimator is still model-unbiased but prediction variance is now

$$Var(\hat{R}\cdot t_x - T) = (\sum_{i\notin s} x_i)^2 Var(\hat{R}) + \sigma^2 \sum_{i\notin s} v(x_i)$$

$$= (\sum_{i\notin s} x_i)^2 \frac{\sigma^2 \sum_{i\in s} v(x_i)}{(\sum_{i\in s} x_i)^2} + \sigma^2 \sum_{i\notin s} v(x_i)$$

$$= \sigma^2 \left(\frac{(N-n)^2 \bar{x}_r^2}{n^2 \bar{x}_s^2} \sum_{i\in s} v(x_i) + \sum_{i\notin s} v(x_i) \right)$$

$$Var(\hat{R}\cdot t_x - T) = \sigma^2\left(\frac{(N-n)^2\,\bar{x}_r^2}{n\bar{x}_s^2}\bar{v}_s + (N-n)\bar{v}_r\right)$$

$$= \sigma^2 N^2\,\frac{1-f}{n}\Big((1-f)\bar{v}_s(\bar{x}_r/\bar{x}_s)^2 + f\cdot\bar{v}_r\Big)$$

$$\bar{v}_s = \sum\nolimits_{i\in s} v(x_i)/n \quad \text{and} \quad \bar{v}_r = \sum\nolimits_{i\notin s} v(x_i)/(N-n)$$

Moreover, $E(\hat{\sigma}^2) \neq \sigma^2$:

$$E(\hat{\sigma}^2) = \frac{1}{n-1}\sum\nolimits_{i\in s}\frac{1}{x_i}E(Y_i - \hat{R}\cdot x_i)^2$$

$$= \sigma^2\left[(v/x)_s + \frac{1}{n-1}\{(v/x)_s - \bar{v}_s/\bar{x}_s\}\right],\ (v/x)_s = \frac{1}{n}\sum\nolimits_{i\in s} v(x_i)/x_i$$

Robust variance estimator for the ratio estimator

$$Var(\hat{R}\cdot t_x - T) = \sigma^2 N^2 \frac{1-f}{n}\left((1-f)\bar{v}_s(\bar{x}_r/\bar{x}_s)^2 + f\cdot\bar{v}_r\right)$$

$$= \sigma^2 N^2 \frac{1-f}{n}\left(\bar{v}_s(\bar{x}_r/\bar{x}_s)^2 + f\cdot\{\bar{v}_r - \bar{v}_s(\bar{x}_r/\bar{x}_s)^2\}\right)$$

$$\approx \sigma^2\bar{v}_s\cdot N^2\frac{1-f}{n}(\bar{x}_r/\bar{x}_s)^2,$$

the leading term in the prediction variance

and: $\sigma^2\bar{v}_s = \frac{1}{n}\sum_{i\in s}\sigma^2 v(x_i) = \frac{1}{n}\sum_{i\in s}Var(Y_i)$

$$\sigma^2\bar{v}_s = \frac{1}{n}\sum_{i\in s}E(Y_i-\beta x_i)^2 = E\{\frac{1}{n}\sum_{i\in s}(Y_i-\beta x_i)^2\}$$

Suggests we may use:

$$\hat{\sigma}^2_{rob}\bar{v}_s = \frac{1}{n-1}\sum_{i\in s}(Y_i - \hat{R}x_i)^2$$

Leading to the robust variance estimator:

$$\hat{V}_{rob}(\hat{R}\cdot t_x - T) = (\bar{x}_r / \bar{x}_s)^2 \cdot N^2 \frac{1-f}{n} \cdot \frac{1}{n-1}\sum_{i\in s}(Y_i - \hat{R}x_i)^2$$

Compared to the *design* variance estimator in SRS:

$$\hat{V}_{SRS}(\hat{R}\cdot t_x) = (\bar{x} / \bar{x}_s)^2 \cdot N^2 \frac{1-f}{n} \cdot \frac{1}{n-1}\sum_{i\in s}(Y_i - \hat{R}x_i)^2$$

General approach to robust variance estimation

1. Find robust estimators of $Var(Y_i)$, that does not depend on model assumptions about the variance

2. $\hat{T} = \sum_{i \in s} w_{is} Y_i$

   $$Var(\hat{T} - T) = \sum_{i \in s} (w_{is} - 1)^2 Var(Y_i) + \sum_{i \notin s} Var(Y_i)$$

3. For $i \in s : \hat{V}(Y_i) = (Y_i - \hat{\mu}_i)^2$

   $\hat{\mu}_i$ estimate $E(Y_i)$

4. Estimate only leading term in the prediction variance, typically dominating, or estimate the second term from the more general model

- Reference to robust variance estimation:
- Valliant, Dorfman and Royall (2000):

  Finite Population Sampling and Inference. A Prediction Approach, ch. 5

However, even more importantly, what if the (linear, model) predictor is misspecified?

Model-assisted approach

- Design-based perspective to inference
- Assume the existence of relevant auxiliary variables, known for all units in the population
- Use modeling to motivate efficiency improvement of the basic HT-estimator.
- Design consistency remains an estimation criterion
- Basic idea:

Suppose $\hat{y}_i = \hat{\beta} x_i$ is a regression-based "estimate" for each y_i in the population. Here, x_i is known for the whole population

$$t = \sum_{i=1}^{N} \hat{y}_i + \sum_{i=1}^{N} (y_i - \hat{y}_i) \text{ and } e = \sum_{i=1}^{N} e_i, \quad \text{where } e_i = (y_i - \hat{y}_i).$$

$$\hat{e}_{HT} = \sum_{i \in s} \frac{e_i}{\pi_i}$$

Final estimator, the generalised regression estimator:

$$\hat{t}_{reg} = \sum_{i=1}^{N} \hat{\beta} x_i + \hat{e}_{HT}$$

Alternative expression:

$$\hat{t}_{reg} = \sum_{i \in s} \frac{y_i}{\pi_i} + \hat{\beta}(t_x - \sum_{i \in s} \frac{x_i}{\pi_i}), \qquad t_x = \sum_{i=1}^{N} x_i$$

$$\hat{t}_{reg} = \hat{t}_{y,HT} + \hat{\beta}(t_x - \hat{t}_{x,HT})$$

Robustness against predictor misspecification

Assume $x_i\beta$ to be conditional expectation of Y_i

It is misspecified if $E(Y_i - x_i\beta) \neq 0$

Let $e_i = y_i - x_i\beta$ and e the corresponding population total

Let $\hat{\beta}$ be an estimator of β, and $\hat{e}_i = y_i - x_i\hat{\beta}$

We have $\hat{e}_i \to \mathrm{e_i}$, as $\hat{\beta} \to E_p(\hat{\beta}) = \beta$, *in probability*

Can write $t = X\beta + e$ and X the population total of x_i

such that $\hat{t}_{reg} = X\hat{\beta} + \hat{e}_{HT} \to t$, *in probability*

even if $E(Y_i - x_i\beta) \neq 0$

Simple random sample

$$\hat{t}_{reg} = N\bar{y}_s + \hat{\beta}(t_x - N\bar{x}_s)$$

Model : The Y_i's are independen t and

$$Y_i = \beta x_i + \varepsilon_i, \ E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2 x_i$$

$\Rightarrow$ Best linear unbiased estimator : $\hat{\beta} = \bar{y}_s / \bar{x}_s$

$$\Rightarrow \hat{t}_{reg} = N\bar{y}_s + \frac{\bar{y}_s}{\bar{x}_s} t_x - N\bar{y}_s = t_x \frac{\bar{y}_s}{\bar{x}_s}, \text{ the ratio estimator}$$

In general with this "ratio model", in order to get approximately design-unbiased estimators:

Can regard β - estimate as an estimate of $\sum_{i=1}^{N} y_i / \sum_{i=1}^{N} x_i$

Numerator is estimated by $\hat{t}_{y,HT} = \sum_{i \in s} y_i / \pi_i$

Denominator is estimated by $\hat{t}_{x,HT} = \sum_{i \in s} x_i / \pi_i$

$$\Rightarrow \text{use } \hat{\beta}_\pi = \hat{t}_{y,HT} / \hat{t}_{x,HT} = \frac{\sum_{i \in s} y_i / \pi_i}{\sum_{i \in s} x_i / \pi_i}$$

$$\Rightarrow \hat{t}_{reg} = t_x \hat{\beta}_\pi = \sum_{i=1}^{N} \hat{y}_i$$

where $\hat{y}_i = \hat{\beta}_\pi x_i$

Variance and variance estimation

Reference: Särndal, Swensson and Wretman : Model Assisted Survey Sampling (1992, ch. 6), Springer

- Regression estimator is approximately unbiased
- Variance estimation:

The sample residuals: $e_i = y_i - \hat{y}_i \ , i \in s$

where $\hat{y}_i = x_i \hat{\beta}_\pi$

If $|s| = n$, fixed in advance:

$$\hat{V}(\hat{t}_{reg}) = \sum_{i \in s} \sum_{\substack{j \in s, \\ j > i}} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2$$

Approximate 95% CI, for large *n*, *N*-*n*:

$$\hat{t}_{reg} \pm 1.96\sqrt{\hat{V}(\hat{t}_{reg})}$$

- Remark: In SSW (1992,ch.6), an alternative variance estimator is mentioned that may be preferable in many cases

Common mean model

$E(Y_i) = \beta$, $Var(Y_i) = \sigma^2$ and the $Y_i's$ are uncorrelated

The ratio model with x_i =1.

$$\hat{\beta}_\pi = \hat{t}_{y,HT} / \hat{t}_{x,HT} = \frac{\sum_{i \in s} y_i / \pi_i}{\sum_{i \in s} 1/\pi_i} = \tilde{y}_s = \frac{\hat{t}_{y,HT}}{\hat{N}}$$

$$\hat{t}_{reg} = t_x \hat{\beta}_\pi = N\hat{\beta}_\pi = N\tilde{y}_s$$

This is the modified H-T estimator

Typically much more efficient than the H-T estimator

$$e_i = y_i - \tilde{y}_s$$

$$\hat{V}(N\tilde{y}_s) = \sum_{i \in s} \sum_{\substack{j \in s, \\ j > i}} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2$$

Alternatively,

$$\hat{V}^*(N\tilde{y}_s) = (N / \hat{N})^2 \sum_{i \in s} \sum_{\substack{j \in s, \\ j > i}} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j} \right)^2$$

Remarks:

1. The model-assisted regression estimator has often the form

$$\hat{t}_{reg} = \sum_{i=1}^{N} \hat{y}_i$$

2. The prediction approach makes it clear: no need to estimate the observed y_i

3. **Any** estimator can be expressed on the “prediction form:

$$\hat{t} = \sum_{i \in s} y_i + \hat{z}_{\hat{t}}$$

$$\text{letting } \hat{z}_{\hat{t}} = \hat{t} - \sum_{i \in s} y_i$$

4. Can then use this form to see if the estimator makes any sense