Introduction to Sampling Design

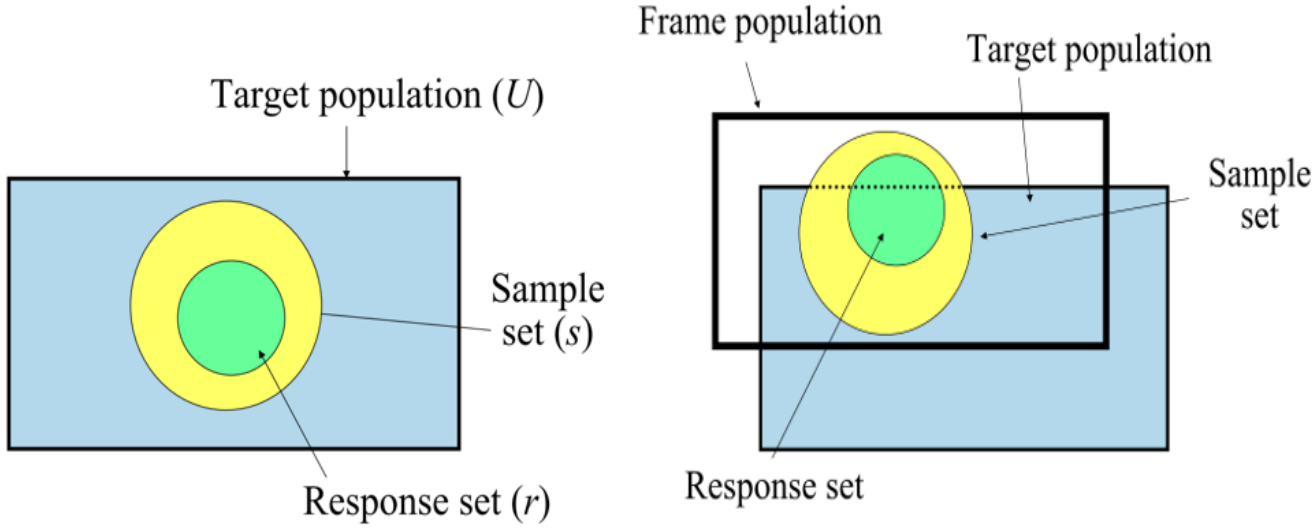
- How to construct a sampling frame
- Sample size determination and allocation

Cenker Burak METİN

Head of Survey and Questionnaire Design Group

How to construct a sampling frame

- ✓ For probability sampling, a list of units is required from which sample is selected from.
- ✓ In an ideal situation, the sampling frame must be identical to target population
- ✓ The coverage, completeness, timeliness(updated), information content and accuracy of the frame are critical factors
- ✓ Frame defects are ***Undercoverage, Overcoverage, Duplication, Misclarification***



Sample size determination and allocation

How to determine the appropriate sample size?

To response this question first ask these questions:

1. Which are the most important study variables, and the parameters to be estimated?
2. Is there any guess about the (statistical) distribution of the study variables?
3. What is the level of precision one would like to have for the parameter estimates?
4. What are the most important domains where the estimates must be provided and how precisely?
5. Are there any specific questions which must be taken in to account?
6. What will be the anticipated nonresponse rate?
7. What are the financial and time constraints?

Sample size determination and allocation

- ✓ The precision of the survey estimates and the sample size are interrelated.

  The sample size ↑ the sampling variance ↓
- ✓ The precision of an estimate $\hat{t}$, may be expressed in terms of the allowable standard error, SE($\hat{t}$), the margin of error, z $\times$ SE($\hat{t}$) , or the coefficient of variation SE($\hat{t}$) / $\hat{t}$.
- ✓ Sample size determination includes the specification of desired precision
- ✓ Sample size determination attempts to control for sampling errors and for nonresponse
- ✓ NSOs should consider some questions before deciding on the appropriate level of precision for survey estimates.

Sample size determination and allocation

1. How will the survey estimates be used? How much sampling variance is acceptable in the survey estimates? How much uncertainty can be tolerated?

   Is margin of error of ±6% with 95% confidence suitable or not

2. Are estimates required for subgroups (domains) of the survey population?

   In addition to producing survey estimates at the national level, provincial estimates may be required

3. How big is the sampling variance relative to the survey estimate?

Sample size determination and allocation

✓ For a given level of precision, the sampling variance (or standard error) formula for a SRS is generally used to calculate the sample size.

$$\widehat{SE}(\hat{\bar{Y}}) = \sqrt{(1-\frac{n}{N})}(\frac{\hat{s}}{\sqrt{n}})$$

Setting the required margin of error to e, then

$$e = z\sqrt{(1-\frac{n}{N})}(\frac{\hat{s}}{\sqrt{n}}) \qquad n = \frac{z^2\hat{s}^2}{e^2+\frac{z^2\hat{s}^2}{N}}$$

where z depends on the confidence level.

✓ For more complex sample designs need to use design effect (deff).

✓ **The design effect** is the ratio of the sampling variance of an estimator under a given design to the sampling variance of an estimator under SRS of the same sample size.

✓ Therefore, for a simple random sample design, deff = 1, and usually deff ≤ 1 for a stratified sample design and deff ≥ 1 for a cluster sample design.

Sample size determination and allocation

An important consideration in determining the efficiency of stratified sampling is the way in which the total sample size, n, is allocated to each stratum.

The allocation or distribution of the sample, n, to the L strata can be carried out using one of two criteria:

- ✓ The total sample size can be determined then distributed across the strata (called **fixed sample size**),
- ✓ The sample size required in each stratum to meet a precision requirement can be determined and summed to determine the total sample size (called **fixed coefficient of variation**, if the precision requirement is expressed in terms of the coefficient of variation).

Sample size determination and allocation

Fixed Sample Size

- ✓ A fixed sample size n is allocated to the strata in a specified manner.
- ✓ The proportion of the sample allocated to the h^{th} stratum is denoted as $a_h = {}^{n_h}/_{n}$, where each a_h is between 0 and 1 inclusively (i.e., $0 \leq a_h \leq 1$) and the sum of the a_h's is equal to 1 (i.e. $\sum_{h=1}^{L} a_h = 1$).
- ✓ For each stratum h, the sample size n_h is equal to the product of the total sample size n and the proportion a_h of the sample coming from that particular stratum:

$$n_h = n \times a_h$$

- ✓ For example, if a stratum has a proportion a_h =½, then half of the entire sample is allocated to that stratum.

Sample size determination and allocation

Fixed Coefficient of Variation

The alternative to fixing the sample size n, is to determine the sample size required in each stratum n_h, given a certain level of precision for the overall estimate. This implies finding the sample size n_h (h = 1,2, … , L) for each stratum, so that the coefficient of variation of the overall estimate does not exceed the desired value CV.

For example, consider the estimate of a total, $\hat{Y}$, from a stratified simple random sample. The equation for the coefficient of variation of an estimated total from a stratified sample can be manipulated into the following expression for the total sample size, n.

$$n = \frac{\sum_{h=1}^{L} N_h^2 S_h^2 / a_h}{CV^2 Y^2 + \sum_{h=1}^{L} N_h S_h^2}$$

- ✓ a_h is the proportion of the sample allocated to the stratum;
- ✓ CV is the required coefficient of variation of Y;

Sample size determination and allocation

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h}$$

$n_h = \tilde{n}\,[\,K^2 + (\,1 - K^2\,)\;M_h^{2}\;]^{1/2}$

n_h = sample size at h_{th} strata

$\tilde{n}$ = average sample size per strata

K^2 = measure of the relative importance

$M_h = H.N_h/N = H.W_h$

M_h= Effect of weight of strata to total strata number

W_h= Weight of strata (N_h / N)

H = Number of strata

$n_{min} = K.\tilde{n}$ sample size for the smallest strata

Proportional Allocation

Tabaka	Nh	nh
A	1500	150
B	400	40
C	800	80
D	300	30
N=	**3000**	300
n=	**300**	

Optimum Allocation

Tabaka	Nh	Sh	Nh*Sh	nh
A	1500	360000	540000000	280
B	400	5500	2200000	1
C	800	40000	32000000	17
D	300	14000	4200000	2
N=	**3000**		578400000	**300**
n=	**300**			

Compromise Allocation

STRATA	Nh	Nh/N	Mh	nh
A	1500	0.5	2	109
B	400	0.133333	0.533333	60
C	800	0.266667	1.066667	74
D	300	0.1	0.4	57
N=	**3000**	**1**	**4**	**301**
ñ=	**72**			
K=	**0.75**			
n_{min}=	**54**			

References

Franklin, S., & Walker, C. (2010). Survey methods and practices. Statistics Canada. *Social Survey Methods Division, Ottawa*. (Originally published in October 2003)

Lohr, S. (2009). Sampling design and analysis (No. 519.52 L64).

Banning, R., Camstra, A., & Knottnerus, P. (2012). Sampling theory. Sampling design and estimation methods. Statistics Netherlands. The Hague, 87.

Lehtonen, R., & Djerf, K. (2008). Survey sampling reference guidelines: introduction to sample design and estimation techniques.

Estimation theory for sample surveys

- **Design weights and HT Estimator**
- **Nonresponse adjustment**

Cenker Burak METİN

Head of Survey and Questionnaire Design Group

Design weights and HT Estimator

- The probability of selecting each unit is not always equal
- **Weighting** is one of the best way to obtain effective results
- The negative impacts of the non-sampling errors on estimates can also be eliminated
- The **design weight**s are constructed, subject to sampling design, to reflect the differences of inclusion probabilities on the estimation.
- Define **design weight** (or sampling weight) for any sampling design, to be the reciprocal of the inclusion probability

$$d_k = \frac{1}{\pi_k}$$

- The **design weight** of unit k can be interpreted as the number of population units represented by unit k.

Design weights and HT Estimator

- Due to coverage and non-response problems, the efficiency of the estimates decrease.
- These estimates may not be consistent to reliable external sources.
- Therefore, some adjustments to the design weights is required.
- Five steps: Calculate **design weights**, **adjust** these weights to compensate for **non-response**, **calibrate the weights to known totals** obtained from the external data sources, **trimming and scaling** of the weights.

Design weights and HT Estimator

- **Why** is **the weighting** important?
- Is there any problem for the estimates if we do not use weights when the design is not self weighted?
- Let the estimates of the population mean for the same sample be $\bar{y}$ (unweighted) and $\bar{y}_d$ (weighted)

$$\bar{y} = \frac{\sum_{j=1}^{n} y_j}{n}$$

$$\bar{y}_d = \frac{\sum_{j=1}^{n} d_j y_j}{\sum_{j=1}^{n} d_j}$$

- For unequal probability sampling, generally $V(\bar{y}) < V(\bar{y}_d)$
- But, according to the correlation between the design weights and the characteristics of interest $\bar{y}$ **will be biased**

Design weights and HT Estimator

- For most of the sampling strategy $V(\bar{y})$ and $V(\bar{y}_d)$ converges to 0 when n increase.
- On the other hand, **Bias($\bar{y}$) does not converge to 0** while **Bias($\bar{y}_d$) converges.**
- **Verma (2014)** claimed that $\bar{y}_d$ also can be used for decreasing the sampling variance.
- Is it conflict? Let's look 2 different sampling design

Strata 1 ($d_i = 1$)			Strata 2 ($d_i = 3$)				
Samples	$\sum_{j:1}^{n} y_i$	$\sum_{j:1}^{n} d_i y_i$	Samples	$\sum_{j:1}^{n} y_i$	$\sum_{j:1}^{n} d_i y_i$		
1;2;3	6	6	5	5	15		
1;2;4	7	7	6	6	18		
1;3;4	8	8	7	7	21		
2;3;4	9	9	8	8	24		
16 possible means for $\bar{y}$							
2,75	3,00	3,00	3,25	3,25	3,25	3,50	3,50
3,50	3,50	3,75	3,75	3,75	4,00	4,00	4,25
16 possible means for $\bar{y}_d$							
3,50	3,67	3,83	4,00	4,00	4,17	4,33	4,50
4,50	4,67	4,83	5,00	5,00	5,17	5,33	5,50
$\bar{Y}$ = 4,50							
$E(\bar{y})$ = 3,50		$V(\bar{y})$ = 0,16			$MSE(\bar{y})$ = 1,16		
$E(\bar{y}_d)$ = 4,50		$V(\bar{y}_d)$ = 0,35			$MSE(\bar{y}_d)$ = 0,35		

Design weights and HT Estimator

- For the second sampling design X auxiliary information used as MOS.

$$P(select\ k\ on\ first\ draw) = \varphi_k \quad (1)$$

$$Pr(l\ chosen\ second\ /\ \ k\ chosen\ first\ ) = \frac{\varphi_l}{1-\varphi_k} \quad (2)$$

$$\pi_{kl} = \varphi_k \frac{\varphi_l}{1-\varphi_k} + \varphi_l \frac{\varphi_k}{1-\varphi_l}$$

x_k	φ_k	y_k	d_k	Samples	$\sum_{k:1}^{n} y_k$	$\sum_{k:1}^{n} d_k y_k$	π_{kl}
1	1/16	5	5,26	5;9	14	152	0,0173
2	2/16	9	2,70	5;12	17	144	0,0269
3	3/16	12	1,85	5;100	105	240	0,1458
10	10/16	100	1,11	9;12	21	136	0,0556
				9;100	109	232	0,2976
				12;100	112	224	0,4567
6 possible means for $\bar{y}$							
7,00	8,50	**10,5**	**52,50**	54,50	56,00		
6 possible means for $\bar{y}_d$							
6,36	6,82	10,22	**21,56**	**35,54**	44,97		
$\bar{Y}$ = 31,5							
$E(\bar{y})$ = 31,5			$V(\bar{y})$ = 523,42			$MSE(\bar{y})$ = 523,42	
$E(\bar{y}_d)$ = 20,9			$V(\bar{y}_d)$ = 219,59			$MSE(\bar{y}_d)$ = 331,73	

Design weights and HT Estimator

- **Horvitz-Thompson** (1952) introduced an unbiased estimator for T (sum of Y) for any design, with or without replacement.
- The Horvitz-Thompson estimator (or π-estimator) of a total

$$\hat{t}_y^{HT} = \sum_{k \in S} \frac{y_k}{\pi_k} = \sum_{k \in S} d_k y_k$$

where,

$$\pi_k = \Pr(k \in S) = \sum_{k \in S} p(s)$$

and it is the selection or inclusion probability, and

$$d_k = \frac{1}{\pi_k}$$

is the sampling weight, for k ∈ S.

Design weights and HT Estimator

- The Horvitz-Thompson estimator is unbiased: E $(\hat{t}_y^{HT}) = t$
- The variance of the estimator is given by:

$$V(\hat{t}_y^{HT}) = \sum_{k,l \in U} (\pi_{kl} - \pi_k \pi_l)\left(\frac{y_k}{\pi_k}\right)\left(\frac{y_l}{\pi_l}\right)$$

where

$$\boldsymbol{\pi_k = \Pr(k \in S) = \sum_{k \in S} p(s)}$$

and

$$\boldsymbol{\pi_{kl} = Pr(k, l \in S) = \sum_{k,l \in S} p(s)}$$

The notation,

$$\boldsymbol{\Delta_{kl} = \pi_{kl} - \pi_k \pi_l}$$

Leads to more compact expression of this variance.

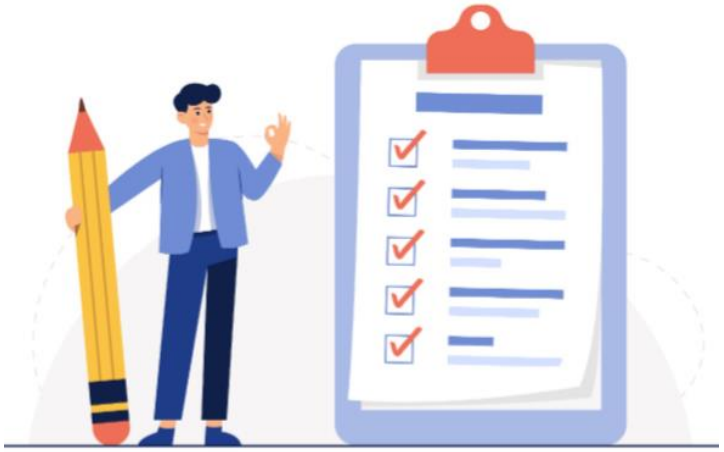
Design weights and HT Estimator

- The H-T variance is estimated by :

$$\hat{V}(\hat{t}_y^{HT}) = \sum_{k,l \in S} \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}} \left(\frac{y_k}{\pi_k}\right)\left(\frac{y_l}{\pi_l}\right)$$

If $\pi_{kl} > 0$ for all $k, l \in U$ then $\hat{V}(\hat{t}_y^{HT})$ is an unbiased estimator of $V(\hat{t}_y^{HT})$.

- ✓ Is direct estimator.
- ✓ Is a general estimator for a population total.
- ✓ Is an unbiased estimator under unequal sampling.
- ✓ Can be used for any probability sampling plan.
- ✓ Both sampling with and without replacement.



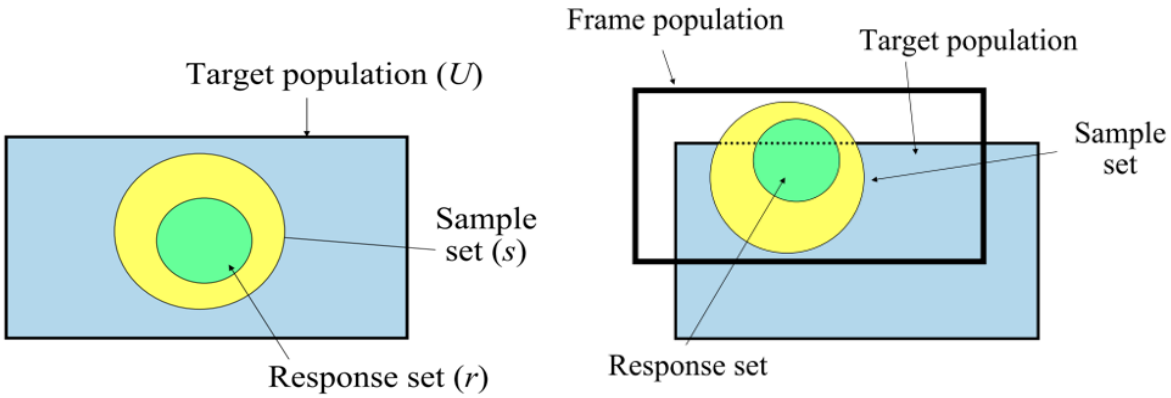
Nonresponse adjustment

If you don't provide the required information from the selected units which are eligible in the sample, then nonresponse problem occurs.

- **Nonresponse problem**
  - ✓ occurs in every survey
  - ✓ may cause estimates to be biased due to selective nonresponse
  - ✓ increase the variance in order to smaller sample sizes
  - ✓ is not easy to reduce.
  - ✓ is not easy to correct for the effect on the estimates.
- **Potential solutions of nonresponse problem**
  - ✓ Try to reduce nonresponse in the fieldwork.
  - ✓ Study to correct for nonresponse after the fieldwork.
  - ✓ Increase the usage of auxiliary variables.

Nonresponse adjustment

- **Main causes of nonresponse**
  - ✓ Not able (due to language problems)
  - ✓ Non-contact
  - ✓ Refusal
- **Minimizing nonresponse in the fieldwork**
  - ✓ Call back
  - ✓ Send information letter before fieldwork
  - ✓ Give assurance of confidentiality
  - ✓ Effective questionnaire design.



Nonresponse adjustment

- **Other methods to reduce nonresponse**
  - ✓ Multilingual interviewers
  - ✓ Translating questionnaires
  - ✓ Increase the number of contact
  - ✓ Longer fieldwork period
  - ✓ More evening calls than daytime calls
  - ✓ Personalizing the letter of invitation
  - ✓ Mentioning the duration of the survey
  - ✓ Interviewer training
  - ✓ Mixed-mode data collection **!!!**
  - ✓ Proxy respondents **!!!**

Nonresponse adjustment (Response models)

- **Fixed Response Model**
  - ✓ There is an assumption that the population consists of two strata (Respondents and Non-respondents)
  - ✓ Sample units in the Respondents stratum always provide required information.
  - ✓ On the contrary, Nonrespondents stratum never give information.

- **Random Response Model**
  - ✓ Each units of the sample has an unknown probability to respond.
  - ✓ The response probability is different for each unit.

Nonresponse adjustment (Response models)

- **Fixed Response Model**

✓ Response indicators $R_1, R_2, \ldots, R_N$, with

- $R_k = 1$ if element k is in Respondents stratum
- $R_k = 0$ if element k is in Non-respondent stratum

Response stratum	Nonresponse stratum
$N_R = \sum_{k=1}^{N} R_k$	$N_{NR} = \sum_{k=1}^{N} (1 - R_k)$
$\bar{Y}_R = \frac{1}{N_R} \sum_{k=1}^{N} R_k Y_k$	$\bar{Y}_{NR} = \frac{1}{N_{NR}} \sum_{k=1}^{N} (1 - R_k) Y_k$

$$N = N_R + N_{NR}$$

$$\bar{Y} = N_R \bar{Y}_R + N_{NR} \bar{Y}_{NR}$$

Source: Nonresponse in Household Surveys (ESTP Training Program)

Nonresponse adjustment (Response models)

- **Fixed Response Model**

✓ Sample indicators $a_1, a_2, \ldots, a_N$, with

- $a_k = 1$ if element k is in the sample
- $a_k = 0$ if element k is not in the sample

Response stratum	Nonresponse stratum
$n_R = \sum_{k=1}^{N} a_k R_k$	$n_{NR} = \sum_{k=1}^{N} a_k (1 - R_k)$
$\bar{y}_R = \frac{1}{n_R} \sum_{k=1}^{N} a_k R_k Y_k$	

$$n = n_R + n_{NR}$$

Source: Nonresponse in Household Surveys (ESTP Training Program)

$$\bar{y}_R = \frac{1}{n_R} \sum_{k=1}^{N} a_k R_k Y_k \qquad E(\bar{y}_R) = \bar{Y}_R \qquad B(\bar{y}_R) = \bar{Y}_R - \bar{Y} = \frac{N_{NR}}{N} \left(\bar{Y}_R - \bar{Y}_{NR} \right) = QK,$$

Nonresponse adjustment (Response models)

- **Random Response Model**
  - ✓ Each element k has an unknown response probability ρk
  - ✓ Response indicators R_1, R_2, …, R_N, with
    - $R_k = 1$ if element k responds
    - $R_k = 0$ if element k does not respond
    - $P(R_k = 1) = \rho_k$, $P(R_k = 0) = 1 - \rho_k$

$$E(\bar{y}_R) \approx \tilde{\bar{Y}} = \frac{1}{N}\sum_{k=1}^{N}\frac{\rho_k}{\bar{\rho}}Y_k \qquad B(\bar{y}_R) = \tilde{\bar{Y}} - \bar{Y} = \frac{R_{\rho Y} S_\rho S_Y}{\bar{\rho}}$$

  - ✓ The bias of the estimator is determined by
    - Correlation $R_{\rho Y}$ between response behaviour and target variable
    - Variation of response probabilities
    - Mean of response probabilities (expected response rate)

Nonresponse adjustment (Response models)

- **Simulation example for random model**

✓ Let's say we divide our population two responding groups according to response probability (ρ). Target variable (Y) will be correlated to auxiliary variable(X) which have two different response probability ρ_1 and ρ_2.

✓ N_1=5000 and N_2=5000, Y_1~N(2500,600) and Y_2~N(2000,600)

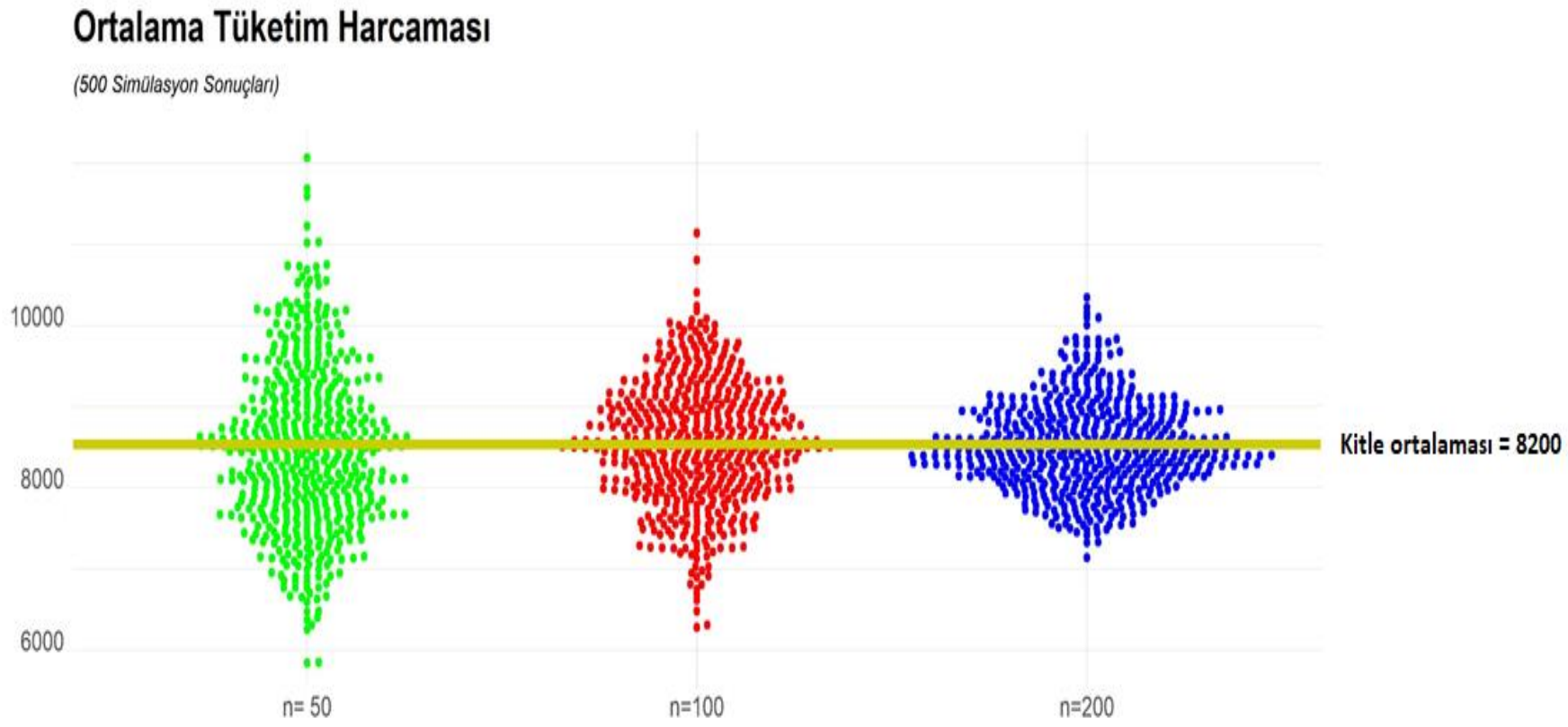
✓ Assume that $\bar{\rho}$ = %70, %80, %90 and %100 respectively.

Tablo 1. Y_1~N(2500,600), Y_2~N(2000,600) olmak üzere ortalama tahmin edicisinin HKO, sapma, varyans ve göreli sapma değerleri

		n=500				n=1000				n=2000			
		HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma	HKO	Sapma	Varyans	Göreli Sapma
$\bar{\rho}$=1	ρ_1=ρ_2=1	819,5692	-0,0082	819,5691	0	374,662	0,0244	374,6614	0,0002	166,4523	0,0211	166,4518	0,0003
$\bar{\rho}$=0.7	ρ_1=0.5 ρ_2=0.9	6275,989	-73,7089	842,9841	86,5681	5630,645	-72,1962	418,3491	92,5701	5612,811	-73,7542	173,129	96,9155
	ρ_1=0.6 ρ_2=0.8	2169,779	-36,7569	818,7084	62,2677	1588,385	-34,8158	376,2479	76,3126	1398,402	-35,0626	169,0183	87,9135
	ρ_1=0.7 ρ_2=0.7	1189,08	0,0035	1189,08	0	552,0083	0,0463	552,0062	0,0004	255,6198	0,0098	255,6197	0
$\bar{\rho}$=0.8	ρ_1=0.65 ρ_2=0.95	2877,022	-45,287	826,1123	71,2859	2921,111	-50,2332	397,7352	86,3841	2299,684	-46,1759	167,469	92,7177
	ρ_1=0.7 ρ_2=0.9	1796,468	-31,3284	814,9992	54,6333	1399,28	-31,841	385,43	72,4551	1065,699	-29,9206	170,4596	84,0049
	ρ_1=0.75 ρ_2=0.85	1086,545	-16,2581	822,2181	24,3273	609,2229	-15,3264	374,3251	38,557	426,267	-16,0577	168,4171	60,4902
	ρ_1=0.8 ρ_2=0.8	1026,897	0,0184	1026,897	0	478,9631	-0,0206	478,9627	0,0001	216,2103	0,0015	216,2103	0
$\bar{\rho}$=0.9	ρ_1=0.8 ρ_2=1	1661,444	-29,2543	805,6273	51,5104	1119,975	-27,096	385,783	65,5543	931,0403	-27,6018	169,1798	81,829
	ρ_1=0.85 ρ_2=0.95	976,3757	-13,2145	801,7537	17,8847	556,0694	-13,7752	366,3136	34,1245	356,719	-13,5829	172,2241	51,72
	ρ_1=0.9 ρ_2=0.9	893,6387	0,0286	893,6379	0,0001	429,3543	0,0327	429,3533	0,0002	192,5299	-0,0248	192,5293	0,0003

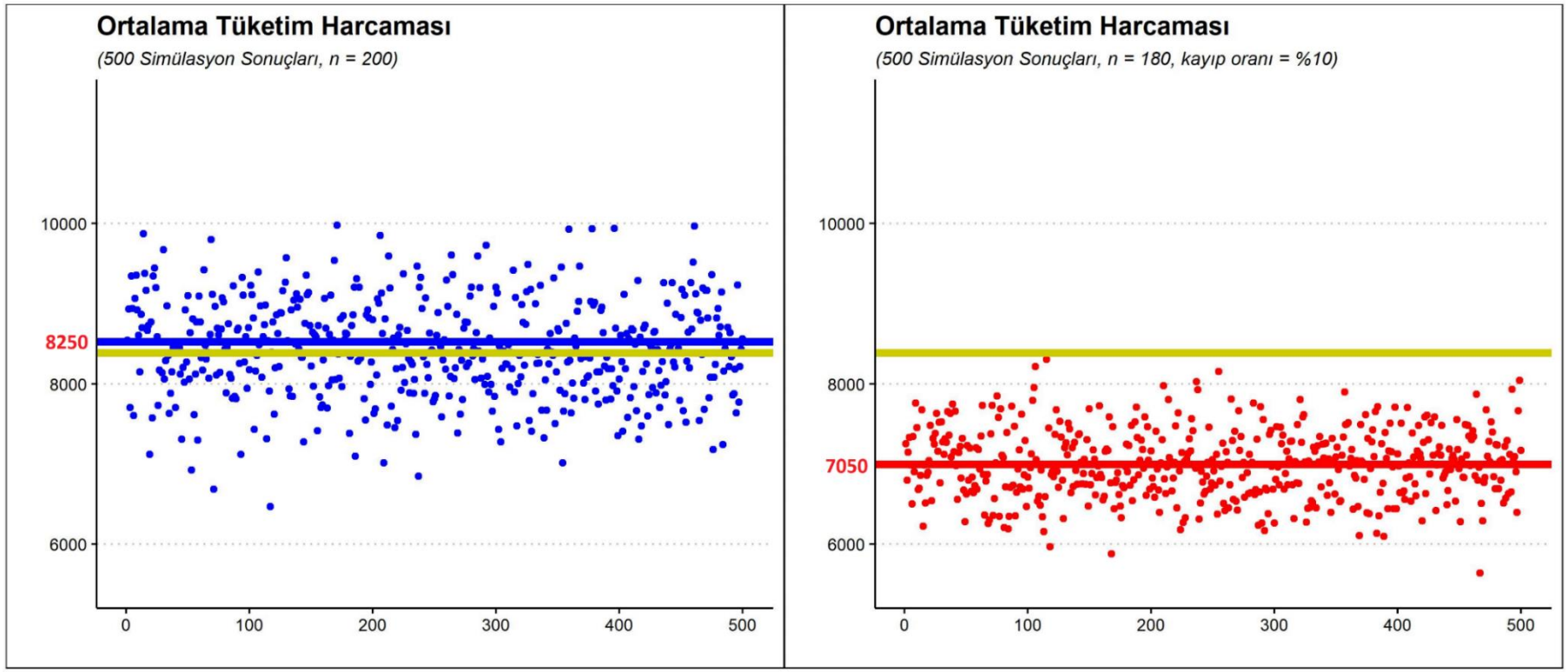
Nonresponse adjustment (Effect of selective nonresponse)

- **Sampling error**



Nonresponse adjustment (Effect of selective nonresponse)

- **Selective nonresponse effect**



Nonresponse adjustment

Missing Data Mechanisms

- ✓ Any analysis of data involving item or unit nonresponse requires some assumption about the missing data mechanism
  - Partition Y into an observed and an unobserved part

    $$Y = (Y_{obs}, Y_{mis})$$

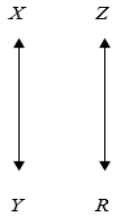
  - Distribution of missingness is characterized by the conditional distribution of R given Y

    $$P(R \mid Y) = P(R \mid Y_{obs}, Y_{mis})$$

Nonresponse adjustment

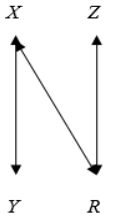
Missing Completely At Random (MCAR)

$$P(R \mid Y) = P(R)$$



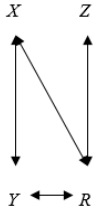
Missing At Random (MAR)

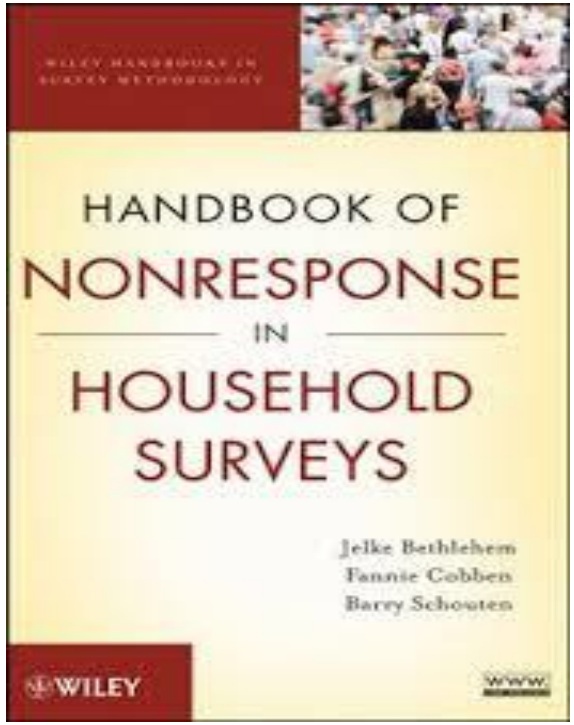
$$P(R \mid Y) = P(R \mid Y_{obs})$$



Not Missing At Random (NMAR)

$$P(R \mid Y) = P(R \mid Y_{obs}, Y_{mis})$$





Nonresponse adjustment

- ✓ To sum up, nonresponse adjustment factors should be used to reduce the effect of differences in response rates achieved in different parts of the sample.
- ✓ These factors can only be estimated in relation to characteristics that are known both respondent and non-respondent parts.
- ✓ Adjustment for non-response is important when nonresponse rates are high and vary for different parts.
- ✓ **The general strategy for nonresponse adjustment** is
  - Identify respondents who are similar to nonrespondents in terms of auxiliary information that is available both of two.
  - Increase the base weights of respondents by nonresponse adjustment factor for representing the similar nonrespondents.

Nonresponse adjustment

- Most common nonresponse adjustment method is cell-weighting (or called as **Response Homogenity Group [RHG]**) since not much information known about non-respondents.
- However, if auxiliary information are available alternative methods can provide more effective adjustment strategy.
- **Non-response adjustment with RHG**
  - ✓ Sample is divided to l=1,2...L group in order to an auxiliary information.(substrata, cluster, geographical or demographic characteristic)
  - ✓ It is assumed that response probabilities of the units in each group are the same and ρ_l
  - ✓ For the group l sample size is n_l, response set is c_l, and the number of respondents is m_l
  - ✓ The estimate of the response probability will be $\hat{\rho}_{k(l)} = m_l/n_l$ and the inverse of this
  - ✓ $\phi_{k(l)} = \frac{1}{\hat{\rho}_{k(l)}}$ l=1,2...L ve j=1,2...c_l used as nonresponse adjustment factor.

$$\hat{t}_y^{RHG} = \sum_l \sum_{k \in c_l} d_k \phi_{k(l)} y_k = \sum_l \sum_{k \in c_l} \frac{1}{\pi_k} \frac{1}{\hat{\theta}_{k(l)}} y_k = \sum_l \sum_{k \in c_l} d_j \frac{n_l}{m_l} y_k$$

References

Franklin, S., & Walker, C. (2010). Survey methods and practices. Statistics Canada. *Social Survey Methods Division, Ottawa*. (Originally published in October 2003)

Lohr, S. (2009). Sampling design and analysis (No. 519.52 L64).

Banning, R., Camstra, A., & Knottnerus, P. (2012). Sampling theory. Sampling design and estimation methods. Statistics Netherlands. The Hague, 87.

Lehtonen, R., & Djerf, K. (2008). Survey sampling reference guidelines: introduction to sample design and estimation techniques.

Verma, V. (2014). Sampling: An Introduction, University of Siena, Siena, September 2014.

Metin, C. B., Şahin Tekin, S. T., & Özdemir, Y. A. (2021). Restricted calibration and weight trimming approaches for estimation of the population total in business statistics. Journal of Applied Statistics, 48(13-15), 2658-2672.

ESTP Training Notes for Nonresponse in Household Surveys, Eurostat

Kalton, G., Flores-Cervantes, I. (2003). Weighting Methods. Journal of Official Statistics, Vol.19, No. 2,s. 81-97.