Design and Analysis of Experiments embedded in Probability Samples

Jan A. van den Brakel [1 2]
jbrl@cbs.nl

IASS Webinar, 22 September, 2022

[1] Statistics Netherlands, Department of Statistical Methods
[2] Maastricht University, Department of Quantitative Economics

Outline

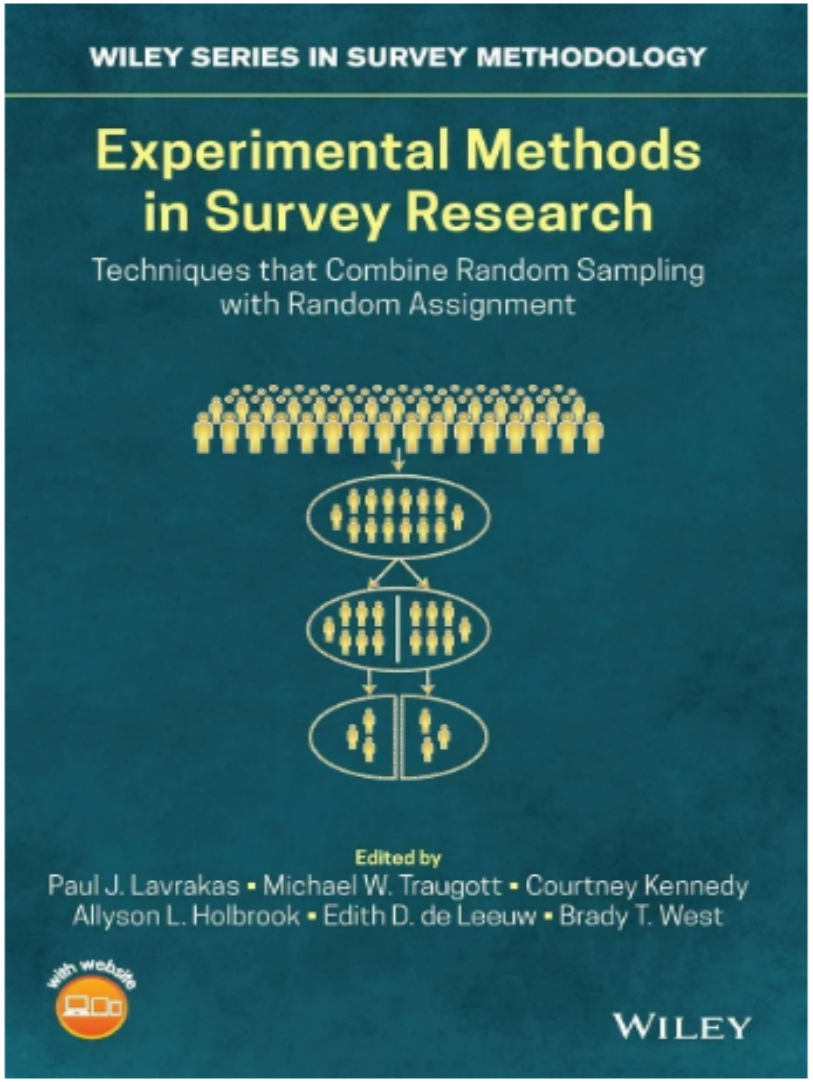
Introduction

Probability sampling and randomized experiments

- Two separated fields of applied statistics
- Unique similarity: design plays a crucial role
- Similarities and differences: Fienberg and Tanur (1987, 1988, 1989)
- Kish (1965), pp 595:
  ”The separation of sample surveys from experimental designs is an accident in the recent history of science, and it will diminish.”

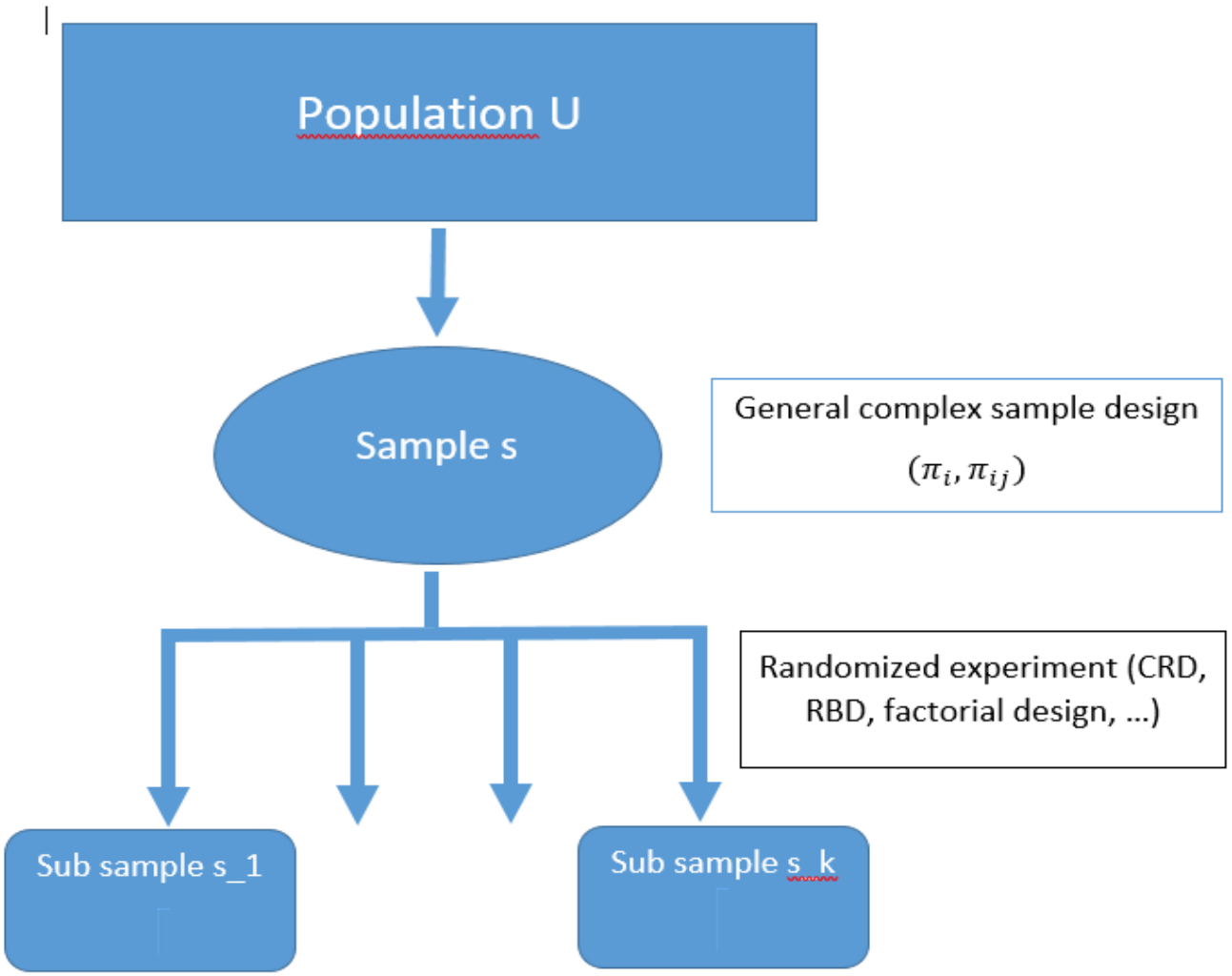
Introduction

Response to Kish:



Introduction

Embedded experiments:



Introduction

Embedded experiments:

- Probability sample from a finite population
- Randomize the sampling units over different treatments

Purpose:

- Estimation population parameters under different survey implementations
- Testing hypotheses about differences or treatment effects

Applications in survey methodology and official statistics:

- Improving survey process of repeated samples, e.g. new questionnaire, data collection modes
- Detecting and quantifying discontinuities
- Quantify non-sampling error sources

Introduction

Literature:

- Interpenetrating subsampling: Mahalanobis (1946), Hartley and Rao (1978)
- Split-ballot designs: Cochran (1977, section 13.15)
- Embedded experiments: Fienberg and Tanur (1987, 1988, 1989, 1996)
- Lavrakas et al (2019)
- ...

Introduction

Embedding randomized experiments in probability samples:

- Combine strong external validity with strong internal validity
- Design-based or model-based inference?
- Design-based inference generalizes results to a target population
- This paper: design-based inference frame work for
  - single factor experiments (CRD, RBD)
  - factorial desings (CRD, RBD)

Design considerations

Clear specifications of / decisions on

- Type and number of treatment factors and levels
- Hypotheses to be tested
  - Main effects
  - Interaction effects
- A prespecified set of target variables
- Minimum required sample size to observe
  - a prespecified treatment effect,
  - at a prespicified significane and power level

Design considerations

- Experimental design (Fienberg and Tanur, 1987, 1988, 1989)
  - Completely Randomized Design (CRD)
  - Randomized Block Design (RBD):
    - strata,
    - clusters,
    - PSU's
    - interviewers
  - Factorial designs versus single factor designs
    - Factorial design combines two or more factors in one experiment
    - Efficient (less sampling units)
    - Analyze interactions
  - Experimental units:
    - Ultimate sampling units
    - Cluster of sampling units (households, PSU's, sampling units assigend to the same interviewer)

Design considerations

- Field staff
  - Selection of interviewers
  - Assignment of interviewers to treatments
  - Are interviewers informed that they participate in an experiment
  - Additional training for the different treatments
  - ...

One-factor experiment at K levels

Probability sample:

- Select sample s size n from U of size N $(\pi_i, \pi_{ii'})$

Experimental design:

- Factor A at $K \geq 2$ levels
- Randomize n units over K treatments
- $\rightarrow K$ subsamples s_k of size n_k
- Designs:
  - Completely randomized design
  - Randomized block designs (strata, PSU's, clusters, interviewers as blocks)

Analysis

Purpose:

- Estimation finite population parameters under different survey implementations
- Testing hypothesis about differences between these subsample estimates

Design-based inference that accounts for:

- Sample design
- Experimental design

Analysis - Measurement error model

- Explains differences between finite population parameter under different survey implementations
- $y_{iqk} = u_i + \beta_k + \gamma_q + \epsilon_{ik}$
- Mixed interviewer effect: $\gamma_q = \psi + \varepsilon_q$
- Treatment effects: $\beta_k, k = 1, \ldots, K$
- Model assumptions:

$$E_m(\epsilon_{ik}) = 0, \quad E_m(\varepsilon_q) = 0$$

$$Cov_m(\varepsilon_q, \varepsilon_{q'}) = \begin{cases} \tau_q^2 & : \quad q = q' \\ 0 & : \quad q \neq q' \end{cases}$$

$$Cov_m(\epsilon_{ik}, \epsilon_{i'k'}) = \begin{cases} \sigma_{ik}^2 & : \quad i = i', k = k' \\ 0 & : \quad i \neq i', k = k' \end{cases}$$

$$Cov_m(\epsilon_{ik}, \varepsilon_q) = 0$$

Analysis - Hypotheses testing

- $\bar{Y}_k = 1/N \sum_{i \in U} y_{iqk}$
- $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_K)^t$
- Hypothesis: $H_0 : \mathbf{C}\bar{\mathbf{Y}} = \mathbf{0}$, $H_1 : \mathbf{C}\bar{\mathbf{Y}} \neq \mathbf{0}$
- Main effects: $K - 1$ contrasts:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ . & . & . & . & . \\ 1 & 0 & 0 & \ldots & -1 \end{pmatrix} \equiv (\mathbf{j}_{(K-1)} | -\mathbf{I}_{(K-1)})$$

- Under measurement error model:

  $\mathbf{C}\bar{\mathbf{Y}} = (\beta_1 - \beta_2, \ldots, \beta_1 - \beta_K)^t$
- Wald test: $W = \hat{\bar{\mathbf{Y}}}^t \mathbf{C}^t Cov(\mathbf{C}\hat{\bar{\mathbf{Y}}})^{-1} \mathbf{C}\hat{\bar{\mathbf{Y}}}$

Analysis - First order inclusion probabilities

Completely randomized designs:

- $P(i \in s_k|s) = n_k/n$
- First order inclusion probability s_k: $\pi_i^* = \pi_i(n_k/n)$

Randomized block designs:

- s divided in B blocks of size n_b
- Randomize n_b units over K subsamples of size n_{bk}
- $P(i \in s_k|s) = n_{bk}/n_b$
- First order inclusion probability s_k: $\pi_i^* = \pi_i(n_{bk}/n_b)$

Analysis - Parameter estimation

General regression estimator for $\bar{Y}_k$ (Särndal et al. (1992)):

- $\hat{\bar{Y}}_{k;r} = \hat{\bar{Y}}_{k;\pi} + \hat{\mathbf{b}}_k^t(\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_\pi)$
  - $\hat{\bar{Y}}_{k;\pi} = \frac{1}{N}\sum_{i=1}^{n_k} \frac{y_{ik}}{\pi_i^*}$,
  - $\hat{\bar{\mathbf{X}}}_\pi = \frac{1}{N}\sum_{i=1}^{n_k} \frac{\mathbf{x_i}}{\pi_\mathbf{i}^*}$,
  - $\hat{\mathbf{b}}_k = \left(\sum_{i=1}^{n_k} \frac{\mathbf{x_i}\mathbf{x_i^t}}{\omega_i^2\pi_i^*}\right)^{-1} \sum_{i=1}^{n_k} \frac{\mathbf{x_i} y_{ik}}{\omega_i^2\pi_i^*}$

General regression estimator for $\bar{\mathbf{Y}}$:

- $\hat{\bar{\mathbf{Y}}}_\mathbf{R} = (\hat{\bar{Y}}_{1;r}, \ldots, \hat{\bar{Y}}_{K;r})^t$

Analysis - Variance estimation

Covariance matrix of $\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}$

Sources of variation:

- measurement error model (m)
- sample design (s)
- experimental design (e)

$$Cov(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = Cov_m E_s E_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) + E_m Cov_s E_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) + E_m E_s Cov_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}})$$

Analysis - Variance estimation

- $Cov_m E_s E_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = \frac{1}{N^2}\sum_{i=1}^{N} \mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t$
- $E_m Cov_s E_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = \frac{1}{N^2}\sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1\right) \mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t$
- $E_m E_s Cov_e(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = E_m E_s \mathbf{CDC}^t - \frac{1}{N^2}\sum_{i=1}^{N} \frac{\mathbf{C}\boldsymbol{\Sigma}_i\mathbf{C}^t}{\pi_i}$
- where
  - $\boldsymbol{\Sigma}_i = Cov_m(\boldsymbol{\epsilon}_i), \qquad \boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iK})^t$
  - $\mathbf{D} = \mathbf{Diag}(d_1, \ldots, d_K)$

Collecting results:

- $Cov(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = E_m E_s \mathbf{CDC}^t$

Analysis - Variance estimation

Randomized block design:

$$d_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_b - 1} \sum_{i=1}^{n_b} \left(\frac{n_b \; e_{ik}}{N\pi_i} - \frac{1}{n_b} \sum_{i'=1}^{n_b} \frac{n_b \; e_{i'k}}{N\pi_{i'}} \right)^2$$

$$e_{ik} = y_{ik} - \mathbf{b}_k^t \mathbf{x}_i$$

Estimator: $\widehat{Cov}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$

$$\hat{d}_k = \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{bk} - 1} \sum_{i=1}^{n_{bk}} \left(\frac{n_b \; \hat{e}_{ik}}{N\pi_i} - \frac{1}{n_{bk}} \sum_{i'=1}^{n_{bk}} \frac{n_b \; \hat{e}_{i'k}}{N\pi_{i'}} \right)^2$$

$$\hat{e}_{ik} = y_{ik} - \hat{\mathbf{b}}_k^t \mathbf{x}_i$$

Completely randomized design:

$B = 1$, $n_b = n$, $n_{bk} = n_k$

Analysis - Variance estimation

Variance structure:

- Assumption of the measurement error model:
  - additive treatment effects equal for all elements in the population
  - measurement errors between sampling units are independent
- Weighting model that meets the condition that there is a vector $\mathbf{a}$ such that $\mathbf{a}^t\mathbf{x}_i = 1$ of all $i \in U$
- Super imposition of experimental design on the sampling design
- Variances of contrasts

Analysis - Wald test

Wald statistic

- $W = \hat{\bar{\mathbf{Y}}}_{\mathbf{R}}{}^{t}\mathbf{C}^{t}(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^{t})^{-1}\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}$ with $\hat{\mathbf{D}} = \mathbf{Diag}(\hat{d}_1, \ldots, \hat{d}_K)$
- If $\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}} \rightarrow \mathcal{N}\left((\mathbf{C}\bar{\mathbf{Y}}), Cov(\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}})\right)$ then $W \rightarrow \chi^2_{(K-1)}$

Analysis - Hypotheses about ratio's

- $\mathbf{R} = (R_1, \ldots, R_K)^t$ with $R_k = \frac{\bar{Y}_k}{\bar{Z}_k}$
- Hypothesis: $H_0 : \mathbf{CR} = \mathbf{0}$, $H_1 : \mathbf{CR} \neq \mathbf{0}$
- Wald test: $W = \hat{\mathbf{R}}^t \mathbf{C}^t Cov(\mathbf{C}\hat{\mathbf{R}})^{-1} \mathbf{C}\hat{\mathbf{R}}$
- with
  - $\hat{\mathbf{R}} = (\hat{R}_1, \ldots, \hat{R}_K)^t$, $\qquad \hat{R}_k = \frac{\hat{\bar{Y}}_{k;r}}{\hat{\bar{Z}}_{k;r}}$
  - $\widehat{Cov}(\mathbf{C}\hat{\mathbf{R}}) = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$, $\qquad \hat{\mathbf{D}} = \mathbf{Diag}(\hat{d}_1, \ldots, \hat{d}_K)$
  - RBD:
    $$\hat{d}_k = \frac{1}{\hat{\bar{Z}}^2_{k;r}} \sum_{b=1}^{B} \frac{1}{n_{bk}} \frac{1}{n_{bk}-1} \sum_{i=1}^{n_{bk}} \left(\frac{n_b \ \hat{e}_{ik}}{N\pi_i} - \frac{1}{n_{bk}} \sum_{i'=1}^{n_{bk}} \frac{n_b \ \hat{e}_{i'k}}{N\pi_{i'}} \right)^2$$
    $$\hat{e}_{ik} = (y_{ik} - \hat{\mathbf{b}}^t_{y;k}\mathbf{x}_i) - \hat{R}_{k;r}(z_{ik} - \hat{\mathbf{b}}^t_{z;k}\mathbf{x}_i)$$
  - CRD: $B = 1$, $n_b = n$, $n_{bk} = n_k$

Analysis - randomizing clusters over treatments

- Population U of size N
  - M PSU's of size N_j (and $N = \sum_{j=1}^{M} N_j$)
- Two stage sample s
  - First stage: m PSU's with selection probabilities π_j^I
  - Second stage: n_j SSU's with selection probabilities $\pi_{i|j}^{II}$
  - $n = \sum_{j=1}^{m} n_j$
- Experimental design:
  - Randomize m PSU's over K subsamples s_k of size m_k
  - CRD:
    - $P(j \in s_k|s) = \frac{m_k}{m}$
    - $\pi_j^{*I} = P(j \in s_k) = \frac{m_k}{m}\pi_j^I$
  - RBD: m PSU's grouped in B blocks of size m_b
    - $P(j \in s_k|s) = \frac{m_{bk}}{m_b}$
    - $\pi_j^{*I} = P(j \in s_k) = \frac{m_{bk}}{m_b}\pi_j^I$

Analysis - randomizing clusters over treatments

Measurement error model:

- $y_{ijqk} = u_{ij} + \beta_k + \gamma_q + \epsilon_{ijk}$
- Mixed interviewer effect: $\gamma_q = \psi + \varepsilon_q$
- Treatment effects: $\beta_k, k = 1, \ldots, K$
- Model assumptions:

$$E_m(\epsilon_{ijk}) = 0 \quad E_m(\varepsilon_q) = 0,$$

$$Cov_m(\varepsilon_q, \varepsilon_{q'}) = \begin{cases} \tau_q^2 & : \quad q = q' \\ 0 & : \quad q \neq q' \end{cases}$$

$$Cov_m(\epsilon_{ijk}, \epsilon_{i'j'k'}) = \begin{cases} \sigma_{ijk}^2 + \sigma_{jk}^2 & : \quad i = i', j = j', k = k' \\ \sigma_{jk}^2 & : \quad i \neq i', j = j', k = k' \\ 0 & : \quad i \neq i', j \neq j', k = k' \end{cases}$$

$$Cov_m(\epsilon_{ijk}, \varepsilon_q) = 0$$

Analysis - randomizing clusters over treatments

- Hypothesis: $H_0 : \mathbf{C}\bar{\mathbf{Y}} = \mathbf{0}$, $H_1 : \mathbf{C}\bar{\mathbf{Y}} \neq \mathbf{0}$
- Wald test: $W = \hat{\hat{\mathbf{Y}}}_{\mathbf{R}}{}^{t}\mathbf{C}^{t}\widehat{Cov}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}})^{-1}\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}$
- $\hat{\hat{\mathbf{Y}}}_{\mathbf{R}} = (\hat{\hat{Y}}_{1;r}, \ldots, \hat{\hat{Y}}_{K;r})^t$
- $\hat{\hat{Y}}_{k;r} = \hat{\hat{Y}}_{k;\pi} + \hat{\mathbf{b}}_k^t(\bar{\mathbf{X}} - \hat{\hat{\mathbf{X}}}_{\pi})$
  - $\hat{\hat{Y}}_{k;\pi} = \frac{1}{N}\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{y_{ijk}}{\pi_j^{*I}\pi_{i|j}^{II}}$,
  - $\hat{\hat{\mathbf{X}}}_{\pi} = \frac{1}{N}\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}}{\pi_j^{*I}\pi_{i|j}^{II}}$,
  - $\hat{\mathbf{b}}_k = \left(\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^t}{\omega_{ij}^2\pi_j^{*I}\pi_{i|j}^{II}}\right)^{-1}\sum_{j=1}^{m_k}\sum_{i=1}^{n_j}\frac{\mathbf{x}_{ij}y_{ijk}}{\omega_{ij}^2\pi_j^{*I}\pi_{i|j}^{II}}$

Analysis - randomizing clusters over treatments

- $\widehat{Cov}(\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}) = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t, \qquad \hat{\mathbf{D}} = \mathbf{Diag}(\hat{d}_1, \ldots, \hat{d}_K)$
- RBD:

$$\hat{d}_k = \sum_{b=1}^{B} \frac{1}{m_{bk}} \frac{1}{m_{bk}-1} \sum_{j=1}^{m_{bk}} \left(\frac{m_b \ \hat{e}_{ik}}{N\pi_j^I} - \frac{1}{m_{bk}} \sum_{j'=1}^{m_{bk}} \frac{m_b \ \hat{e}_{j'k}}{N\pi_{j'}^I} \right)^2$$

$$\hat{e}_{jk} = \sum_{i=1}^{n_j} \frac{(y_{ijk} - \hat{\mathbf{b}}_k^t \mathbf{x}_{ij})}{\pi_{i|j}^{II}}$$

- CRD: $B = 1$, $m_b = m$, $m_{bk} = m_k$
- Randomizing clusters of sampling units assigned to the same interviewer
- Extension to analyis of ratio's

Analysis - simulation

Artificial population

			Value target parameter			
Stratum	Nr PSU's	Nr SSU's	Mean	St.Dev.	Min	Max
1	70	6,250	22,183	12,001	7,607	50,915
2	130	18,250	6,128	1,866	3,007	10,490
3	250	85,000	1,407	732	512	3,248
Total	450	109,500	3,380	5,803	512	50,915

Analysis - simulation

Sample design: stratified two-stage sampling

Stratum	Nr PSU's	Nr SSU's
1	25	900
2	30	1,080
3	50	1,800
Total	105	3,780

Analysis - simulation

Experimental design

Experimental design		Treatment effects			
		β_1	β_2	β_3	β_4
CRD	RBD	0	0	0	0
CRD	RBD	0	20	40	60
CRD	RBD	0	40	80	120
CRD	RBD	0	80	160	240

Analysis - simulation

Simulation:

- Repeatedly drawing samples from the finite population
- For each replicate generate measurement errors with variances proportional to the target variables
- Randomize the SSU's of the sample over 4 treatments with a CRD and RBD
- Add treatment effects
- Estimate contrasts, variances and evaluate the Waldtest
- Number of replicates: $R = 100{,}000$

Analysis - simulation

Analysis of the simulation:

- For each replicate r:
  - Hájek's ratio estimator for sample mean: $\hat{Y}_k = (\sum_{i \in S_k} \frac{1}{\pi_i^*})^{-1} \sum_{i \in S_k} \frac{y_i}{\pi_i^*}$, $k = 1, \ldots, 4$.
  - Sample means: $\hat{\mathbf{Y}}^r = (\hat{Y}_1^r, \ldots, \hat{Y}_4^r)^t$
  - Contrasts $\mathbf{C}\hat{\mathbf{Y}}^r$ with variance: $\mathbf{C}\hat{\mathbf{D}}^r\mathbf{C}^t$
  - Wald statistic: $W^r = (\mathbf{C}\hat{\mathbf{Y}}^r)^t(\mathbf{C}\hat{\mathbf{D}}^r\mathbf{C}^t)^{-1}(\mathbf{C}\hat{\mathbf{Y}}^r)$
- Means over the R =100,000 resamples:
  - $\tilde{\mathbf{Y}} = \frac{1}{R}\sum_{r=1}^{R} \hat{\mathbf{Y}}^r$
  - Contrasts $\mathbf{C}\tilde{\mathbf{Y}}$
  - Variance: $\mathbf{C}\tilde{\mathbf{D}}\mathbf{C}^t = \frac{1}{R}\sum_{r=1}^{R} \mathbf{C}\hat{\mathbf{D}}^r\mathbf{C}^t$
  - Wald statistic: $\tilde{W} = \frac{1}{R}\sum_{r=1}^{R} W^r$

Analysis - simulation

Analysis of the simulation:

- Point estimates: $\mathbf{C}\tilde{\mathbf{Y}} \approx \mathbf{C}\beta$
- Variance: $\mathbf{C}\tilde{\mathbf{D}}\mathbf{C}^t$ should be approximately equal to:

$$\mathbf{CVC}^t = \frac{1}{R}\sum_{r=1}^{R}\mathbf{C}(\hat{\mathbf{Y}}^r - \tilde{\mathbf{Y}})(\hat{\mathbf{Y}}^r - \tilde{\mathbf{Y}})^t\mathbf{C}^t$$

- Distribution of the Wald test:
  - If $\mathbf{C}\hat{\mathbf{Y}} \to \mathcal{N}(\mathbf{C}\beta, \mathbf{CVC}^t)$, then $W \to \chi^2_{[K-1][\delta]}$ with n.c.p. $\delta = (1/2)(\mathbf{C}\beta)^t(\mathbf{CVC}^t)^{-1}(\mathbf{C}\beta)$
  - Power of the Waldtest: $P(\chi^2_{[K-1][\delta]} > \chi^2_{[1-\alpha][K-1]})$ should be approximately equal to simulated power:

$$P^{sim}(W) = \frac{1}{R}\sum_{r=1}^{R} I(W^r > \chi^2_{[1-\alpha][K-1]})$$

  - $\tilde{W} \approx E(\chi^2_{[K-1][\delta]}) = (K-1) + 2\delta$

Analysis - simulation

Results RBD $\boldsymbol{\beta} = (0, 80, 160, 240)^t$
Population mean: 3,380

Subsamples			Contrasts			
k	β_k	$\tilde{Y}_k$	$k - k'$	$\mathbf{C}\tilde{\mathbf{Y}}$	$\mathbf{CVC}^t$	$\mathbf{C}\tilde{\mathbf{D}}\mathbf{C}^t$
1	0	3,390			diagonal elements	
2	80	3,470	1-2	-80	6,204	6,180
3	160	3,550	1-3	-160	6,210	6,181
4	240	3,630	1-4	-240	6,214	6,181

Analysis - simulation

Results RBD $\boldsymbol{\beta} = (0, 80, 160, 240)^t$

	Wald statistic	
α	$P(W)$	$P^{sim}(W)$
0.050	0.77340	0.77712
0.025	0.68135	0.68789
0.010	0.55796	0.56701

$\tilde{W}$ = 13.4859
δ = 5.1331
$E(\chi^2_{[K-1][\delta]})$ = 13.2662

Analysis - simulation

Results for one randomly chosen resample
RBD $\boldsymbol{\beta} = (0, 80, 160, 240)^t$

Design-based Wald statistic RBD								
Subsamples			Contrasts			Wald statistic		
k	β_k	$\hat{Y}_k$	$k-k'$	$\hat{Y}_k - \hat{Y}_{k'}$	$\sqrt{\hat{d}_k + \hat{d}_{k'}}$	W	df	p
1	0	3,395						
2	80	3,420	1-2	-25	81.247	9.93	3	0.019
3	160	3,515	1-3	-120	80.697			
4	240	3,626	1-4	-231	82.383			

Analysis - simulation

Results for one randomly chosen resample
RBD $\boldsymbol{\beta} = (0, 80, 160, 240)^t$

Standard ANOVA RBD: Descriptives				
Subsamples			Contrasts	
k	β_k	$\hat{Y}_k$	$k - k'$	$\hat{Y}_k - \hat{Y}_{k'}$
1	0	8,815		
2	80	8,150	1-2	665
3	160	8,566	1-3	249
4	240	8,746	1-4	69

Analysis - simulation

Results for one randomly chosen resample
RBD $\beta = (0, 80, 160, 240)^t$

Standard ANOVA RBD: F-test				
Source	*df*	*MSS*	*F*	*p*
Between blocks	2	1.6773 E+11		
Between treatments	3	84,377,277	1.99	0.1126
Residual	3,774	42,310,035		
Total	3,779	131,089,505		

$K \times L$ factorial experiments

Probability sample:

- Select sample s size n from U of size N $(\pi_i, \pi_{ii'})$

Experimental design:

- Factor A at $K \geq 2$ levels
- Factor B at $L \geq 2$ levels
- Randomize n units over KL treatment combinations
- $\rightarrow$ KL subsamples s_{kl} of size n_{kl}
- Designs:
  - Completely randomized design
  - Randomized block designs (strata, PSU's, clusters, interviewers as blocks)

$K \times L$ factorial experiments

Measurement error model

- $y_{iqkl} = u_i + \beta_{kl} + \gamma_q + \epsilon_{ikl}$
- Mixed interviewer effect: $\gamma_q = \psi + \varepsilon_q$
- Treatment effects ANOVA: $\beta_{kl} = A_k + B_l + AB_{kl}$
- Restrictions:
  $\sum_{k=1}^{K} A_k = 0, \quad \sum_{l=1}^{L} B_l = 0,$
  $\sum_{k=1}^{K} AB_{kl} = 0, \quad \sum_{l=1}^{L} AB_{kl} = 0,$
- Model assumptions:

$$E_m(\epsilon_{ik}) = 0, \quad E_m(\varepsilon_q) = 0$$

$$Cov_m(\varepsilon_q, \varepsilon_{q'}) = \begin{cases} \tau_q^2 & : \quad q = q' \\ 0 & : \quad q \neq q' \end{cases}$$

$$Cov_m(\epsilon_{ik}, \epsilon_{i'k'}) = \begin{cases} \sigma_{ik}^2 & : \quad i = i', k = k' \\ 0 & : \quad i \neq i', k = k' \end{cases}$$

$$Cov_m(\epsilon_{ik}, \varepsilon_q) = 0$$

$K \times L$ factorial experiments

Hypotheses:

- $\bar{Y}_{kl} = 1/N \sum_{i \in U} y_{iqkl}$
- $\bar{\mathbf{Y}} = (\bar{Y}_{11}, \ldots, \bar{Y}_{kl}, \ldots, \bar{Y}_{KL})^t$
- Hypothesis: $H_0 : \mathbf{C}\bar{\mathbf{Y}} = \mathbf{0}$, $H_1 : \mathbf{C}\bar{\mathbf{Y}} \neq \mathbf{0}$
  - Main effects A: $K-1$ contrasts
    $\mathbf{C_A} = \frac{1}{L}\left(\mathbf{j}_{(K-1)} | -\mathbf{I}_{(K-1)}\right) \otimes \mathbf{j}_L^t \equiv \frac{1}{L}\tilde{\mathbf{C}}_\mathbf{A} \otimes \mathbf{j}_L^t$
  - Main effects B: $L-1$ contrasts
    $\mathbf{C_B} = \frac{1}{K}\mathbf{j}_K^t \otimes \left(\mathbf{j}_{(L-1)} | -\mathbf{I}_{(L-1)}\right) \equiv \frac{1}{K}\mathbf{j}_K^t \otimes \tilde{\mathbf{C}}_\mathbf{B}$
  - Interactions: $(K-1)(L-1)$ contrasts
    $\mathbf{C_{AB}} = \left(\mathbf{j}_{(K-1)} | -\mathbf{I}_{(K-1)}\right) \otimes \left(\mathbf{j}_{(L-1)} | -\mathbf{I}_{(L-1)}\right) = \tilde{\mathbf{C}}_\mathbf{A} \otimes \tilde{\mathbf{C}}_\mathbf{B}$

$K \times L$ factorial experiments

Under measurement error model:

- $\mathbf{C_A}\bar{\mathbf{Y}} = \mathbf{C_A}\beta = (A_1 - A_2, \ldots, A_1 - A_K)^t$
- $\mathbf{C_B}\bar{\mathbf{Y}} = \mathbf{C_B}\beta = (B_1 - B_2, \ldots, B_1 - B_L)^t$
- $\mathbf{C_{AB}}\bar{\mathbf{Y}} = \mathbf{C_{AB}}\beta = (AB_{11} - AB_{12} - AB_{21} + AB_{22}, \ldots,$
  $AB_{11} - AB_{1L} - AB_{21} + AB_{2L}, \ldots,$
  $AB_{11} - AB_{12} - AB_{K1} + AB_{K2}, \ldots,$
  $AB_{11} - AB_{1L} - AB_{K1} + AB_{KL})^t$

Wald test: $W = \hat{\bar{\mathbf{Y}}}^t \mathbf{C}^t Cov(\mathbf{C}\hat{\bar{\mathbf{Y}}})^{-1} \mathbf{C}\hat{\bar{\mathbf{Y}}}$

$K \times L$ factorial experiments

Inclusion probabilities

- CRD: s_{kl}: $\pi_i^* = \pi_i(n_{kl}/n)$
- RBD: s_{kl}: $\pi_i^* = \pi_i(n_{bkl}/n_b)$

General regression estimator for $\bar{Y}_{kl}$:

- $\hat{\bar{Y}}_{kl;r} = \hat{\bar{Y}}_{kl;\pi} + \hat{\mathbf{b}}_{kl}^t(\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_\pi)$
  - $\hat{\bar{Y}}_{kl;\pi} = \frac{1}{N}\sum_{i=1}^{n_{kl}} \frac{y_{ikl}}{\pi_i^*}$,
  - $\hat{\bar{\mathbf{X}}}_\pi = \frac{1}{N}\sum_{i=1}^{n_{kl}} \frac{\mathbf{x}_i}{\pi_i^*}$,
  - $\hat{\mathbf{b}}_{kl} = \left(\sum_{i=1}^{n_{kl}} \frac{\mathbf{x}_i\mathbf{x}_i^t}{\omega_i^2\pi_i^*}\right)^{-1} \sum_{i=1}^{n_{kl}} \frac{\mathbf{x}_i y_{ikl}}{\omega_i^2\pi_i^*}$
- $\hat{\bar{\mathbf{Y}}}_{\mathbf{R}} = (\hat{\bar{Y}}_{11;r}, \ldots, \hat{\bar{Y}}_{kl;r}, \ldots, \hat{\bar{Y}}_{KL;r})^t$

$K \times L$ factorial experiments

Randomized block design:

$$Cov(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = E_m E_s \mathbf{C}\mathbf{D}\mathbf{C}^t$$

$$d_{kl} = \sum_{b=1}^{B} \frac{1}{n_{bkl}} \frac{1}{n_b - 1} \sum_{i=1}^{n_b} \left(\frac{n_b \; e_{ikl}}{N\pi_i} - \frac{1}{n_b} \sum_{i'=1}^{n_b} \frac{n_b \; e_{i'kl}}{N\pi_{i'}} \right)^2$$

$$e_{ikl} = y_{ikl} - \mathbf{b}_{kl}^t \mathbf{x}_i$$

Estimator: $\widehat{Cov}(\mathbf{C}\hat{\hat{\mathbf{Y}}}_{\mathbf{R}}) = \mathbf{C}\hat{\mathbf{D}}\mathbf{C}^t$

$$\hat{d}_{kl} = \sum_{b=1}^{B} \frac{1}{n_{bkl}} \frac{1}{n_{bkl} - 1} \sum_{i=1}^{n_{bkl}} \left(\frac{n_b \; \hat{e}_{ikl}}{N\pi_i} - \frac{1}{n_{bkl}} \sum_{i'=1}^{n_{bkl}} \frac{n_b \; \hat{e}_{i'kl}}{N\pi_{i'}} \right)^2$$

$$\hat{e}_{ikl} = y_{ikl} - \hat{\mathbf{b}}_{kl}^t \mathbf{x}_i$$

Completely randomized design:

$B = 1$, $n_b = n$, $n_{bkl} = n_{kl}$

$K \times L$ factorial experiments

Wald statistic

- $W = \hat{\bar{\mathbf{Y}}}_{\mathbf{R}}{}^{t}\mathbf{C}^{t}(\mathbf{C}\hat{\mathbf{D}}\mathbf{C}^{t})^{-1}\mathbf{C}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}$ with $\hat{\mathbf{D}} = \mathbf{Diag}(\hat{d}_{11}, \ldots, \hat{d}_{KL})$
- Main effects A: $\mathbf{C_A}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}$ with $W \sim \chi^2_{K-1}$
- Main effects B: $\mathbf{C_B}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}$ with $W \sim \chi^2_{L-1}$
- Interaction effects between A and B: $\mathbf{C_{AB}}\hat{\bar{\mathbf{Y}}}_{\mathbf{R}}$ with $W \sim \chi^2_{(K-1)(L-1)}$

Extensions factorial designs

- Higher order factorial designs
  - Similar parameter and variance estimators
  - Different definitions for contrast matrices
- Hypotheses about ratio's of estimated population totals
  - Adaption variance estimators
- Experimental units clusters of sampling units
  - E.g. households, or persons assigned to the same interviewer
  - Adaption variance estimators

Special cases

- Self-weighted sample design and balanced experimental designs
  - $K \times L$ factorial CRD:
    Wald test main effect equal to F-test ANOVA two-way layout
  - $K \times L$ factorial RBD:
    Wald test main effect equal to F-test ANOVA three-way layout
- Single factor designs and self-weighted sample design
  - CRD: Wald test equal to F-test ANOVA one-way layout
  - RBD: Wald test equal to F-test ANOVA two-way layout
- Single factor designs with two levels
  - Design-based version of t-test
  - Self-weighted sample design: Welch t-test or t-test

Application

Mixed-mode experiment in the Dutch Crime Victimization Survey (CVS)

- Survey design:
  - Annual survey
  - Publication domains: 25 police regions
  - Stratified sample using police regions as strata
  - Sample size:
    - Minimum 750 respondents per domain
    - Net response of about 19,000 respondents
    - With a response of 65% a gross sample of 30,500
  - Data collection: mixed mode using CATI and CAPI
  - Inference: general regression estimator

Application

Mixed-mode experiment Dutch CVS:

- Purpose: effect of sequential mixed mode design on response rates and target parameters
- Sequential mixed-mode design
  - Starts with Web Interviewing
  - Non-respondent follow-up with CATI and CAPI
- Embedded experiment:
  - Control group: regular CVS with 30,500 persons (net response 19,000)
  - Treatment group: 3750 persons (net repsonse 2350)
  - RBD using strata as block variables

Application

Mixed-mode experiment Dutch CVS:

- Target variables:
  - Satispol: percentage of people satisified with police performance
  - Nuisance: Mean perceived amount of irritation due to antisocial behaviour (10 point scale)
  - Propvic: Percetage of people victem of property crime
  - Violvic: Percentage of people victem of violent crime
  - Repvic: Victimization rate of reported crimes

Application

Observable differences at 5% significance and different power levels

	Observ. dif. Bonferroni t-test					
	Applied design 19,000 versus 2350			**Balanced design** 10,675 versus 10,675		
	Power			Power		
Variable	50%	80%	90%	50%	80%	90%
Satispol	4.55	6.49	7.51	2.85	4.07	4.70
Nuisance	0.06	0.09	1.00	0.04	0.05	0.06
Propvic	2.06	2.94	3.41	1.29	1.84	2.13
Violvic	2.30	3.29	3.81	1.44	2.06	2.38
Repvic	4.49	6.41	7.41	2.81	4.01	4.64

Application

Results

Var.	Control		Treatment		Difference		
	$\hat{\bar{Y}}_1$	$\sqrt{\hat{d}_1}$	$\hat{\bar{Y}}_2$	$\sqrt{\hat{d}_2}$	$\hat{\Delta}^{(*)}$	$t^{(**)}$	p
Satisp	55.46	(0.75)	53.95	(2.35)	1.51	0.610	0.542
Nuis	2.94	(0.01)	3.19	(0.03)	-0.25	-6.958	0.000
Propv	16.02	(0.38)	14.47	(1.06)	1.55	1.377	0.169
Violv	8.47	(0.34)	8.71	(1.00)	-0.23	-0.220	0.826
Repv	35.95	(0.74)	34.36	(2.20)	1.60	0.492	0.689

$*): \quad \hat{\Delta} = \hat{\bar{Y}}_1 - \hat{\bar{Y}}_2$

$**): \quad t = \frac{\hat{\Delta}}{\sqrt{\hat{d}_1 + \hat{d}_2}}$

Discussion

Embedding randomized experiments in probability samples:

- Potentially strong internal and external validity
- Useful to generalize results to large target population
- Design-based inference for a large set of designs
- Many applications in official statistics
- Software: X-tool (Blaise component)

Acknowledgement

- Prof. S.E. Fienberg, Prof P. Kooiman, Dr. R. Renssen
- Dr. D. Binder
- Referees

Thank you for your attention

- Brakel, J.A. van den (2019). Design based analysis of experiments embedded in probability pamples, in P.J. Lavrakas et al. (eds.) *Experimental Methods in Survey Research*, Ch. 23, Wiley & Sons, Hoboken, USA.
- Brakel, J.A. van den (2013). Design based analysis of factorial designs embedded in probability samples, *Survey Methodology*, vol. 39, pp. 323-349.
- Brakel, J.A. van den, (2008). Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey, *Journal of the Royal Statistical Society, Series A*, vol. 171, pp. 581-613
- Brakel, J.A. van den and R.H. Renssen, (2005). Analysis of experiments embedded in complex sampling designs, *Survey Methodology*, vol. 31, nr. 1, pp. 23-40.
- Chipperfield, J. and P. Bell, (2010). Embedded experiments in repeated and overlapping surveys., *Journal of the Royal Statistical Society, A series*, vol. 173, pp. 51-66.