Statistiska centralbyrån Statistics Sweden

Aspects of Responsive Design for the Swedish Living Conditions Survey (LCS)

Peter Lundquist and Carl-Erik Särndal

Nürnberg 2011

Introduction

A short theoretical presentation

- Balance indicators and distance measure in the paper by Särndal in JOS 2011
- R-indicators by Schouten, Bethlehem et al. in the RISQ-project www.R-indicator.eu
- Responsive design Groves and Heeringa in JRSS A 2006
- Empirical results for the Swedish LCS 2009

Introduction: The Swedish background

- High demands on response rate (from clients)
- Expensive to chase respondents
- Studies on LFS, LCS and HF raises questions about the value of today’s field work strategy
- Panel surveys have other needs (measures over time) than one-time surveys

LCS 2009

LCS is a telephone survey. It is used in this study without supplement, this gives a simple random sample design (SRS) *from Swedish RTP* with a sample size of n = 8,220

- Response rate ordinary field work (5 w): 60.4%
- Final response rate after follow-up (+3,3w) : 67.4%

The same data collection strategy is used in the follow-up.

Relative difference: *RDF*

Register variables (known for s) used as y-variables

Standard x-vector: {Phone, High education, Four Age-groups, Property ownership, Swedish origin}

$$\hat{Y}_{CAL} = \sum_r d_k m_k y_k \qquad m_k = \left(\sum_s d_k \mathbf{x}_k\right)' \left(\sum_r d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k$$

$$\hat{Y}_{FUL} = \sum_s d_k y_k \qquad d_k = 1/\pi_k \quad \text{where} \quad \pi_k = \Pr(k \in s)$$

$$RDF = 100(\hat{Y}_{CAL} - \hat{Y}_{FUL}) / \hat{Y}_{FUL}$$

Statistiska centralbyrån Statistics Sweden

The LCS 2009 data collection:

Progression of the response rate P (in per cent) and of RDF for three selected register variables. Computations are based on the standard x-vector.

Step in the data collection	$100 \times P$	*RDF* Sickness benefits	Income	Employed
Attempt 1	12.8	10.5	-0.05	-1.3
Attempt 2	24.6	3.3	-1.1	-2.0
Attempt 3	32.8	1.6	-0.4	0.2
Attempt 8	53.0	1.0	2.4	2.4
End ordinary field work	60.4	-0.9	3.3	2.9
Final	67.4	-3.6	2.9	3.1

Statistiska centralbyrån Statistics Sweden

Balance indicators

Matrix language is needed because of the multivariate nature of $\mathbf{x}_k$. Let $\mathbf{D} = \bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s = (D_1, ..., D_j, ..., D_J)'$. Under perfect balance, $\mathbf{D} = \mathbf{0}$, the zero vector. But normally, $\mathbf{D} \neq \mathbf{0}$.

A univariate measure, ***lack of balance***, is filled by the quadratic form

$$\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D} = (\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s)'\mathbf{\Sigma}_s^{-1}(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_s)$$

where $\bar{\mathbf{x}}_r = \sum_r d_k \mathbf{x}_k / \sum_r d_k$ and $\bar{\mathbf{x}}_s = \sum_s d_k \mathbf{x}_k / \sum_s d_k$ and the weighting matrix is $\mathbf{\Sigma}_s = \sum_s d_k \mathbf{x}_k \mathbf{x}_k' / \sum_s d_k$.

Increased mean differences D_j tend to increase $\mathbf{D}'\mathbf{\Sigma}_s^{-1}\mathbf{D}$.

Balance indicators

Statistiska centralbyrån Statistics Sweden

It can be shown (Särndal, 2011) that $0 \le \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} \le Q-1$ where $Q = 1/P$.

Hence, $(\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D})/(Q-1)$ measures lack of balance on a unit interval scale.

We examine several **balance indicators** measured on the unit interval scale and such that the value "1" implies perfect balance. The first is

$$BI_1 = 1 - \sqrt{\frac{\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}}{Q-1}}$$

Because $P(1-P) \le 1/4$, an alternative indicator also contained in the unit interval is

$$BI_2 = 1 - 2P\sqrt{\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}}$$

Balance indicators alt. description

$$BI_1 = 1 - S_{\hat{\theta}} \Big/ \sqrt{P(1-P)} \quad , \; BI_2 = 1 - 2S_{\hat{\theta}}$$

can be interpreted through the variance of estimated response probabilities $\hat{\theta}_k$ for $k \in s$,

$$S^2_{\hat{\theta}} = \sum_s d_k (\hat{\theta}_k - \bar{\hat{\theta}}_s)^2 / \sum_s d_k$$

if ordinary linear least squares is used, the estimates are for $k \in s$,

where $\hat{\theta}_k = \mathbf{x}'_k \mathbf{b}$ with

$$\mathbf{b} = (\sum_s d_k \mathbf{x}_k \mathbf{x}'_k)^{-1} (\sum_r d_k \mathbf{x}_k) \, .$$

R-indicators

Statistiska centralbyrån Statistics Sweden

The *R*-indicators use logistic regression fit to obtain

$$\hat{\theta}_{k,\log} = \exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})/[1+\exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})] \text{ for } k \in s.$$

The (unadjusted) *R*-indicator is

$$R = 1 - 2\,S_{\hat{\theta},\log}$$

The adjusted R-indicator is given in for example the [RISQ manual]

Distance between resp. and non-resp.

The distance measure: $dist = [(\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_{s-r})' \mathbf{\Sigma}_s^{-1} (\bar{\mathbf{x}}_r - \bar{\mathbf{x}}_{s-r})]^{1/2}$

where $\bar{\mathbf{x}}_r = \sum_r d_k \mathbf{x}_k / \sum_r d_k$ and $\bar{\mathbf{x}}_{s-r} = \sum_{s-r} d_k \mathbf{x}_k / \sum_{s-r} d_k$ and the weighting matrix is $\mathbf{\Sigma}_s = \sum_s d_k \mathbf{x}_k \mathbf{x}'_k / \sum_s d_k$.

$$BI_1 = 1 - \sqrt{P(1-P)} \times dist \quad , \quad BI_2 = 1 - 2P(1-P) \times dist.$$

Indicators computed on the LCS 2009 data collection

Progression of the response rate P (in per cent), the balance indicators BI_1, BI_2, R unadjusted and R adjusted, and the distance measure *dist*. Computations are based on the standard x-vector.

Step in data collection	$100 \times P$	BI_1	BI_2	R unadj.	R adjusted	*dist*
Attempt 1	12.8	0.855	0.904	0.902	0.905	0.433
Attempt 2	24.6	0.802	0.829	0.829	0.831	0.460
Attempt 3	32.8	0.779	0.793	0.794	0.796	0.470
Attempt 8	53.0	0.751	0.752	0.758	0.760	0.499
End ordinary field work	60.4	0.738	0.744	0.752	0.754	0.536
Final	67.4	0.717	0.735	0.742	0.743	0.603

Lack of balance – special case

The quadratic form $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ has a particularly useful expression when the vector $\mathbf{x}_k$ is defined in terms of J mutually exclusive and exhaustive traits or characteristics.

The trait of unit k is then uniquely coded by the J-vector $\mathbf{x}_k = (\gamma_{1k},...,\gamma_{jk},...,\gamma_{Jk})' = (0,...,1,...,0)'$ (with a single entry “1”).

Lack of balance – special case

For trait j, let $W_{js} = \sum_{s_j} d_k / \sum_s d_k$ be that trait's share of s, and $W_{jr} = \sum_{r_j} d_k / \sum_r d_k$ its share of r.

Then the **lack of balance** (or imbalance) is a sum of non-negative terms expressed as

$$\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D} = \sum_{j=1}^{J} C_j = \sum_{j=1}^{J} \frac{(W_{jr} - W_{js})^2}{W_{js}}$$

Lack of balance – special case

The experimental $\mathbf{x}$-vector is defined by the complete crossing of:

- *Education level* (high, not high),
- *Property ownership* (owner, non-owner),
- *Country of origin* (Sweden, other).

This defines eight mutually exclusive and exhaustive groups.

Values of the eight terms C_j of $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ (multiplied by 100). Experimental x-vector defined by crossing of Education (high, not high), Property ownership (owner, non-owner) and Country of origin (Sweden, other).

Group characteristic			$100 \times C_j$						
			Ordinary fieldwork attempt				Follow-up attempt		
Education	Property ownership	Origin	1	5	12	End	1	4	Final
Not high	Non-owner	Abroad	1.49	1.44	1.26	1.23	1.25	1.16	1.18
Not high	Non-owner	Sweden	0.00	0.06	0.11	0.11	0.08	0.07	0.07
Not high	Owner	Abroad	0.06	0.01	0.00	0.00	0.00	0.00	0.00
Not high	Owner	Sweden	0.72	0.24	0.21	0.19	0.17	0.17	0.18
High	Non-owner	Abroad	1.28	0.39	0.29	0.26	0.25	0.23	0.22
High	Non-owner	Sweden	0.11	0.26	0.25	0.24	0.21	0.20	0.23
High	Owner	Abroad	0.18	0.01	0.03	0.03	0.03	0.02	0.04
High	Owner	Sweden	0.29	0.58	0.64	0.66	0.62	0.53	0.44
		$100 \times \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$	4.13	2.99	2.78	2.72	2.61	2.37	2.36

Laboratory experiment

Experiments in retrospect, each based on an *experimental data collection strategy* consisting of:

- a suitably chosen $\mathbf{x}$-vector, (the *experimental - vector*, with known value for every sampled unit,
- one or more specified *intervention points*, with a *stopping rule* for each intervention point.

Laboratory experiment: Strategy 1

Statistics Sweden

Statistiska centralbyrån

Response rates in per cent at three points in the LCS 2009 data collection

Group characteristic (Experimental x-vector)			Response rate in per cent			Individuals in sample
			After 12 calls	2 follow-up calls	Final	
No high education	No house	Born abroad	37.5	41.8	44.6	847
No high education	No house	Born Sweden	54.6	59.8	64.6	3210
No high education	House	Born abroad	58.5	62.3	66.8	171
No high education	House	Born Sweden	63.0	67.6	73.2	2036
High education	No house	Born abroad	39.4	44.9	48.7	236
High education	No house	Born Sweden	66.8	71.6	77.6	816
High education	House	Born abroad	68.1	73.6	81.9	72
High education	House	Born Sweden	72.2	77.4	81.5	832

Experimental strategy 1; the eight terms C_j of $\mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$ (multiplied by 100) at three points in the data collection.

Group characteristic			Value of $100 \times C_j$ at point		
Education	Property ownership	Origin	Attempt 12 ordinary	Attempt 2 follow-up	Final
Not high	Non-owner	Abroad	1.26	1.06	0.94
Not high	Non-owner	Sweden	0.11	0.03	0.00
Not high	Owner	Abroad	0.00	0.00	0.00
Not high	Owner	Sweden	0.21	0.24	0.08
High	Non-owner	Abroad	0.29	0.21	0.16
High	Non-owner	Sweden	0.25	0.07	0.02
High	Owner	Abroad	0.03	0.01	0.00
High	Owner	Sweden	0.64	0.31	0.17
		$100 \times \mathbf{D}'\boldsymbol{\Sigma}_s^{-1}\mathbf{D}$	2.78	1.93	1.39

Laboratory experiment

Statistiska centralbyrån Statistics Sweden

Response rate (in per cent), balance indicator and distance measure. The computations are based on experimental x-vector.

Experimental Strategy	$100 \times P$	BI_1	*dist*
After 12 calls	57.7	0.805	0.394
2 follow-up calls	61.5	0.824	0.361
Final	63.9	0.843	0.326

Laboratory experiment

Experimental strategy 2:

- Same $\mathbf{x}$-vector,
- 60 % response gives 5 intervention points.

Experimental strategy 3:

- Same $\mathbf{x}$-vector,
- 50% response gives 5 intervention points.

Laboratory experiment

Statistics Sweden

Statistiska centralbyrån

The experimental strategies compared with the actual LCS 2009: Response rate (in per cent), *RDF*, BI_1, *dist* and reduction (in per cent) of the number of call attempts. Computations are based on the ordinary x-vector.

End of data collection	100×P	*RDF* Sickness allowance	*RDF* Income	*RDF* Employed	BI_1	*dist*	*Reduction in %*
Actual 2009 LCS	67.4	-3.6	2.9	3.1	0.717	0.603	0.0
Strategy 1	63.9	-1.6	2.7	3.0	0.765	0.489	8.2
Strategy 2	58.9	-1.2	2.6	3.2	0.787	0.433	20.2
Strategy 3	50.3	1.0	1.0	2.0	0.808	0.383	36.4

Ende…

Statistiska centralbyrån Statistics Sweden