Survey data analysis
Week 10:
“Designing weights”

© Peter Lugtig

Literature today

- Exercise last week
- How to design weights
- Bias-variance trade-off
  - Adjustment error
- Paradata
  - Short lecture
  - How to design and select covariates for weights?
- Class discussion
- Exercise on calibration, raking

Exercise nonresponse weights

- Discussion
  - Is weighting succesful?
  - What happens to bias?
    - MAR or MNAR?
  - What happens to standard errors

Weighting methods

- 1. Propensity score weighting
  - Often using frame information from sample
  - Last week
- 2. post-stratification
- 3. linear weighting (GREG)
- 4. Raking
  - 2-4 often based on population statistics

Nonresponse bias

<u>Deterministic</u>

It is a function of the nonresponse rate *M/N* and the difference between the respondents' *r* and the nonrespondents' *m* population values.

$$B(\bar{y}_r) = \left(\frac{M}{N}\right)\left(\bar{Y}_r - \bar{Y}_m\right)$$

<u>Probabilistic</u>

It is a function of the correlation σ of the survey outcome y with the response propensity ρ and the mean response propensity measured in the target population (Bethlehem 2002).

$$B(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}}$$

Note that nonresponse bias is always estimate-specific!

Propensity weighting

Deterministic

It is a function of the nonresponse rate *M/N* and the difference between the respondents' *r* and the nonrespondents' *m* population values.

$$B(\bar{y}_r) = \left(\frac{M}{N}\right)\left(\bar{Y}_r - \bar{Y}_m\right)$$

Probabilistic

It is a function of the correlation σ of the survey outcome y with the response propensity ρ and the mean response propensity measured in the target population (Bethlehem 2002).

$$B(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}}$$

Propensity-score weights

For propensity-score weights (logistic regression) models estimate the response propensity (predicted probability) of each sample unit given a set of covariates.

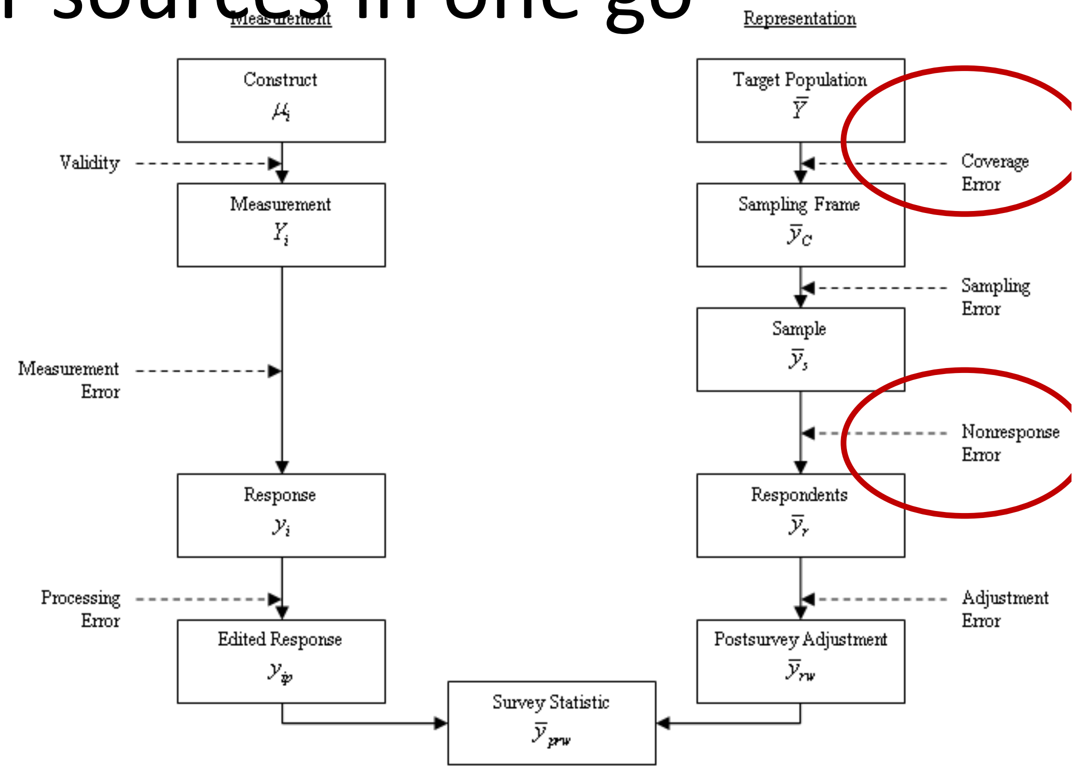
- Response rate for all linear combinations of for example:
  - response[0;1] ~ gender+age+region+typehouse

Weight is the scaled inverse of the predicted response propensity of each sample unit.

- Design weight = sample inclusion probability
- Propensity weight = participation probability

Post-stratification

- Deterministic approach
- Uses population statistics
- Correct multiple error sources in one go
  - Nonresponse
  - Coverage



Post-stratification: fictitious example

Survey distribution

Age	Gender	
	Male	Female
16-25	6.1%	5.9%
26-35	12.8%	10.7%
36-45	19.5%	16.9%
46-55	10.0%	9.5%
56-65	3.9%	4.7%

Population distribution

Age	Gender	
	Male	Female
16-25	10.0%	10.0%
26-35	10.0%	10.0%
36-45	10.0%	10.0%
46-55	10.0%	10.0%
56-65	10.0%	10.0%

Male respondents aged 16-25 receive a weight of
w = 0.1/0.061 = 1.639

Post-stratification: fictitious example

Survey distribution

Age	Gender	
	Male	Female
16-25	6.1%	5.9%
26-35	12.8%	10.7%
36-45	19.5%	16.9%
46-55	10.0%	9.5%
56-65	3.9%	4.7%

Population distribution

Age	Gender	
	Male	Female
16-25	10.0%	10.0%
26-35	10.0%	10.0%
36-45	10.0%	10.0%
46-55	10.0%	10.0%
56-65	10.0%	10.0%

Female respondents aged 36-45 receive a weight of
w = 0.1/0.169 = 0.592

Post-stratification: fictitious example

Survey distribution

Age	Gender: Male	Gender: Female
16-25	6.1%	5.9%
26-35	12.8%	10.7%
36-45	19.5%	16.9%
46-55	10.0%	9.5%
56-65	3.9%	4.7%

Population distribution

Age	Gender: Male	Gender: Female
16-25	10.0%	10.0%
26-35	10.0%	10.0%
36-45	10.0%	10.0%
46-55	10.0%	10.0%
56-65	10.0%	10.0%

In R (using survey library):

```
PostStratify(design= svy.unweighted, strata = agegender, population=agegender.dist
```

When does weighting work?

1. Target variables vary little within strata
2. Response probabilities vary little within strata
3. Target variable and response probabilities are not correlated within strata (=MAR)

Example Unit Nonresponse:

- Income varies little within gender and age (male/female)
- Given gender and age, there is little variation in response rates
- Given gender and age, there is no correlation between probability of response and income

When does weighting work?

1. Target variables vary little within strata
2. Response probabilities vary little within strata
3. Target variable and response probabilities are not correlated within strata (=MAR)

Example Unit Nonresponse:

- Income varies little within gender and age (male/female)
- Given gender and age, there is little variation in response rates
- Given gender and age, there is no correlation between probability of response and income

Unlikely

How to weight better…

- Use more variables in weighting
  - Gender, age, …
- To improve relation X-> R and X->Y

- Problems:
- 1. There may be empty strata
  - Combine empty strata
  - Use fewer variables
- 2. The population distribution across all variables may not be available

Raking
aka multiplicative weighting

Raked weights adjust the individual distributions of key survey variables to known joint population distributions.

Raking is an iterative process.

Raking: fictitious example

Survey distribution

Gender	
Male	52.3%
Female	47.7%

Age	
16-25	12.0%
26-35	23.5%
36-45	36.4%
46-55	19.5%
56-65	8.6%

Population distribution

Gender	
Male	50.0%
Female	50.0%

Age	
16-25	20.0%
26-35	20.0%
36-45	20.0%
46-55	20.0%
56-65	20.0%

Raking: fictitious example

Survey distribution

Gender	
Male	52.3%
Female	47.7%

Age	
16-25	12.0%
26-35	23.5%
36-45	36.4%
46-55	19.5%
56-65	8.6%

Population distribution

Gender	
Male	50.0%
Female	50.0%

Age	
16-25	20.0%
26-35	20.0%
36-45	20.0%
46-55	20.0%
56-65	20.0%

The survey distributions are iteratively adjusted to the age and gender population distributions until an equilibrium is reached.

Raking: fictitious example

Survey distribution

Gender	
Male	52.3%
Female	47.7%

Age	
16-25	12.0%
26-35	23.5%
36-45	36.4%
46-55	19.5%
56-65	8.6%

Population distribution

Gender	
Male	50.0%
Female	50.0%

Age	
16-25	20.0%
26-35	20.0%
36-45	20.0%
46-55	20.0%
56-65	20.0%

Raking in R with survey library:

```
rake(design = svy.unweighted,
sample.margins = list(~sex, ~age), population.margins = list(sex.dist, age.dist))
```

Alternative: Linear weighting

- Aka *Generalized Regression Estimator (GREG) or calibration*
- Fixes problem of empty population cells
- Fixes problem of population distribution

	Male	Female	Total
Young	23	15	38
Middle	16	17	33
Old	13	16	39
Total	52	48	100

Sample

	Male	Female	Total
Young	?	?	43
Middle	?	?	30
Old	?	?	27
Total	51	49	100%

Population

With thanks to Jelke Bethlehem for this example

Linear weighting

Assume there are p continuous auxiliary variables

- Similar to propensity score weighting (other link-function)
- Required: vector of population means
- Estimate: conditional RR (most likely) for combinations of categories of p
- Best value for B (least squares): $B = \left(\sum_{k=1}^{N} X_k X_k'\right)^{-1} \left(\sum_{k=1}^{N} X_k Y_k\right)$
- Sample-based estimate (full response): $b = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i\right)$

With thanks to Jelke Bethlehem for this example

Linear weighting

Gender	Age	X1	X2	X3	X4	X5	X6
Male	Young	1	1	0	1	0	0
Male	Middle	1	1	0	0	1	0
Male	Old	1	1	0	0	0	1
Female	Young	1	0	1	1	0	0
Female	Middle	1	0	1	0	1	0
Female	Old	1	0	1	0	0	1
Sample prop		-	.52	.48	.38	.33	.39
Population prop.		1.00	.51	.49	.43	.30	.27
Weights		1.00					

- Weight for X2 (as in raking):
  - Naive - .51/.52 = .98 (But correlated with x4-x6

Linear weighting

Gender	Age	X1	X2	X3	X4	X5	X6
Male	Young	1	1	0	1	0	0
Male	Middle	1	1	0	0	1	0
Male	Old	1	1	0	0	0	1
Female	Young	1	0	1	1	0	0
Female	Middle	1	0	1	0	1	0
Female	Old	1	0	1	0	0	1
Sample prop		-	.52	.48	.38	.33	.39
Population prop.		1.00	.51	.49	.43	.30	.27
Weights		1.00					

- Weight for X2 (as in raking):
  - Minimize total distance across variables
  - .503/ .52 = .967. Weight = -.033

Linear weighting

Linear weighting computations

Gender	Age	X1	X2	X3	X4	X5	X6
Male	Young	1	1	0	1	0	0
Male	Middle	1	1	0	0	1	0
Male	Old	1	1	0	0	0	1
Female	Young	1	0	1	1	0	0
Female	Middle	1	0	1	0	1	0
Female	Old	1	0	1	0	0	1
Sample prop		-	.52	.48	.38	.33	.39
Population prop.		1.00	.51	.49	.43	.30	.27
Weights		1.00	-.033	.033	.161	-.095	-.066

- Weight for young female = 1 +.033+.161=1.185

Linear weighting

Linear weighting computations

Gender	Age	X1	X2	X3	X4	X5	X6
Male	Young	1	1	0	1	0	0
Male	Middle	1	1	0	0	1	0
Male	Old	1	1	0	0	0	1
Female	Young	1	0	1	1	0	0
Female	Middle	1	0	1	0	1	0
Female	Old	1	0	1	0	0	1
Sample prop		-	.52	.48	.38	.33	.39
Population prop.		1.00	.51	.49	.43	.30	.27
Weights		1.00	-.033	.033	.161	-.095	-.066

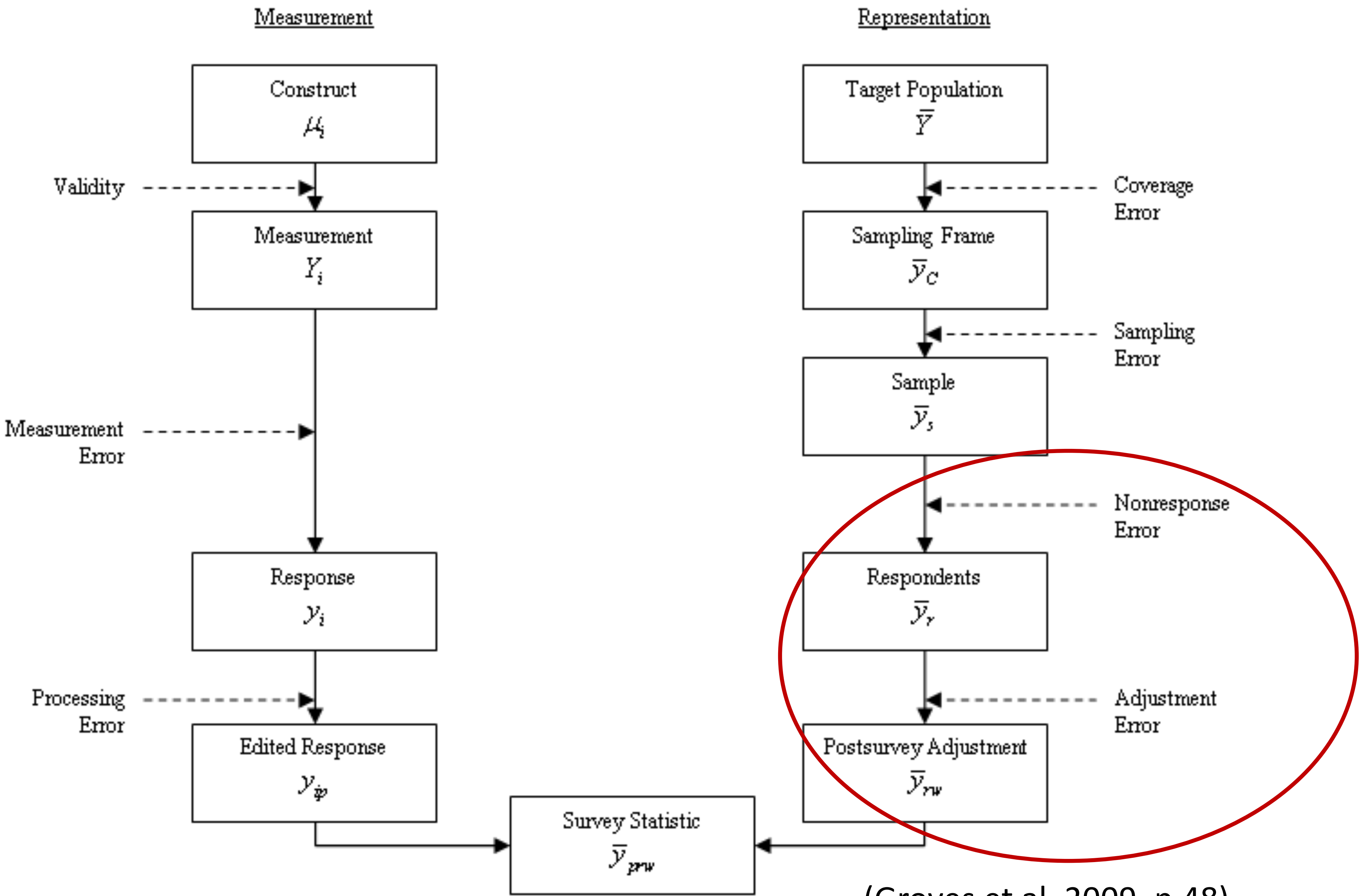
In R using survey library:

Calibrate(design=svy.unweighted, formula = ~age+gender , population=c(sex.dist,age.dist)

GREG or Raking?

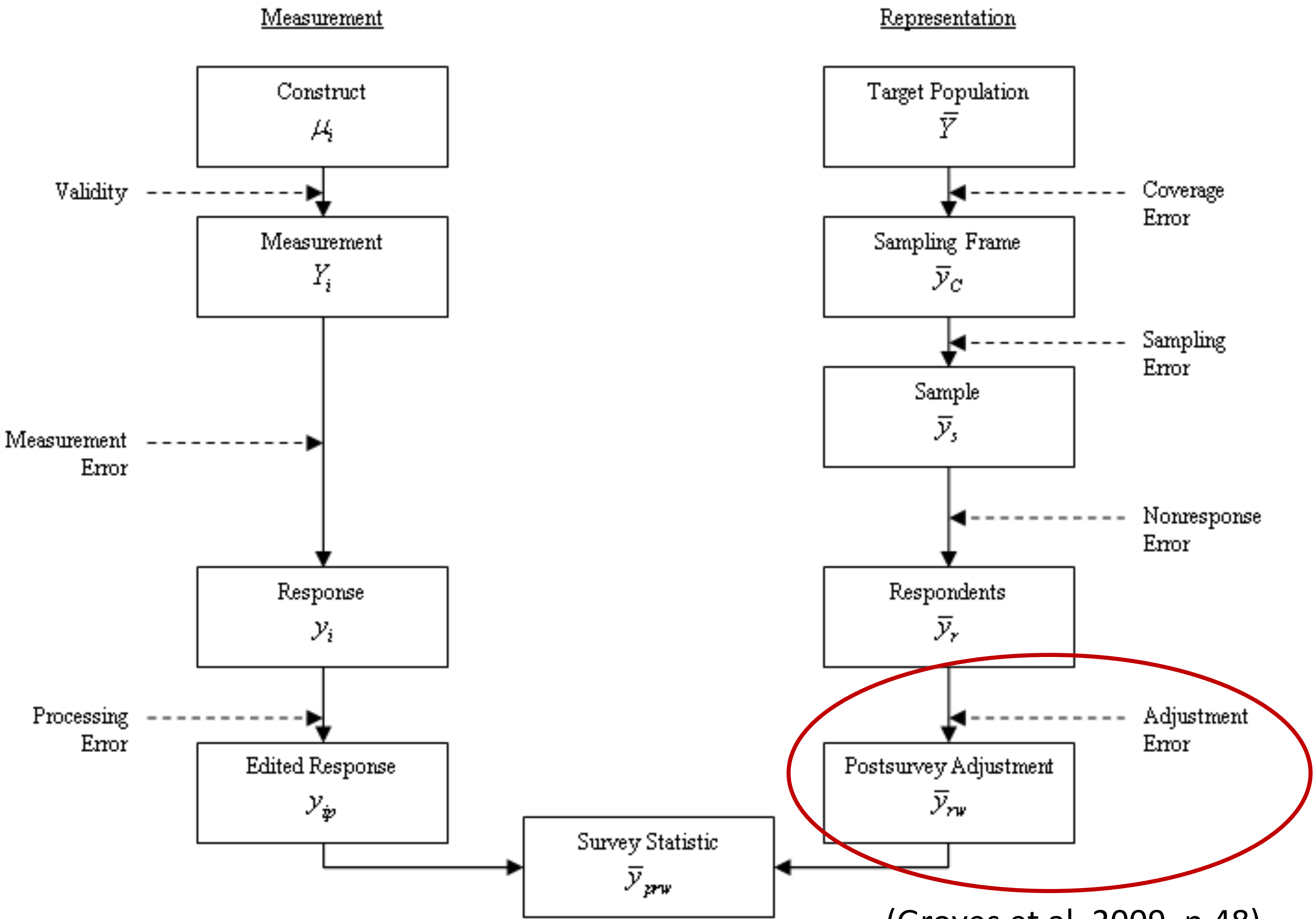
- GREG
  - Clear model
  - Variance of estimators can be computed
  - Can include continuous weighting variables (e.g. age)
  - Assumptions of regression model (normality, linearity)
  - Weights may become negative (causing computational problems)
- Raking
  - No model, no assumptions
  - No variance estimation
  - Only categorical variables

Total Survey Error (TSE) Framework



(Groves et al. 2009, p.48)

Total Survey Error (TSE) Framework



(Groves et al. 2009, p.48)

Bias-variance trade-off

Bias

- Sample mean $\overline{y} = \frac{1}{n}\sum^{n} y_i$
- Weighted mean $\overline{y_w} = \frac{1}{n}\sum^{n} w_i y_i$

Variance

- If there is no correlation between survey weights and the characteristic to be estimated, maximum increase in variance of the mean is (Kish, 1965):

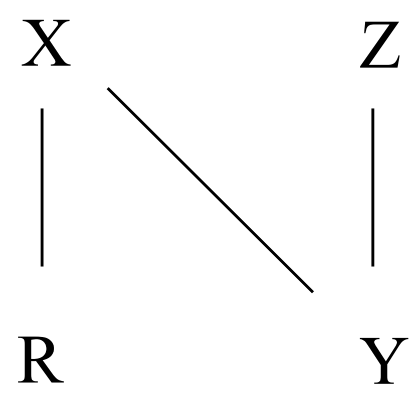
$$UWE = 1 + cv^2(w_i)$$

$$cv = standard\ deviation(w_i)/mean(w_i)$$

Succesful NR weighting

- When (NR) weighting is succesful
  - Little and Vartivarian (2005)

	X -> Y weak	X -> Y strong
X-> R weak	Nothing happens	Still bias Variance deflation
X -> R strong	Still bias Variance inflation	Bias reduction Variance reduction



MAR

Bias-variance trade-off(2)

- To avoid the inflation of variance
  - Outlier trimming
  - Trimming weights >3 to 3
  - Propensity score weighting
    - Use 5 bands of propensity scores

Designing weights

- Sample level information
  - Location, gender, age, ….
  - ?
- Population level information
  - LOTS of potential variables
  - Gender, age, education, ethnicity, region, income
  - Membership of union, newspaper readership, politically active, visited the Efteling, etc,

Designing weights

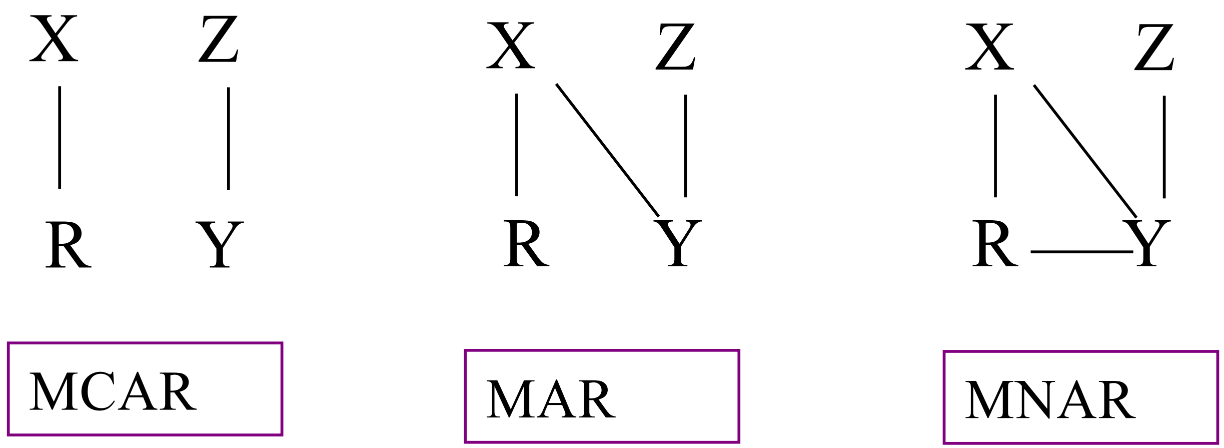
- Sample level information
  - Location, gender, age, ...
  - ? ---- paradata
- Population level information
  - LOTS of potential variables
  - Gender, age, education, ethnicity, region, income
  - Membership of union, newspaper readership, politically active, visited the Efteling, etc,

Paradata

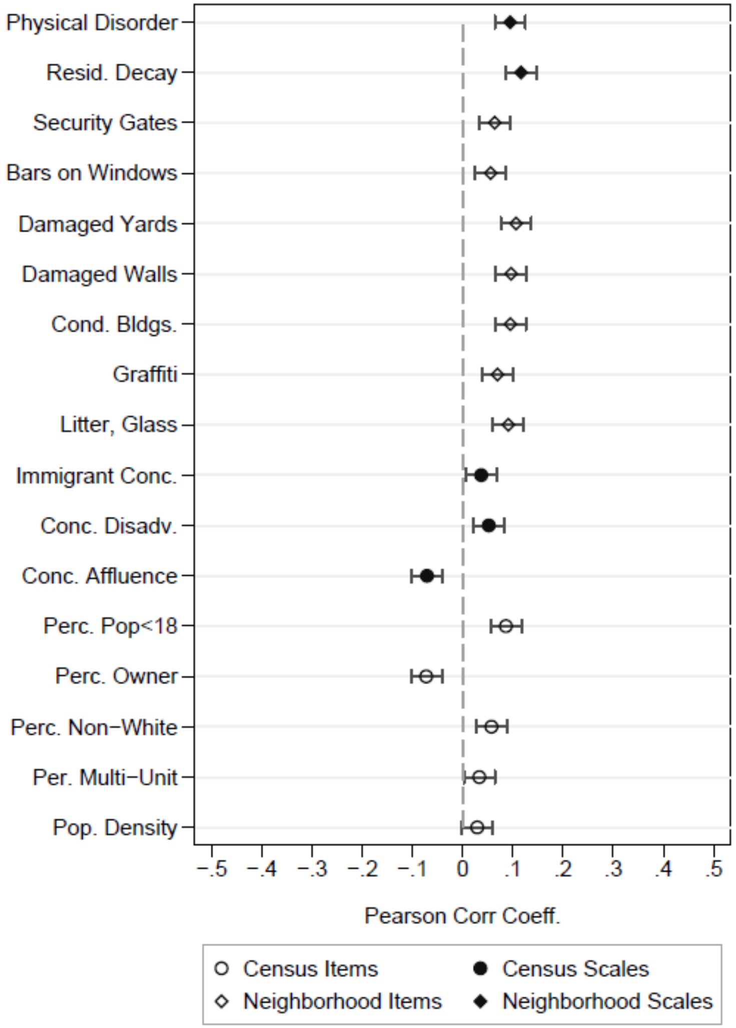
- By-product of doing research
- Surveys
  - F2f: Interviewer observations
    - Or Recordings
  - CATI: Call record data
  - WEB: Browser data, Response timings, section timings, evaluation questions, etc.

Paradata – so what?

- Interviewer observations
  - Useful in nonresponse corrections
    - Strong link X -> Y



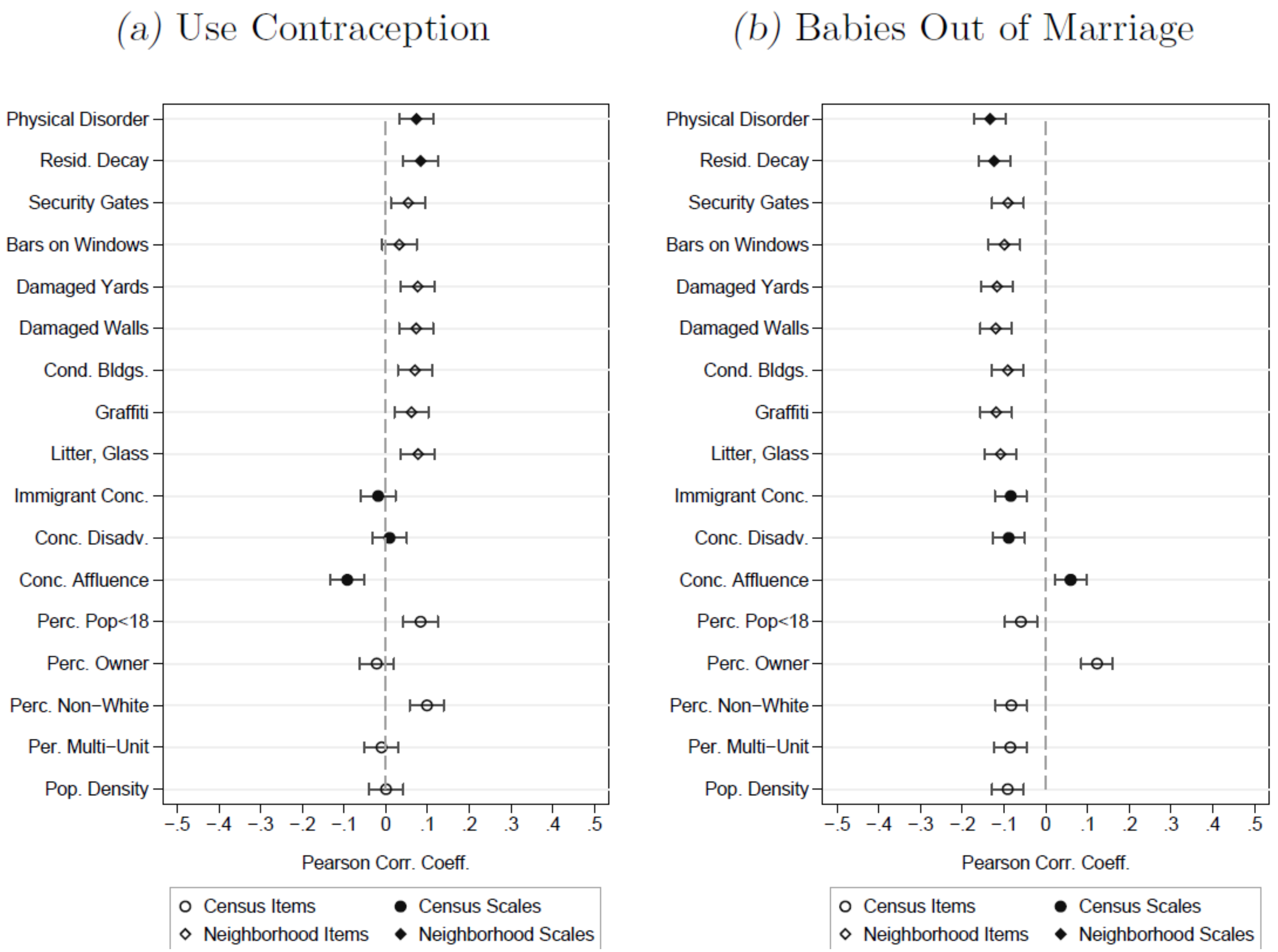
Interviewer observations example Casas-Cordero (2010)



Relations X -> R.
Bivariate (logistic) relations rather weak,
… but strong link with Y?

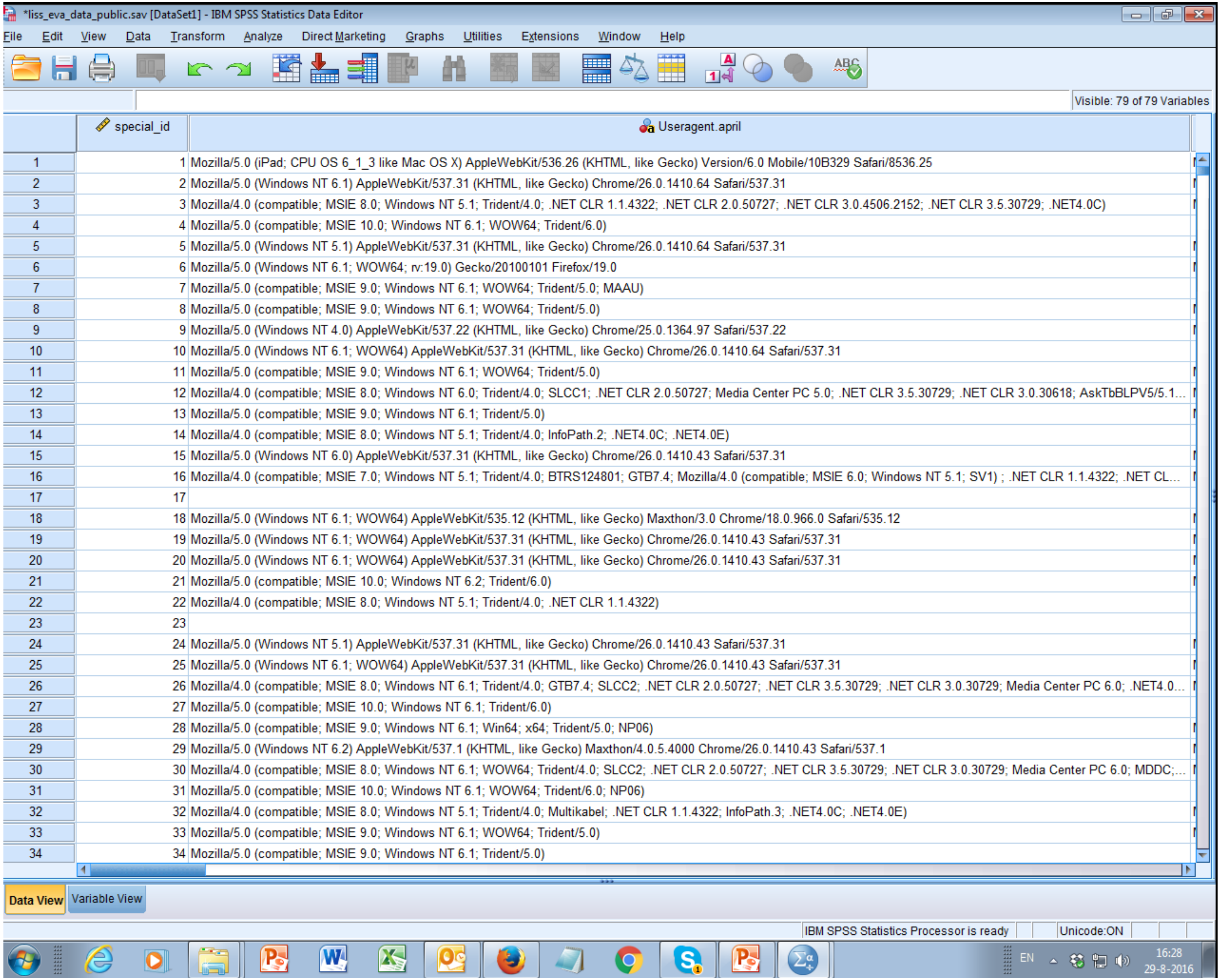
- Figure taken from Casas-Cordero (2010)

Interviewer observations (2)



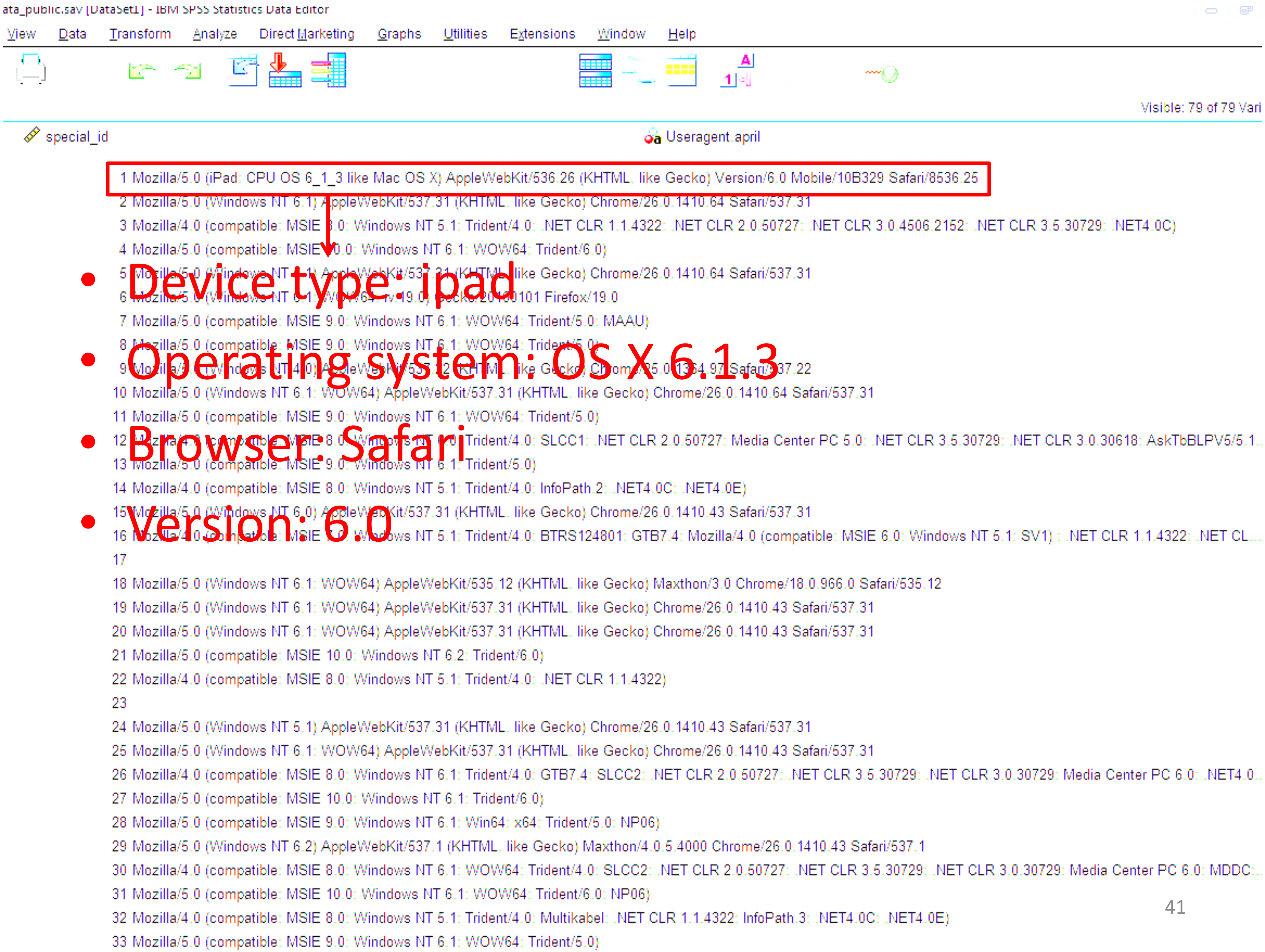
- Relation X-> Y. Figure taken from Casas-Cordero (2010)

Paradata – web



	special_id	Useragent.april
1	1	Mozilla/5.0 (iPad; CPU OS 6_1_3 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10B329 Safari/8536.25
2	2	Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.64 Safari/537.31
3	3	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET CLR 3.0.4506.2152; .NET CLR 3.5.30729; .NET4.0C)
4	4	Mozilla/5.0 (compatible; MSIE 10.0; Windows NT 6.1; WOW64; Trident/6.0)
5	5	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.64 Safari/537.31
6	6	Mozilla/5.0 (Windows NT 6.1; WOW64; rv:19.0) Gecko/20100101 Firefox/19.0
7	7	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; MAAU)
8	8	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)
9	9	Mozilla/5.0 (Windows NT 4.0) AppleWebKit/537.22 (KHTML, like Gecko) Chrome/25.0.1364.97 Safari/537.22
10	10	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.64 Safari/537.31
11	11	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)
12	12	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .NET CLR 3.5.30729; .NET CLR 3.0.30618; AskTbBLPV5/5.1...
13	13	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)
14	14	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; InfoPath.2; .NET4.0C; .NET4.0E)
15	15	Mozilla/5.0 (Windows NT 6.0) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.43 Safari/537.31
16	16	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.0; BTRS124801; GTB7.4; Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1) ; .NET CLR 1.1.4322; .NET CL...
17	17	
18	18	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/535.12 (KHTML, like Gecko) Maxthon/3.0 Chrome/18.0.966.0 Safari/535.12
19	19	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.43 Safari/537.31
20	20	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.43 Safari/537.31
21	21	Mozilla/5.0 (compatible; MSIE 10.0; Windows NT 6.2; Trident/6.0)
22	22	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; .NET CLR 1.1.4322)
23	23	
24	24	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.43 Safari/537.31
25	25	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.31 (KHTML, like Gecko) Chrome/26.0.1410.43 Safari/537.31
26	26	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; Trident/4.0; GTB7.4; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0...
27	27	Mozilla/5.0 (compatible; MSIE 10.0; Windows NT 6.1; Trident/6.0)
28	28	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0; NP06)
29	29	Mozilla/5.0 (Windows NT 6.2) AppleWebKit/537.1 (KHTML, like Gecko) Maxthon/4.0.5.4000 Chrome/26.0.1410.43 Safari/537.1
30	30	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; MDDC;...
31	31	Mozilla/5.0 (compatible; MSIE 10.0; Windows NT 6.1; WOW64; Trident/6.0; NP06)
32	32	Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; Multikabel; .NET CLR 1.1.4322; InfoPath.3; .NET4.0C; .NET4.0E)
33	33	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)
34	34	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)

User Agent strings:

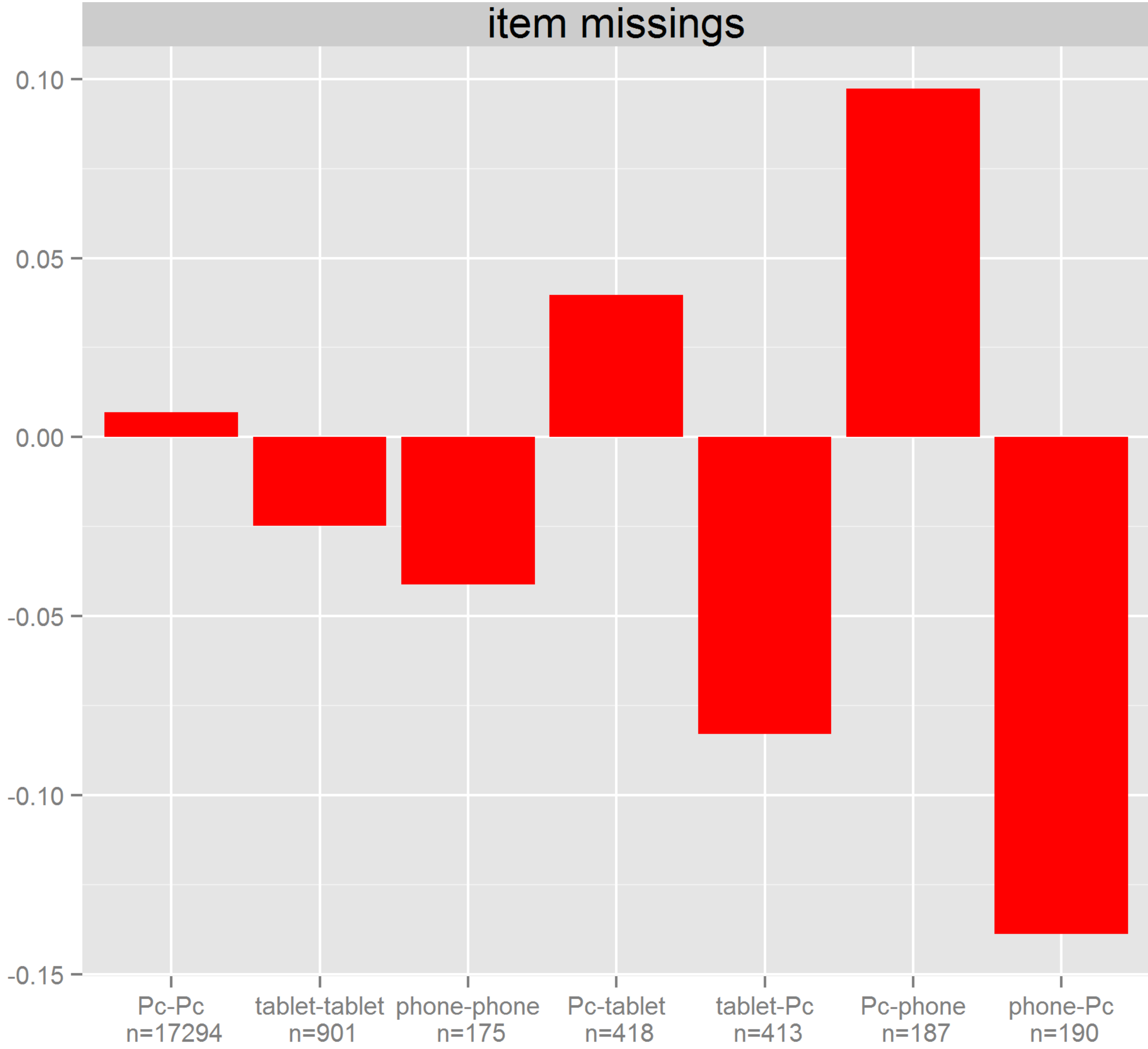


- Device type: ipad
- Operating system: OS X 6.1.3
- Browser: Safari
- Version: 6.0

User agent strings
Example by Lugtig and Toepoel (2015)

% Device used in month (t)		% Device used in May					N	% of wave respondents
		PC	Tablet	Phone	No participation			
April	PC	77.4	1.1	0.4	21.1		4966	90.2
	Tablet	19.3	24.4	0.2	56.1		435	7.8
	Phone	33.9	4.6	30.3	31.2		109	2.0
	No participation	37.6	2.1	0.8	59.4		715	-

User agent strings
Example by Lugtig and Toepoel (2)



Paradata +++

- Taken from Toepoel and Lugtig (2014)
  - more in week "designed big data"

How to use paradata

- Scheduling what to do next -> reduce costs
  - Call scheduling, interviewer visits
- Planning next wave of data collection
  - Longitudinal, or repeated surveys
- For corrections!
  - Use interviewer observations to weight (or impute)
  - Use type of nonresponse and make separate models.
    - Noncontact <- age + #hh members + work
    - Refusal <- level of education + gender

Designing weights in R

1. Design your study such that
- You get a rich frame (paradata)
- You ask about known population statistics that predict Y

2. Find variables (x) that predict both R and Y
- propensity score weighting
- Poststratify
- calibrate (linear weighting)
- Rake

Combining weights

Designing weights can have several stages. For example:

1. The design weight is derived from the sampling frame (and the household selection during fieldwork).

2. Propensity-score weights are calculated using information from the sampling frame and the response indicator.

3. The data are post-stratified to known population distributions.

Weights can be combined : W1 * W2 * W3 = Wt

- Often only the total Weight included in public datasets
  - No detailed information on sampling design or nonresponse correlates
- Or svydesign for sampling plan, use weights for nonresponse and coverage

Class exercise

- Website of Statistics Netherlands: statline
  - https://opendata.cbs.nl/statline/#/CBS/en/
- Find auxiliary variables for your scenario
  - Should predict your Y
  - … and nonresponse
- 15 minutes in groups of 4

How to use variables in R…

- ‘survey’ library
- Imagine: auxiliary data for sex, age
  - PostStratify(design= svy.unweighted, strata = agegender, population=agegender.dist
  - rake(design = svy.unweighted, sample.margins = list(~sex, ~age), population.margins = list(sex.dist, age.dist))
  - Calibrate(design=svy.unweighted, formula = ~age+gender , population=c(sex.dist,age.dist)

Next week

- Finish exercise on creating weights
- Two weeks on imputation (by Stef van Buuren)
  - readings

Now: work on exercise

- Exercise on creating your own weights
  - Poststratification
  - Propensity weighting
  - Raking

- All with the survey package