RRTCS package. Application of Horvitz model to a real survey

Beatriz Cobo Rodríguez, María del Mar Rueda García, Antonio Arcos Cebrián

2021-04-21

##Horvitz model In the Horvitz model, the randomized response device presents to the sampled person labelled i a box containing a large number of identical cards, with a proportion $p, (0 < p < 1)$ bearing the mark A and the rest marked B (an innocuous attribute whose population proportion α is known).
The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A or if i bears B and the card drawn is marked B. Otherwise z_i takes the value 0.

$$z_i = \begin{cases} y_i \rightarrow \begin{cases} \text{if the card is marked } A \text{ and you really have to } A \\ \text{if the card is marked } B \text{ and you really have to } B \end{cases} \\ 0 \rightarrow \text{ if the mark of the card does not match with your real characteristic} \end{cases}$$

##Horvitz model example The file data "HorvitzDataRealSurvey.rda" stores a sample, extracted by simple random sampling without replacement, of size 710. With these data we estimate the prevalence of students who have sensitive characteristics set out below.
The randomized response technique used is Horvitz model (Horvitz et al, 1967 and Greenberg et al., 1969) with parameter $p = 0.5$

###Horvitz model survey design In this model an additional questionnaire with the following characteristics is required:

- The questions should not be sensitive.
- The questions should not be related to the theme of "sensitive" questions.
- For each of the questions that measure qualitative variables, the answers should have a known probability.

The survey is

To estimate results, you must have the response probabilities of non-sensitive questions:

###Horvitz model technique description In order to give more confidence to the respondent, he was given a booklet of instructions:

- Step 1: Flip a coin and did not show anyone the side obtained.
- Step 2: The question you must answer in each game depends on the side obtained on the coin:
  - If you get head, answer the question preceded by the word HEAD
  - If you get tail, answer the question preceded by the word TAIL

This procedure must be repeated for each question.

The respondent only had to put on the question sheet a cross in the chosen answers, and then he deposited the survey sheet in an urn.

We store the answers to senstive questions in data.frame (HorvitzDataRealSurvey.rda).

###Horvitz function We use the Horvitz function to obtain the estimates:

HEAD	Have you ever copied in an exam?	YES	NO
TAIL	Were you born in July?		

HEAD	Have you ever fought with a teacher?	YES	NO
TAIL	Does your ID number end in 2?		

HEAD	Have you been bullied?	YES	NO
TAIL	Were you born of 1 to 20 of the month?		

HEAD	Have you ever bullied someone?	YES	NO
TAIL	Does your ID number end in 5?		

HEAD	Have you ever taken drugs on the campus?	YES	NO
TAIL	Were you born of 15 to 25 of the month?		

HEAD	Have you had sex on the premises of the university?	YES	NO
TAIL	Were you born in April?		

Figure 1: Sin titulo

#	**Question**	**Response probability**
1	Were you born in July?	1/12
2	Does your ID number end in 2?	1/10
3	Were you born of 1 to 20 of the month?	20/30
4	Does your ID number end in 5?	1/10
5	Were you born of 15 to 25 of the month?	10/30
6	Were you born in April?	1/12

Figure 2: Sin titulo

```r
library("RRTCS")
N=10777
n=710
data(HorvitzDataRealSurvey)
p=0.5
alpha=c(1/12,1/10,20/30,1/10,10/30,1/12)
pi=rep(n/N,n)
cl=0.95
out1=Horvitz(HorvitzDataRealSurvey$copied,p,alpha[1],pi,"mean",cl,N)
out1
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$copied, p = p, alpha = alpha[1],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.083
##
## Estimation: 0.8406103
## Variance: 0.001389716
## Confidence interval (95%)
##     Lower bound: 0.767545
##     Upper bound: 0.9136756
```

```r
out2=Horvitz(HorvitzDataRealSurvey$fought,p,alpha[2],pi,"mean",cl,N)
out2
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$fought, p = p, alpha = alpha[2],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.1
##
## Estimation: 0.4070423
## Variance: 0.001045196
## Confidence interval (95%)
##     Lower bound: 0.3436776
##     Upper bound: 0.4704069
```

```r
out3=Horvitz(HorvitzDataRealSurvey$bullied,p,alpha[3],pi,"mean",cl,N)
out3
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$bullied, p = p, alpha = alpha[3],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.67
##
```

```
## Estimation: 0.1220657
## Variance: 0.001337415
## Confidence interval (95%)
##     Lower bound: 0.05038851
##     Upper bound: 0.1937429
```

```
out4=Horvitz(HorvitzDataRealSurvey$bullying,p,alpha[4],pi,"mean",cl,N)
out4
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$bullying, p = p, alpha = alpha[4],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.1
##
## Estimation: 0.128169
## Variance: 0.0005597858
## Confidence interval (95%)
##     Lower bound: 0.08179667
##     Upper bound: 0.1745414
```

```
out5=Horvitz(HorvitzDataRealSurvey$drug,p,alpha[5],pi,"mean",cl,N)
out5
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$drug, p = p, alpha = alpha[5],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.33
##
## Estimation: 0.1286385
## Variance: 0.000991658
## Confidence interval (95%)
##     Lower bound: 0.06691805
##     Upper bound: 0.1903589
```

```
out6=Horvitz(HorvitzDataRealSurvey$sex,p,alpha[6],pi,"mean",cl,N)
out6
```

```
##
## Call:
## Horvitz(z = HorvitzDataRealSurvey$sex, p = p, alpha = alpha[6],
##     pi = pi, type = "mean", cl = cl, N = N)
##
## Qualitative model
## Horvitz model for the mean estimator
## Parameters: p=0.5; alpha=0.083
##
## Estimation: 0.06596244
## Variance: 0.000383954
## Confidence interval (95%)
```

```
##     Lower bound: 0.02755745
##     Upper bound: 0.1043674
```

##References Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.