A Log-Linear Randomized-Response Model to Account for Cheating

M.J.L.F. Cruyff[1], A. van den Hout[2], P.G.M. van der Heijden[1] and U. Böckenholt[3]

[1] Utrecht University, The Netherlands, m.cruijff@fss.uu.nl
[2] MRC, Cambridge, United Kingdom, ardo.vandenhout@mrc-bsu.cam.ac.uk
[3] McGill Unversity, Montreal, Canada, ulf.bockenholt@mcgill.ca

Abstract: Randomized response (RR) is an interview technique designed to eliminate response bias when sensitive questions are asked. In RR the answer depends to a certain degree on the outcome of a randomizing device. Although RR elicits more honest answers than direct questions, respondents do not always follow the instructions and in the sense that they answer *no* regardless of the outcome of the randomizing device. In this paper we present a log-linear randomized-response model that accounts for this kind of cheating. The main results of this model are (1) an estimate of the probability of cheating; (2) log-linear parameters estimates describing the associations between RR variables and; (3) prevalence estimates of the sensitive behavior that are corrected for cheating. We illustrate the model with an example.

Keywords: randomized response; log-linear model; cheating parameter.

1 Introduction

Most people are reluctant to publicly answer questions about sensitive topics, like drug or alcohol (ab)use, sexuality or anti-social behavior. As a result, respondents may refuse to give the embarrassing answer and the stigmatizing behavior is often underreported. Randomized Response (RR) is an interview technique that is especially developed to eliminate this kind of evasive response bias (Warner, 1965). In RR the answer to the sensitive question is to a certain extent determined by a randomizing device, like a pair of dice or the draw of card. Since the outcome of the device is known only to the respondent, confidentiality is guaranteed. A meta-analysis shows that RR yields more valid prevalence estimates than direct-questioning designs (Lensvelt-Mulders, Hox, Van der Heijden and Maas, 2005).
Despite the protection of the respondents' privacy, RR does not completely eliminate the response bias. Several studies have shown that the RR design is susceptible to cheating, in the sense that the respondent does not answer in accordance with the outcome of the randomizing device. For example, in an experimental setting, Edgell, Himmelfarb and Duchan (1982) used an

RR design of which the outcomes of the randomization procedure were fixed in advance. It appeared that to a question about having experiences with homosexuality, 25% of the respondents who had to answer *yes* by design cheated. In another study, Van der Heijden, Van Gils, Bouts and Hox (2000) applied different interview techniques to subjects who were identified as having committed social welfare fraud. Although the RR condition elicited more admittances of fraud than direct-questioning or computer-assisted self-interviews, a substantial proportion of the subjects still denied having committed fraud. Finally, after completing a computer-assisted RR survey, most respondents stated they had found it hard to give a false *yes* response and some admitted having cheated (Boeije and Lensvelt-Mulders, 2002).
Böckenholt and van der Heijden (2004) propose an item randomized-response model to correct for cheating. In this paper we present a log-linear modeling approach to account for cheating. Chen (1989) and Van den Hout and Van der Heijden (2004) have presented log-linear randomized-response (LLRR) models to study the associations between RR variables. In the present paper a log-linear randomized cheating (LLRRC) model is specified by the introduction of an extra parameter into the LLRR model to account for cheating. The three main results of LLRRC model are: (1) an estimate of the probability of cheating; (2) log-linear parameter estimates describing the associations between RR variables and; (3) prevalence estimates of the sensitive behavior that are corrected for cheating. The model is illustrated with an example from the Social Welfare Survey conducted in 2000 in the Netherlands (Van Gils, Van der Heijden, Bouts and Hox, 2000, and Lensvelt-Mulders, Van der Heijden, Laudy and Gils, 2006).
The outline of this paper is as follows. In section 2 we present the questions and the RR design used in the Social Welfare Survey. Section 3 introduces the general RR model and shows that identification problems arise when a cheating parameter is introduced. Then the log-linear randomized-response (LLRR) model is introduced as a reparametrization of the general RR model, followed by the log-linear randomized-response cheating (LLRRC) model. This model is not overparametrized because the cheating parameter is incorporated into the model at the expense of the highest-order interaction parameter. We then present the results of the example from the Social Welfare Survey. The example is followed by an investigation of the robustness of the parameter estimates against violations of model assumptions.

2 Survey Data

In the Netherlands employees are insured under various social welfare acts against the loss of income due to redundancy, invalidity or sickness. Recipients of financial benefits have to comply with the rules and regulations that are stipulated in these acts. Violation of the rules is considered fraud and can have serious repercussions for the offender. In 2000 the Dutch De-

partment of Social Affaires conducted a nationwide survey to monitor the degree of noncompliance with respect to these rules.
From the 2000 Social Welfare Survey we present an examples of the LLRRC model. The example concerns the following four questions that were asked to 1,308 beneficiaries of the Disability Benefit Act (DBA):

1 Has a doctor or specialist ever told you that the symptoms your disability classification is based upon have decreased without you informing the Department of Social Services of this change?

2 At a Social Services check-up, have you ever acted as if you were sicker or less able to work than you actually are?

3 Have you yourself ever noticed an improvement in the symptoms causing your disability, for example in your present job, in volunteer work or the chores you do at home, without informing the Department of Social Services of this change?

4 For periods of any length at all, do you ever feel stronger and healthier and able to work more hours without informing the Department of Social Services of this change?

The answers to these questions are denoted by the variables D, E, F and G. The questions were all answered according to the Kuk design (Kuk, 1990 and, Van der Heijden, Van Gils, Bouts and Hox, 2000). In this particular RR design the respondent is given two decks with red and black playing cards. One deck contains 80% red cards and 20% black cards and is called the *yes* deck. The other deck contains 80% black cards and 20% red cards and is called the *no* deck. Each time a sensitive question is asked, the respondent draws one card from both decks and answers the question by naming the color of the card from the deck corresponding to the true answer. So, if the true answer is *yes*, the respondent names the color of the card from the *yes* deck, and if the true answer is *no*, the respondent names the color of the card from the *no* deck. The vector with the observed-response frequencies $yyyy, yyyn, \ldots, nnnn$ is given by $\boldsymbol{n}^* = (43, 22, 10, 34, 20, 31, 40, 93, 30, 29, 40, 91, 60, 86, 146, 533)^t$.

3 The Log-Linear Randomized-Response Cheating Model

3.1 A general Randomized-Response Model

Consider a multivariate RR design with K dichotomous sensitive questions $1, 2, \ldots, K$. Let the true responses be denoted by the random variables $A, B, \ldots$, with realizations $a, b, \ldots \in \{1 \equiv yes, 2 \equiv no\}$, and let the random variable X denote the true-responses profiles $A = a, B = b, \ldots$. Analogously, let X^* denote the observed-response profiles. A general RR model is given by

$$\boldsymbol{\pi}^* = \boldsymbol{P}_K \boldsymbol{\pi}, \tag{1}$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_D)^t$ denotes the true-response profile probabilities. The multivariate matrix $\boldsymbol{P}_K$, with elements $p_{ij} = I\!P(X^* = i|X = j)$, $i, j \in \{1, \ldots, D\}$ denoting the conditional misclassification probabilities, is a transition matrix given by the Kronecker products $\boldsymbol{P}_1 \otimes \boldsymbol{P}_1 \otimes \ldots \otimes \boldsymbol{P}_1$ of the univariate transition matrix $\boldsymbol{P}_1$. For the Kuk design

$$\boldsymbol{P}_1 = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 8/10 & 2/10 \\ 2/10 & 8/10 \end{pmatrix}. \tag{2}$$

3.2 Boundary solutions

The general RR model sometimes exhibits a lack of fit. We refer to such a result as a boundary solution, because in the general RR model a lack of fit is characterized by one or more parameter estimates on the boundary of the parameter space (Van den Hout and Van der Heijden, 2002). A lack of fit is a somewhat unexpected result, because the general RR model is a saturated model in the sense that the number of independent parameters equals the number of independent observed relative frequencies. There are two potential causes for boundary solutions to occur.

Define the probability of observing a *yes* response given that the prevalence of the sensitive characteristic is zero as the chance level. Boundary solutions occur when the relative frequency of a response or response profile is below chance level. One potential cause for a boundary solution is RR sampling variation, by which we mean the random fluctuation in the sample average of the number of red cards. If the prevalence of the sensitive characteristic is zero and less red cards are drawn from the *no* deck than the expected 20%, the relative frequency of *yes* responses will be below chance level. The other potential cause for a boundary solution is cheating. If some of the respondents who have drawn a red card answer *no* instead of the required *yes*, the relative frequency of *yes* responses will drop below chance level if the true prevalence is (near) zero.

3.3 A general Randomized-Response Cheating Model

Cheating can be modeled in the general RR model by the introduction of a cheating parameter θ,

$$\begin{aligned} \boldsymbol{\pi}^* &= (1-\theta)\boldsymbol{P}_K\boldsymbol{\pi} + \theta\mathbf{v} \\ &= \boldsymbol{Q}_K\boldsymbol{\pi} \end{aligned} \tag{3}$$

where θ denotes the probability of cheating, $\mathbf{v}$ is the D-dimensional vector $(0, \ldots, 1)^t$, and the transition matrix $\boldsymbol{Q}_K$ has elements

$$q_{ij} = \begin{cases} (1-\theta)p_{ij} & \text{for } i \neq D, j \in \{1, \ldots, D\} \\ (1-\theta)p_{ij} + \theta & \text{for } i = D, j \in \{1, \ldots, D\} \end{cases} \tag{4}$$

Notice however that model (3) is not identified, in the model there is one more parameter than the observed number of independent relative response frequencies.

3.4 The Log-Linear Randomized-Response Cheating Model

The log-linear randomized-response (LLRR) model is presented by Chen (1989) in the context of misclassification of categorical data and is further developed by Van den Hout and Van der Heijden (2004). Consider the true-response variables A, B and C. The saturated log-linear model $[ABC]$ is given by

$$\log \pi_{abc} = \lambda_0 + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{bc}^{BC} + \lambda_{abc}^{ABC}, \tag{5}$$

where the λ terms are constrained to sum to zero over any subscript. The variables A, B and C are not directly observed, but are linked with the observed data through the misclassification probabilities. The log-linear model linking the true-responses with the observed data is obtained by defining $\eta_j = \log \pi_j, j \in \{1, \ldots, D\}$, and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_D)^t = \boldsymbol{M}\boldsymbol{\lambda}$, where $\boldsymbol{M}$ is the $D \times r$ design matrix of the model and $\boldsymbol{\lambda}$ is the $r \times 1$ parameter vector, for $r \in \{1, 2, \ldots, D\}$.
Model (5) is easily adapted to accommodate cheating by replacing the elements p_{ij} of transition matrix $\boldsymbol{P}_K$ by the elements q_{ij} of transition matrix $\boldsymbol{Q}_K$ defined in (4). However, as for the general RR model the incorporation of a cheating parameter leads to an overparameterized model. We solve this problem by constraining the highest-order interaction parameter to zero. In a design with K variables, constraining the K-factor interaction parameter ensures that the hierarchical structure of the model is preserved. In the case of the three variables A, B and C, the unconstrained LLRRC model (i.e. the model with all parameters except the K-factor interaction parameter), $\theta, [AB, AC, BC]$ is given by

$$\log \pi_{abc} = \lambda_0 + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{bc}^{BC}. \tag{6}$$

The kernel of the log-likelihood of the LLRRC model is given by

$$\ell(\boldsymbol{\lambda}, \theta | \boldsymbol{n}^*, \boldsymbol{P}_K) = \sum_{i=1}^{D} n_i^* \log \left(\sum_{j=1}^{D} q_{ij} e^{\eta_j} \right), \tag{7}$$

where the terms q_{ij} refer to the elements of the transition matrix $\boldsymbol{Q}_K$ defined in (4). Apart from the transition matrix, model (6) is identical to the LLRR model, with $\eta_j = \log \pi_j, j \in \{1, 2, \ldots, D\}$, $\boldsymbol{M}$ the $D \times r$ design matrix, $\boldsymbol{\lambda}$ the $r \times 1$ parameter vector, and $\boldsymbol{\eta} = \boldsymbol{M}\boldsymbol{\lambda}$. In the unconstrained LLRR model $r \in \{1, 2, \ldots, D-1\}$, and the dimensions of the parameter vector $\boldsymbol{\lambda}$ and the design matrix $\boldsymbol{M}$ can be further reduced by constraining interaction parameters to be zero or equal to each other.

4 Example

Table 1 presents a summary of the process of model selection. The table reports the likelihood-ratio statistics L^2 that were obtained from fitting various LLRR models and LLRRC models. The table also presents the cheating parameter estimates for the LLRRC models.

TABLE 1. Model Selection and Cheating Parameter Estimates

Model	$\widehat{\theta}$	L^2	df
H0: $[DEFG]$	-	37.3	0
H1: $[DEF, DEG, DFG, EFG]$	-	38.9	1
H2: $\theta, [DEF, DEG, DFG, EFG]$	.15 (.03)	7.1	0
H3: $\theta, [DE, DF, DG, EF, EG, FG]$	.13 (.05)	7.1	4
H4: $\theta, [DE, DF, DG, EF, EG, FG]$ [a]	.13 (.05)	36.2	9
H5: $\theta, [DE, EF, FG]$ [b]	.15 (.03)	8.4	8
H6: $\theta, [D, E, F, G]$	.27 (.02)	82.9	10

a. equality constraints on all interaction parameters
b. equality constraints $\lambda_{DE} = \lambda_{EF}$

The models H0 and H1 are LLRR models. The saturated LLRR model H0 fit poorly. In the model H1 the highest-order interaction parameter λ_{defg}^{DEFG} is constrained to zero. The slight deterioration in fit indicates the absence of substantial K-factor interaction in the data when cheating is not taken into account. Models H2 to H6 are LLRRC models. The unconstrained LLRRC model H2 has likelihood-ratio statistic of 7.1. Elimination of all 3-factor interaction parameters in model H3 does not affect the fit. The model H5, with equality of the interaction parameters λ_{DE} and λ_{EF}, is the most parsimonious model, with an estimated probability of cheating of 0.15. Model H4 and model H6 illustrate that the fit deteriorates when more constraints are imposed.

TABLE 2. Estimated 2-way Interactions

Model	Interaction	Odds Ratio	$\widehat{\lambda}$
H5	$DE = EF$	181.3	1.30 (.30)
	FG	29.2	.84 (.23)

The associations between the individual variables are presented in Table 2 for the LLRRC model H5. The odds ratios suggest that associations between the variables D and E and the variables E and F in the Health example are very strong, implicating that fraud with respect variable to E is almost always associated with fraud with respect to variables D and F. The variables F and G are also strongly associated.

Table 3 reports the univariate fraud estimates with corresponding confidence intervals. The confidence intervals are obtained with the parametric bootstrap method. When comparing the results of the LLRR and LLRRC models, it can be seen that the correction for cheating has a substantial effect on the univariate fraud estimates.

TABLE 3. Estimated Fraud Probabilities and 95% Confidence Intervals

Model	D	E	F	G
H0	.07 (.05,.11)	.08 (.06,.12)	.11 (.09,.15)	.16 (.12,.20)
H5	.10 (.07,.17)	.11 (.08,.17)	.15 (.11,.21)	.25 (.20,.32)

5 Robustness of the Model

In this section we evaluate the robustness of the cheating parameter estimate and the univariate prevalence estimates against non-zero K-factor interaction. We fitted the unconstrained LLRRC model to the data sets $\boldsymbol{n}_{(-1)}, \boldsymbol{n}_{(0)}$ and $\boldsymbol{n}_{(1)}$, that were computed using the cheating and log-linear parameters of model H5 supplemented with a 4-factor interaction parameter $\lambda_{defg}^{DEFG} \in \{-1, 0, 1\}$. The results are shown in Table 4. The columns labeled "True" show the true parameter values, and the columns labeled "Est." show the estimates of the unconstrained LLRRC model. If $\lambda_{defg}^{DEFG} = 0$ the estimates are unbiased, and otherwise the bias is relatively small.

TABLE 4. Parameter bias as a function of ignored K-factor interaction

	$\boldsymbol{n}^*_{(0)}$	$\boldsymbol{n}^*_{(-1)}$		$\boldsymbol{n}^*_{(1)}$	
Parameter	True=Est.	True	Est.	True	Est.
θ	.146	.146	.142	.146	.149
π_1 (D)	.104	.135	.133	.094	.095
π_1 (E)	.110	.151	.149	.096	.097
π_1 (F)	.150	.189	.187	.137	.138
π_1 (G)	.249	.539	.535	.151	.153

References

Böckenholt, U., and Van der Heijden, P.G.M. (2004). Measuring noncompliance in insurance benefit regulations with randomized response methods for multiple items. In: A. Biggeri, E. Dreassi, C. Lagazio and M. Marchi, In (Eds.), 19th International Workshop on Statistical Modelling. Florence, Italy, pp. 106-110.

Boeije, H. and Lensvelt-Mulders, G. (2002). Honest by chance: A qualitative interview study to clarify respondents' (non)-compliance with computer-assisted randomized response. *Bulletin de Methodologie Sociologique* **75**, 24-39.

Chen, T.T. (1989). A review of methods for misclassified categorical data in epidemiology, *Statistics in Medicine* **8**, 1095-1106.

Edgell, S.E., Himmelfarb S., and Duncan, K.L. (1982). Validity of forced response in a randomized response model. *Sociological Methods and Research* **11**, 89-110.

Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biometrika* 77:436-438.

Lensvelt-Mulders G.J.L.M., Hox, J.J., van der Heijden, P.G.M., and Maas, C. (2005). Meta-analysis of randomized response research, thirty-five years of validation, *Sociological Methods & Research* **33**, 319-348.

Lensvelt-Mulders G.J.L.M., Van der Heijden, P.G.M., Laudy, O., and Van Gils, G. (2006). A validation of a computer assisted randomized response survey to estimate the prevalence of fraud in social security, *Journal of the Royal Statistical Association, Series A: Statistics in Society* **169**, 305-318.

Van den Hout, A. and Van der Heijden, P.G.M. (2002). Randomized response, statistical disclosure control and misclassification: a review. *International Statistical Review* **70**, 269-288.

Van den Hout, A. and Van der Heijden, P.G.M. (2004). The analysis of multivariate misclassified data with special attention to randomized response. *Sociological Methods and Research* **32**, 384-410.

Van der Heijden, P.G.M., Van Gils, G., Bouts, J., and Hox, J.J. (2000). A comparison of Randomized Response, Computer-Assisted Self-Interview and Face-To-Face Direct-Questioning. *Sociological Methods and Research* **28**, 505-537.

Warner, S.L. (1965). Randomized response: a survey technique for eliminating answer bias. *Journal of the American Statistical Association* **60**, 63-69.