RANDOMIZED RESPONSE TO MORE THAN ONE QUESTION

Wanzer Drane, Southern Methodist University and
University of Texas Health Science Center at Dallas

If the answers to multiple questions on a randomized response design give rise to contingency tables, then the tests for association are unaffected by the introduction of the randomized response design. This is illustrated with a two by two table which can easily be extended to larger or higher dimensional tables although with tedium.

1. INTRODUCTION

The randomized response technique of obtaining information from humans on sensitive or stigmatizing questions or characteristics was recently dramatized by Campbell and Joiner (1973). In each case below the authors are obtaining estimates of the proportion of a population who possess the sensitive trait while knowing, *a priori*, the proportions responding to a second innocuous question. The originator's question (Warner, 1965) requires a yes-no response. Abul-Ela et al (1967) extended this to a trinomial response; Warner (1971) introduced a linear model to include many questions simultaneously; Greenberg and Kuebler (1971) introduced quantative responses and Moors (1971) published optimal allocations of proportions of sensitive and innocuous questions.

This paper presents a method of using two separate sensitive questions requiring yes/no responses and obtaining the association between them using a standard two by two contingency table. The extension to more questions and higher dimensional tables is straightforward but tedious and quickly unmanageable.

2. TECHNIQUE

Using a lottery system which the potential respondent can understand such as white and blue marbles in a clay pitcher, instruct him to draw a marble and respond truthfully on Question 1 if the marble is white and respond "YES" if another color. Replace the marble; move to the second pitcher and question and repeat the procedure.

Knowing the proportions of white marbles in the two pitchers we can obtain the following:

1. Estimates of proportions of the population sampled answering truthfully to each of the two questions, their confidence intervals and variance estimates,
2. Estimates of conditional and compound probabilities associated with all combinations of responses (Y,Y),...,(N,N),
3. A test for association between the two questions.

3. THEORY

Let

π_j = proportion of population that "would" answer YES to question j; j = 1, 2,

$\pi_{j\cdot i}$ = proportion of population that "would" answer YES to question j given that they "would" answer YES to question i,

$\pi_{j\cdot\bar{i}}$ = proportion of population that "would" answer YES to question j given that they "would" answer NO to question i,

p_i = proportion of white (truthful) "marbles" in pitcher i, and

$q_i = 1 - p_i$.

The probabilities (λ_{ij}) of the four responses are then

$$\begin{aligned}\lambda_{11} = P(YY) &= p_1\pi_1p_2\pi_{2\cdot 1} + p_1\pi_1(1-p_2) \\ &\quad + (1-p_1)p_2\pi_2 + (1-p_1)(1-p_2) \\ &= p_1p_2\pi_{11} + p_1q_2\pi_1 + q_1p_2\pi_2 \\ &\quad + q_1q_2\end{aligned}$$

$$\begin{aligned}\lambda_{12} = P(YN) &= p_1\pi_1p_2(1-\pi_{2\cdot 1}) + (1-p_1)p_2(1-\pi_2) \\ &= p_1p_2\pi_{12} + q_1p_2(1-\pi_2)\end{aligned}$$

$$\begin{aligned}\lambda_{21} = P(NY) &= p_1(1-\pi_1)p_2\pi_{2\cdot\bar{1}} + p_1(1-\pi_1)(1-p_2) \\ &= p_1p_2\pi_{21} + p_1q_2(1-\pi_1)\end{aligned}$$

$$\begin{aligned}\lambda_{22} = P(NN) &= p_1(1-\pi_1)p_2(1-\pi_{2\cdot\bar{1}}) \\ &= p_1p_2\pi_{22}. \end{aligned} \qquad (3.1)$$

If the two questions are statistically independent, then

$$\pi_{2\cdot 1} = \pi_2,\ \pi_{2\cdot\bar{1}} = \pi_2,\ \pi_{12} = \pi_1\pi_2, \text{ etc., and}$$

the four probabilities are factorable as follows:

$$\begin{aligned}\lambda_{11} &= (p_1\pi_1 + q_1)(p_2\pi_2 + q_2) \\ \lambda_{12} &= (p_1\pi_1 + q_1)p_2(1-\pi_2) \\ \lambda_{21} &= p_1(1-\pi_1)(p_2\pi_2 + q_2) \\ \lambda_{22} &= p_1(1-\pi_1)p_2(1-\pi_2). \end{aligned} \qquad (3.2)$$

The marginals in either case are

$$\begin{aligned}\lambda_{1\cdot} &= P(Y,0) = p_1\pi_1 + q_1 \\ \lambda_{2\cdot} &= P(N,0) = p_1(1-\pi_1) \end{aligned} \qquad (3.3)$$

which is symmetric for P(0,Y) and P(0,N).

Sampling $n_{\cdot\cdot}$ people at random from our population we obtain Tables 1A and 1B with estimates of λ being those of the usual two by two table.

TABLE 1

FREQUENCIES AND PROBABILITIES

A. FREQUENCES OF RESPONSES

		Q_2 Y	N	
Q_1	Y	n_{11}	n_{12}	$n_{1\cdot}$
	N	n_{21}	n_{22}	$n_{2\cdot}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

B. PROBABILITIES OF RESPONSE

		Q_2 Y	N	
Q_1	Y	λ_{11}	λ_{12}	$\lambda_{1\cdot}$
	N	λ_{21}	λ_{22}	$\lambda_{2\cdot}$
		$\lambda_{\cdot 1}$	$\lambda_{\cdot 2}$	1

$n_{\cdot\cdot}$ people are sampled and respond to the two questions as in Table 1A with probabilities as in Table 1B.

From (3.3) we obtain estimates of π_i which are known to be

$$\hat{\pi}_1 = \frac{\hat{\lambda}_{1\cdot} - q_1}{p_1}$$

$$\hat{\pi}_2 = \frac{\hat{\lambda}_{\cdot 1} - q_2}{p_2} \tag{3.4}$$

Confidence limits on $\lambda_{1\cdot}$ and $\lambda_{\cdot 1}$ which are binomial parameters map into confidence limits for π_1 and π_2 using (3.4). Furthermore, the variance estimates of $\hat{\pi}_i$ are also straightforward as $\hat{\pi}_i$ is a linear combination of $\hat{\lambda}$. That is,

$$\mathrm{Var}(\hat{\pi}_1) = \lambda_{1\cdot}(1-\lambda_{1\cdot})/(n_{1\cdot}p_1^2)$$

$$\mathrm{Var}(\hat{\pi}_2) = \lambda_{\cdot 1}(1-\lambda_{\cdot 1})/(n_{\cdot 1}p_2^2), \tag{3.5}$$

and are estimated by substituting $\hat{\lambda}$ wherever needed in (3.5).

Finally, if the two questions are not independent, we return to (3.1) and (3.4) to obtain

$$\hat{\pi} = \begin{pmatrix} (1-q_1-q_2)/(p_1p_2) & -q_2(p_1p_2) & -q_1/(p_1p_2) & 0 \\ q_1/(p_1p_2) & 1/(p_1p_2) & q_1/(p_1p_2) & 0 \\ q_2/(p_1p_2) & q_2/(p_1p_2) & 1/(p_1p_2) & 0 \\ 0 & 0 & 0 & 1/(p_1p_2) \end{pmatrix} \times \begin{pmatrix} n_{11}/n_{\cdot\cdot} \\ n_{12}/n_{\cdot\cdot} \\ n_{21}/n_{\cdot\cdot} \\ n_{22}/n_{\cdot\cdot} \end{pmatrix} + \begin{pmatrix} q_1q_2/(p_1p_2) \\ -q_1/(p_1p_2) \\ -q_2/(p_1p_2) \\ 0 \end{pmatrix} \tag{3.6}$$

which is in the form of

$$\hat{\pi} = A\,\hat{\Lambda} + K\,. \tag{3.7}$$

It is known that $\hat{\Lambda}$ is unbiased for Λ, the four entries of the table, with covariance matrix

$$S = \frac{1}{n_{\cdot\cdot}} \begin{pmatrix} \lambda_{11}(1-\lambda_{11}) & -\lambda_{11}\lambda_{12} & -\lambda_{11}\lambda_{21} & -\lambda_{11}\lambda_{22} \\ -\lambda_{11}\lambda_{12} & \lambda_{12}(1-\lambda_{12}) & -\lambda_{12}\lambda_{21} & -\lambda_{12}\lambda_{22} \\ -\lambda_{11}\lambda_{21} & -\lambda_{12}\lambda_{21} & \lambda_{21}(1-\lambda_{21}) & -\lambda_{21}\lambda_{22} \\ -\lambda_{11}\lambda_{22} & -\lambda_{12}\lambda_{22} & -\lambda_{21}\lambda_{22} & \lambda_{22}(1-\lambda_{22}) \end{pmatrix} \tag{3.8}$$

From (3.7) and (3.8) we quickly conclude that the expected value of $\hat{\pi}$ is $A\,\Lambda + K$ with covariance matrix ASA'. Furthermore, any confidence region on Λ will map one-to-one and on to a confidence region for π using (3.6) and (3.7).

It also follows from (3.1) and (3.2) that the usual test for row-column independence will test $\pi_{11} = \pi_1\pi_2$, etc., or the independence between the two sensitive questions.

4. REFERENCES

[1] Abdul-Ela, A.A., B.G. Greenberg and D.G. Horwitz, "A Multi-proportions Randomized Response Model," Journal of the American Statistical Association, 62 (September 1967), 990-1008.

[2] Campbell, C. and B. Joiner, "How to Get the Answer Without Being Sure You've Asked the Question," The American Statistician, 27 (December 1973), 229-231.

[3] Greenberg, B.G., R.R. Kuebler, Jr., J.R. Abernathy and D.G. Horwitz, "Application of Randomized Response Technique in Obtaining Quantitative Data," Journal of the American Statistical Association, 66 (June 1971), 243-250.

[4] Moors, J.J.A., "Optimization of the Unrelated Question Randomized Response Model," Journal of the American Statistical Association, 66 (September 1971), 627-629.

[5] Warner, S.L., "Randomized Response: A Survey Technique For Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (March 1965), 63-69.

[6] Warner, S.L., "Linear Randomized Response Model," Journal of the American Statistical Association, 66 (December 1971) 884-888.