Journal of Modern Applied Statistical Methods

Volume 3 | Issue 1 Article 11

5-1-2004

Respondent-Generated Intervals (RGI) For Recall in Sample Surveys

S. James Press
University of California, Riverside, jpress@ucr.edu

Recommended Citation
Press, S. James (2004) "Respondent-Generated Intervals (RGI) For Recall in Sample Surveys," *Journal of Modern Applied Statistical Methods*: Vol. 3 : Iss. 1 , Article 11.
DOI: 10.22237/jmasm/1083370260
Available at: http://digitalcommons.wayne.edu/jmasm/vol3/iss1/11

Journal of Modern Applied Statistical Methods
May, 2004, Vol. 3, No. 1, 104-116

1538 – 9472/04/$95.00

Respondent-Generated Intervals (RGI) For Recall in Sample Surveys

S. James Press
Department of Statistics
University of California, Riverside

Respondents are asked for both a basic response to a recall-type question, their usage quantity, and are asked to provide lower and upper bounds for the (Respondent-Generated) interval in which their true values might possibly lie. A Bayesian hierarchical model for estimating the population mean and its variance is presented.

Key words: Bayes, bounds, bracketing, range, recall, survey

Introduction

Answers to recall-type questions are frequently required for surveys carried out by governmental agencies. While answers to such questions might become available to the agency at considerable expense and expenditure of time and effort through record checks, if the information is available at all, it is sometimes more expedient and efficient to directly question samples of the subpopulations for which the answers are required. Unfortunately, because respondents frequently differ greatly in their abilities to recall the correct answers to such questions, estimates of the population mean often suffer from substantial response bias, resulting in large non-sampling errors for the population characteristics of interest. A new protocol for asking such recall-type questions in sample surveys is proposed, and an estimation procedure for analyzing the results that can improve upon the accuracy of the usual sample mean is suggested.

This new method is called Respondent-Generated Intervals (RGI). The procedure involves asking respondents not only for a basic answer to a recall-type question (this basic answer is called the "usage quantity"), but also, the respondent is asked for a smallest value his/her true answer could be, and a largest value his/her true answer could be. These values are referred to as the lower and upper bounds provided. It is assumed that the respondent knew the true value at some point but because of imperfect recall, he/she is not certain of the true value, and also, that the respondent is not purposely trying to deceive.

With the RGI protocol it is being implicitly assumed that there is a distinctive recall probability distribution associated with each respondent. To obtain an estimate of the mean usage quantity in a population typically the simple average of the responses from individuals who may have very different abilities to carry out the recall task is formed. But such a simple average may not necessarily account well for typical unevenness in recall ability.

It may be that an improvement upon population estimates can be made by learning more about the different recall abilities, and then taking them into account in the estimation

S. James Press, is Distinguished Professor, University of California-Riverside. Email: jpress@ucr.edu. The author thanks Dr. Monroe G. Sirken and Professor Judith M. Tanur for their encouragement, and Professor Judith M. Tanur for her suggestions for improving the presentation. Drs. Arthur Kennickell and Steven Heeringa provided helpful background materials and discussions about related approaches to similar problems. Thanks are also due to Dr. Diane Miller for helpful comments, and for developing the macros for calculating the RGI weighted average estimator and related quantities developed in this article. Norman M. Bradburn and Roger Tourangeau provided helpful discussions.

process. Ideally, the respondents could be asked many additional questions about their recall of their true answers for the recall question. That would permit many fractile points on each of their recall distributions to be assessed. Owing to the respondent burden of a long questionnaire, the sometimes heavy cost limitations of adding questions to a survey, the cost of added interviewer time, etc., there may sometimes be a heavy penalty imposed for each additional question posed in the survey questionnaire. The RGI protocol proposes adding to the usage quantity just two additional bounds questions and thereby obtains three points on each respondent's recall distribution. The interpretation of these three points is discussed in the section on estimation.

It is being proposed that respondents provide bounds on what they believe the true value for recall-type questions could possibly be. While there are other survey procedures that also request that respondents provide bounds-type information under certain circumstances, such procedures are not quantitatively associated with improved estimators, as is the RGI estimator. Usually these other procedures ask respondents to select their responses from several (analyst-generated) pre-assigned intervals (sometimes called "brackets").

Kennickell, (1997) described the 1995 Survey of Consumer Finances (SCF), carried out by the National Opinion Research Center at the University of Chicago, as including opportunities for the respondents who answered either "don't know", or "refuse", to select from 8 pre-assigned ranges, or to provide their own lower and upper bounds ("volunteered ranges"). These respondents were addressing what are traditionally recognized as sensitive questions about their assets. By contrast with the survey approach taken in the current research where the respondent is asked for both a basic response and lower and upper bounds, in the SCF, the respondent is given a choice to either give a basic response, or to select from one of several pre-assigned ranges, or to provide volunteered bounds. The pre-assigned intervals are supplied on "range cards" designed for situations in which the respondent has indicated that he/she does not desire to provide the specific usage quantity requested.

Another related technique that has been proposed is called unfolding brackets (Heeringa, Hill, & Howard, 1995). In this approach, respondents are asked a sequence of binary ("yes"/ "no") types of bracketing questions that successively narrow the range in which the respondent's true value might lie.

Several issues about these bounds-, or range-related techniques are not yet resolved. Which of these approaches, RGI, Range, Unfolding Brackets, or more traditional techniques yields the best results? How do these methods compare to one another under various circumstances? How do these different options affect response rate?

Schwartz and Paulin (2000) carried out a study comparing response rates of different groups of randomly assigned participants who used either range cards, unfolding brackets, or RGI, with respect to income questions. To include RGI in their study, Schwartz and Paulin used an early manuscript version of RGI. Schwartz and Paulin (2000) found that all three approaches studied reduced item non-response in that all three techniques presented a viable method for obtaining some income information from respondents who might otherwise have provided none.

In fact, 30% of the participants in the study selected RGI as their favorite range technique. The participants "claimed that they liked this technique because it allowed them to have control over their disclosures; the RGI intervals they provided tended to be narrower than pre-defined intervals; the RGI intervals did not systematically increase with income levels (as did the other techniques); RGI was the only technique that prompted respondents to provide exact values rather than ranges; and RGI allowed respondents to feel the most confident in the accuracy of the information they were providing."

Conrad and Brown (1994; 1996) and Conrad, Brown and Cashman (1998) studied strategies for estimating behavioral frequency using survey interviews. Conrad and his colleagues suggested that when respondents are faced with a question asking about the frequency of a behavior (the usage quantity), if that behavior is infrequent, respondents attempt to count the instances; if it is frequent, they attempt to estimate. When the respondents count they tend to underreport, but when they estimate they tend to over-report. This finding may be relevant to RGI reporting.

Methodology

Let y_i, a_i, b_i denote the basic usage quantity response, the lower bound response, and the upper bound response, respectively, of respondent i, i = 1,…,n. Suppose that the y_i's are all normally distributed $N(\theta_i, \sigma_i^2)$, that the θ_i's are exchangeable, and $\theta_i \sim N(\theta_0, \tau^2)$. It is shown in the Appendix, using a hierarchical Bayesian model, that in such a situation, the conditional posterior distribution of the population mean, θ_0, is given by:

$$(\theta_0 \mid \text{data}, \sigma_i^2, \tau^2) \sim N(\tilde{\theta}, \omega^2), \tag{3.1}$$

where the posterior mean, $\tilde{\theta}$, conditional on the data and (σ_i^2, τ^2) is expressible as a weighted average of the usage quantities and the y_i's, and the weights are expressible approximately as simple algebraic functions of the interval lengths defined by the bounds. The conditional posterior variance, ω^2, drives the associated credibility intervals; it is discussed below.

For normally distributed data it is commonly assumed that lower and upper bounds that represent extreme possible values for the respondents can be associated with 3 standard deviations below, and above, the mean, respectively. That interpretation is used to assess values for the σ_i^2 parameters from: $k_1\sigma_i = b_i - a_i \equiv r_i$, the respondent interval lengths. Analogously, a value for τ^2 is assessed from: $k_2\tau = \bar{b} - \bar{a} \equiv r_0$, the average respondent interval length. It will generally be assumed that $k_1 = k_2 = k = 6$ (corresponding to 3 standard deviations above and below the mean). The assumption of "3" standard deviations is examined numerically in the examples section, and is applied more generally in the Appendix.

The conditional posterior mean is shown in the Appendix to be given by:

$$\tilde{\theta} = \sum_1^n \lambda_i y_i, \tag{3.2}$$

where the λ_i's are weights that are given approximately by:

$$\lambda_i \doteq \frac{\left(\dfrac{1}{r_i^2 + r_0^2}\right)}{\sum_1^n \left(\dfrac{1}{r_i^2 + r_0^2}\right)}. \tag{3.3}$$

Note the following characteristics of this estimator:

1. The weighted average in Eq (3.3) is simple and quick to calculate, without requiring any computer-intensive sampling techniques. A simple Minitab macro is available for calculating it.

2. It will be seen in the examples section that if the respondents who give short intervals are also the more accurate ones, RGI will tend to give an estimate of the population mean that has smaller bias than that of the sample mean. In the special case in which the interval lengths are all the same, the weighted average reduces to the sample mean, $\bar{y}$, where the weights all equal (1/n). In any case, the lambda weights are allnon-negative, and must sum to one.

3. The longer the interval a respondent gives, the less weight is applied to that respondent's usage quantity in the weighted average. The length of respondent i's interval seems intuitively to be a measure of his/her degree of confidence in the usage quantity he/she gives, so that the shorter the interval, the greater degree of confidence that respondent seems to have in the usage quantity he/she reports. Of course a high degree of confidence does not necessarily imply an answer close to the true value.

4. The lambda weights can be thought of as a probability distribution over the values of the usage quantities in the sample. So λ_i

represents the probability that $y = y_i$ in the posterior mean.

5. From equation (A23) in the Appendix it is seen that the conditional variance of the posterior distribution is given by:

$$\omega^2 = \frac{1}{\sum_1^n \left(\frac{1}{\sigma_i^2 + \tau^2} \right)} \doteq \frac{1}{\sum_1^n \left(\frac{1}{\frac{(b_i - a_i)^2}{k_1^2} + \frac{r_0^2}{k_2^2}} \right)}. \tag{3.4}$$

As explained in the discussion just above equation (3.2), it will sometimes be taken to be the case that $k_1 = k_2$ = k = 6 (other values of k are also being studied). So if the precision of a distribution is defined as its reciprocal variance, the quantity $\{\frac{r_i^2 + r_0^2}{36}\}$ is the conditional variance in the posterior distribution corresponding to respondent i, and therefore, its reciprocal represents the conditional precision corresponding to respondent i. Summing over all respondent's precisions gives:
total conditional posterior variance

$$= \omega^2 \doteq \frac{1}{\sum_1^n \left(\frac{36}{r_i^2 + r_0^2} \right)}. \tag{3.5}$$

Thus, another interpretation of λ_i is that it is the proportion of the total conditional posterior precision in the data attributable to respondent i.

The variance of the conditional posterior distribution is given in equation. (3.4). The posterior variance is the reciprocal of the posterior total precision. Because the posterior distribution of the population mean, θ_0, is normal, it is straightforward to find credibility intervals for θ_0. For example, a 95% credibility interval for θ_0 is given by:

$$(\tilde{\theta} - 1.96\omega, \tilde{\theta} + 1.96\omega). \tag{3.6}$$

That is,

$$P\{(\tilde{\theta} - 1.96\omega \le \theta_0 \le \tilde{\theta} + 1.96\omega) \big| data\} = 95\%. \tag{3.7}$$

More general credibility intervals for other percentiles are given in the appendix. From eqn. (3.1) it is seen that the posterior distribution of the population mean, θ_0, is normal. It is therefore straightforward to test hypotheses about θ_0 using the Jeffreys procedure for Bayesian hypothesis testing; (Jeffreys, 1961).

The behavior of the RGI Bayesian estimator is illustrated and examined using some numerical examples. It will be seen that for these examples, the way the RGI estimator works is to assign greater weight to the usage quantities of respondents who give relatively short bounding intervals, and less weight to the usage quantities of those who give relatively long intervals. If the respondents who give short intervals are also the more accurate ones, RGI will tend to give an estimate of the population mean that has smaller bias than the sample mean. Also, the credibility intervals will tend to be shorter and closer to the true population values than the associated confidence intervals.

Example 1

Suppose there is a sample survey of size n = 100 in which the RGI protocol has been used. Suppose also that the true population mean of interest is to be estimated, and it is given by $\theta_0 = 1000$. In this example the usage quantities and the respondents' bounds, (a_i, b_i) are fixed at $r_i = b_i - a_i$, $i = 1, \ldots, n$, arbitrarily, whereas in Example 2 it will be assumed that the data are generated randomly. Define $r_0 = \overline{b} - \overline{a}$. This quantity will be used as an assessment for τ, the common standard deviation of θ_i, the mean for respondent i.

Assume that the first 50 respondents all have excellent memories and are quite accurate in their responses. Suppose the intervals these accurate respondents give are:

$$(a_1, b_1), \ldots, (a_{50}, b_{50}) = (975, 975), \ldots, (975, 975).$$

That is, they are all not only pretty accurate, but they all believe that they are accurate, so they respond to the bounds questions with degenerate intervals whose lower and upper bounds are the same. Accordingly, these accurate respondents all report intervals of length $r_i = 0$, and usages of equal amounts, $y_i = 975$ (compared with the true value of 1000).

Next suppose that the last 50 respondents all have poor memories and are inaccurate. They report the intervals:

$$\begin{aligned}&(a_{51}, b_{51}), \ldots, (a_{100}, b_{100}) \\ &= (500,1500), \ldots, (500,1500)\end{aligned},$$

that have lengths of $r_i = 1000$, and they report equal usage quantities of y_i = 550. Their true values, θ_i, may all be different from one another, but assume that they all guess 550. It is now found that: $\bar{a} = 737.5,$ and $\bar{b} = 1237.5,$ so $r_0 = \bar{b} - \bar{a} = 500.$

RGI Bayesian Point Estimate of the Population Mean

The weights are calculated to be given by:

$$\lambda_i = \begin{cases} .0167, & i = 1, \ldots 50 \\ .0033, & i = 51, \ldots, 100 \end{cases}$$

It is easy to check that: $\sum_1^{100} \lambda_i = 1.$ It may now be readily found that the conditional posterior mean RGI estimator of the population mean, θ_0, is given by:

$$\tilde{\theta} = \sum_{i=1}^{100} \lambda_i y_i = 904.167.$$

The corresponding sample mean is given by: $\bar{y} = 762.5$. The numerical error (bias) of the posterior mean is given by 1000 - $\tilde{\theta}$ = 1000-904.167 = 95.833. The numerical error (bias) of the sample mean is given by $1000 - \bar{y} = 1000 - 762.5 = 237.5.$ The RGI estimator has reduced the bias error by 237.5 - 95.833 = 141.667, or about 60%, compared with the standard error of the sample mean.

It is also interesting to compare interval estimates of the population mean by comparing the standard error of $\bar{y}$, with ω, the standard deviation of the posterior distribution of $\tilde{\theta}$. These estimates give rise to the corresponding confidence and credibility intervals for θ_0, respectively.

From Eq (3.4) it may readily be found that for the data in this example, $\omega = 10.76$. It is also easy to check that for the data, the standard deviation of the data is 213.56. So the standard error for a sample of size 100 is 213.56/10, or 21.36. Thus, the RGI estimate of standard deviation is less than half that of the sample mean.

Correspondingly, the length of the 95% credibility interval $2(1.96)\omega$ = 42.18, while the length of the 95% confidence interval is 2(1.96)(21.36) = 83.74. The 95% confidence interval is about twice as long as the 95% credibility interval. The 95% credibility interval is given by: (883.081, 925.253). The 95% confidence interval is given by: (720.63, 804.37). Note in this example that:

1. Neither the RGI credibility interval nor the confidence interval covers the true value of 1000 (all usage quantities were biased downward).

2. The confidence and credibility intervals do not even overlap (but the entire credibility interval is closer to the true value).

3. It is expected to find many situations for which the bias error of the RGI estimator is smaller than that of the sample mean; however, the differences may be more, or less, dramatic compared with their values in this example.

Now examine some variations of the conditions in this example to explore the robustness of the RGI estimator with respect to variations in the assumptions.

Variation 1

Suppose that there were only 30 accurate respondents (instead of the 50 assumed in this example), responding in exactly the same way, and 70 inaccurate respondents (instead of the 50 assumed in the example), the RGI estimate would still have been an improvement in bias error over that of the sample mean, although the improvement in bias error would have been smaller (35.03%).

Variation2

Now take the example to the extreme by supposing that there were only 1 accurate respondent (instead of the original 50 assumed in the example), responding in exactly the same way, and 99 inaccurate respondents (instead of the 50 assumed in the example), the RGI estimate would still have been an improvement in bias error over that of the sample mean, although the improvement in bias error would have been only 9.5%.

Variation 3

How are the population mean estimates affected by the values selected for k_1 and k_2? First recall that as long as k_1 and k_2 are the same, the posterior mean is unaffected by the value of k. However, the posterior variance and the credibility intervals are affected. Continue to take $k_1 = k_2 = k$ but vary the value of k and assume the original split of 50 accurately-reporting respondents and 50 inaccurately-reporting respondents. Table 1 below compares results as a function of the common $k_1 = k_2 = k$ selected.

That is, if a_i denotes the lower bound provided by respondent i, and b_i denotes the same respondent's upper bound, the assessment method being used has been to take $\tau = \frac{1}{n}\sum_{i=1}^{n}(b_i - a_i) = \bar{b} - \bar{a}$, where the bar denotes average. But consider as an alternative the range assessment: $\tau = b_0 - a_0$, where $a_0 \equiv \min_i(a_i), \quad b_0 \equiv \max_i(b_i)$.

Table 1. Effect Of Common Value of "k"

k	ω = posterior standard deviation	95% credibility interval	Length of credibility interval
4	16.14	(872.54, 935.80)	63.26
5	12.91	(878.86, 929.47)	50.61
6	**10.76**	**(883.08, 925.25)**	**42.17**
7	9.22	(886.09, 922.24)	36.15
8	8.07	(888.35, 919.98)	31.63

Examination of Table 1 suggests that for general purposes, selecting a common k and taking it to be k = 6 (bold face) is a reasonable compromise.

Note that the *range of belief* is reflected by the interval $(a_0, b_0) \equiv (500,1500)$. How will the estimates of the population mean be affected? Results are shown in Table 2 for the two different methods for the 50/50 split of accurate and inaccurate respondents used in the original example.

Table 2 demonstrates that in this example, the "average" procedure used for assessing produces better results than the range procedure: there is less bias, smaller posterior variance, a shorter credibility interval, and a credibility interval that is also closer to the true population mean (the population mean in this example was 1000). It is therefore recommended that τ be assessed by using the average, rather than the range procedure.

Example 2

In this example the usage quantities from appropriate normal distributions are simulated while the respondents' bounds are fixed conveniently. Again assume a survey of 100 respondents and again use $k_1 = k_2 = 6$.

Table 2. Comparing Methods For Assessing τ

	posterior mean	posterior variance	95% credibility interval	length of 95% credibility interval
average assessment procedure for τ	904.167	115.741	(883.081, 925.253)	42.172
range assessment procedure for τ	833.333	370.370	(795.614, 871.053)	75.439

Adopt usage quantities that are generated from distinct normal distributions, the average of whose means is $\theta_0 = 1000$, and whose standard deviations are all 300. Such usage data are included within the framework of the model. The actual usage quantities that were generated are given in Tables 3a and 3b. Assign lower and upper bound intervals of (900, 1100) for the 26 usage quantities (out of 100) between 900 and 1100 (usages that lie close to the true population value), and assign lower and upper bound intervals of (200, 1900) for the other 74 usage quantities (those usages that lie further from the true population value). The lower and upper bounds adopted are given in Tables 3a 3b, as are the values of the calculated lambda weights (which sum to one).

Bias Reduction

The sample mean for this example is 964.497. The posterior mean or RGI estimator is 973.816. The bias error for the sample mean is 35.503, while that for the RGI estimator is 26.184. The RGI estimator has reduced the bias by 9.319, or 26.2%.

The standard deviation of the usage quantities is 324.1 while the standard error of the sample mean is 32.4. So a 95% confidence interval is (900.993, 1028.001). It has length 127.008.

The standard deviation of the RGI estimator (standard deviation of the posterior distribution of the population mean estimator) is about 30.0, so a 95% credibility interval is (915.023, 1032.61). It has length 117.587.

The result is that both the 95% confidence interval and the 95% credibility intervals cover the true population values, but the credibility interval is shorter.

Conclusion

A new method for asking recall-type questions in sample surveys has been proposed. The method can substantially reduce the non-sampling bias error compared with the error of the sample mean. It is anticipated that over time, even better techniques will be developed to take advantage of this path to improved estimation accuracy. Such techniques will likely prompt respondents who believe they are accurate in their recollection to provide short bounding intervals, and conversely, the techniques will prompt respondents who are uncertain of the quantity to be recalled to give longer bounding intervals.

The RGI technique may also prove to be less threatening to respondents faced with answering sensitive questions. Respondents who might not answer such questions at all, might be willing at least to provide bounds, thereby increasing response rate. Therefore, there also may be a response rate benefit that accrues from the use of the RGI protocol in surveys containing sensitive questions.

The RGI estimator proposed appears to be robust with respect to variations in the distributions of the data and in the assumptions of the model. The data selected for Example 1 didn’t follow any familiar distribution. What is important is that one or more accurate respondents also gave short bounding intervals that could be used in the weighted average, independently of any distributional assumptions. This robustness property appears to be very promising for survey applications.

Table 3a. Raw Data and Lambda Weights for Normal Data Example

Number	usage	lo-bound	up-bound	lambda	Number	usage	lo-bound	up-bound	lambda
1	518.67	200	1900	0.007033	26	414.04	200	1900	0.007033
2	1428.18	200	1900	0.007033	27	1282.15	200	1900	0.007033
3	1352.09	200	1900	0.007033	28	1317.67	200	1900	0.007033
4	919.02	900	1100	0.018446	29	466.53	200	1900	0.007033
5	572.47	200	1900	0.007033	30	869.69	200	1900	0.007033
6	822.51	200	1900	0.007033	31	820.92	200	1900	0.007033
7	814.42	200	1900	0.007033	32	1223.39	200	1900	0.007033
8	431.61	200	1900	0.007033	33	1330.02	200	1900	0.007033
9	1099.68	900	1100	0.018446	34	1267.04	200	1900	0.007033
10	1318.16	200	1900	0.007033	35	1123.93	200	1900	0.007033
11	704.25	200	1900	0.007033	36	1155.08	200	1900	0.007033
12	918.67	900	1100	0.018446	37	1206.36	200	1900	0.007033
13	1105.79	200	1900	0.007033	38	1082.61	900	1100	0.018446
14	931.64	900	1100	0.018446	39	997.76	900	1100	0.018446
15	1839.33	200	1900	0.007033	40	1205.7	200	1900	0.007033
16	625.11	200	1900	0.007033	41	675.03	200	1900	0.007033
17	1482.88	200	1900	0.007033	42	1642.04	200	1900	0.007033
18	691.66	200	1900	0.007033	43	909.12	900	1100	0.018446
19	1218.58	200	1900	0.007033	44	834.9	200	1900	0.007033
20	761.49	200	1900	0.007033	45	439.11	200	1900	0.007033
21	1041.2	900	1100	0.018446	46	279.14	200	1900	0.007033
22	283.22	200	1900	0.007033	47	996.48	900	1100	0.018446
23	1276.98	200	1900	0.007033	48	237.63	200	1900	0.007033
24	640.76	200	1900	0.007033	49	1284.94	200	1900	0.007033
25	1442.09	200	1900	0.007033	50	1143.45	200	1900	0.007033

Table 3b. Raw Data and Lambda Weights for Normal Data Example

Number	usage	lo-bound	up-bound	lambda	Number	usage	lo-bound	up-bound	lambda
51	906.23	900	1100	0.018446	76	902.57	900	1100	0.018446
52	903.56	900	1100	0.018446	77	706.01	200	1900	0.007033
53	1223.66	200	1900	0.007033	78	953.28	900	1100	0.018446
54	643.64	200	1900	0.007033	79	909.81	900	1100	0.018446
55	1355.94	200	1900	0.007033	80	1097.39	900	1100	0.018446
56	1270.2	200	1900	0.007033	81	1112.48	200	1900	0.007033
57	744.64	200	1900	0.007033	82	1082.91	900	1100	0.018446
58	1036.39	900	1100	0.018446	83	1013.8	900	1100	0.018446
59	673.09	200	1900	0.007033	84	1084.29	900	1100	0.018446
60	1234.4	200	1900	0.007033	85	750.5	200	1900	0.007033
61	435.02	200	1900	0.007033	86	1306.53	200	1900	0.007033
62	1196.57	200	1900	0.007033	87	875.54	200	1900	0.007033
63	849.43	200	1900	0.007033	88	994	900	1100	0.018446
64	1099.91	900	1100	0.018446	89	605.55	200	1900	0.007033
65	1025.81	900	1100	0.018446	90	982.78	900	1100	0.018446
66	1515.97	200	1900	0.007033	91	828.1	200	1900	0.007033
67	646.71	200	1900	0.007033	92	820.58	200	1900	0.007033
68	1119.5	200	1900	0.007033	93	644.87	200	1900	0.007033
69	460.59	200	1900	0.007033	94	630.19	200	1900	0.007033
70	1046.75	900	1100	0.018446	95	762.45	200	1900	0.007033
71	1378.79	200	1900	0.007033	96	1154.98	200	1900	0.007033
72	1267.88	200	1900	0.007033	97	764.34	200	1900	0.007033
73	799.94	200	1900	0.007033	98	984.5	900	1100	0.018446
74	1796.67	200	1900	0.007033	99	973.23	900	1100	0.018446
75	775.25	200	1900	0.007033	100	657.28	200	1900	0.007033

References

Conrad, F. G., & Brown, N. R. (1994). Strategies for Estimating Category Frequency: Effects of Abstractness and Distinctiveness, *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, p. 1345-1350. Erlbaum Associates.

Conrad, F. G., & Brown, N. R (1996). Estimating frequency: A multiple strategy approach. In D. J. Hermann, C., McEvoy, C., Hertzog, P., Hertel, & M. K., Johnson (Eds.) *Basic and Applied Memory Research: Practical Applications, Volume 2.* Mahwah, NJ: Lawrence

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory*, *6*, 339-366.

Heeringa, S. G., Hill, D. H., & Howell, D. H. (1995, June). Unfolding brackets for reducing Item non-response in economic surveys. *Health and Retirement Study Working Paper Series*, Paper No. 94-029, Institute for Social Research, University of Michigan.

Jeffreys, H. (1961). *Theory of Probability*. Third Edition, Oxford: Oxford University Press.

Kennickell, A. B. (1997, January). Using range techniques with CAPI in the 1995 Survey of Consumer Finances. Board Governors of the Federal Reserve System, Washington, D.C. 20551.

Schwartz, L. K., & Paulin, G. D. (2000). Improving Response Rates to Income Questions, *Proceedings of the American. Statistical. Association.* Section on Survey Research Methods, pp. 965-969.

Appendix

In this Appendix, a hierarchical Bayesian model is developed for estimating the posterior distribution of the population mean for data obtained by using the RGI protocol. Suppose respondent i gives a point response y_i, and bounds (a_i, b_i), $a_i \le b_i$, i = 1,…, n, as his/her answers to a factual recall question. Assume:

$$(y_i | \theta_i, \sigma_i^2) \sim N(\theta_i, \sigma_i^2). \quad \text{(A1)}$$

The normal distribution will often be appropriate in situations for which the usage quantity corresponds to a change in some quantity of interest. Assume the means of the usage quantities are themselves exchangeable, and normally distributed about some unknown population mean of fundamental interest, θ_0:

$$(\theta_i | \theta_0, \tau^2) \sim N(\theta_0, \tau^2). \text{ (A2)}$$

Thus, respondent i has a recall distribution whose true value is θ_i (each respondent is attempting to recall a different number of visits to the doctor last year). It is desired to estimate θ_0. Assume $(\sigma_1^2, \ldots, \sigma_n^2, \tau^2)$ are known; they will be assigned later. Denote the column vector of usage quantities by $y = (y_i)$, and the column vector of means by $\theta = (\theta_i)$. Let $\underset{\sim}{\sigma}^2 = (\sigma_i^2)$ denote the column vector of data variances. The joint density of the y_i's is given in summary form by:

$$p(\underset{\sim}{y} | \underset{\sim}{\theta}, \underset{\sim}{\sigma}^2) \propto \exp\left\{(-\frac{1}{2})\sum_1^n \left(\frac{y_i - \theta_i}{\sigma_i}\right)^2\right\}.$$

(A3)

The joint density of the θ_i's is given by:

$$p(\underset{\sim}{\theta} | \theta_0, \tau^2) \propto \exp\left\{(-\frac{1}{2})\sum_1^n \left(\frac{\theta_i - \theta_0}{\tau}\right)^2\right\}.$$

(A4)

So the joint density of $(\underset{\sim}{y}, \underset{\sim}{\theta})$ is given by:

$$p(\underset{\sim}{y}, \underset{\sim}{\theta} | \theta_0, \tau^2, \underset{\sim}{\sigma}^2) = p(\underset{\sim}{y} | \underset{\sim}{\theta}, \underset{\sim}{\sigma}^2) p(\underset{\sim}{\theta} | \theta_0, \tau^2)$$

or, multiplying (A3) and (A4), gives:

$$\begin{aligned} &p(\underset{\sim}{y}, \underset{\sim}{\theta} | \theta_0, \tau^2, \underset{\sim}{\sigma}^2) \\ &\propto \exp\left\{(-\frac{1}{2})\left[\sum_1^n \left(\frac{y_i - \theta_i}{\sigma_i}\right)^2 + \sum_1^n \left(\frac{\theta_i - \theta_0}{\tau}\right)^2\right]\right\} \\ &\propto \exp\left\{(-\frac{A(\underset{\sim}{\theta})}{2})\right\}, \end{aligned}$$

(A5)

where:

$$A(\underset{\sim}{\theta}) \equiv \sum_1^n \left(\frac{y_i - \theta_i}{\sigma_i}\right)^2 + \sum_1^n \left(\frac{\theta_i - \theta_0}{\tau}\right)^2.$$

(A6)

Expand (A6) in terms of the θ_i's by completing the square. This takes some algebra. Then:

$$A(\underset{\sim}{\theta}) = \sum_1^n \left\{\alpha_i \left[\left(\theta_i - \frac{\beta_i}{\alpha_i}\right)^2 + \left(\frac{\gamma_i}{\alpha_i} - \frac{\beta_i^2}{\alpha_i^2}\right)\right]\right\},$$

(A7)

$$\alpha_i = \frac{1}{\sigma_i^2} + \frac{1}{\tau^2}, \quad \beta_i = \frac{y_i}{\sigma_i^2} + \frac{\theta_0}{\tau^2}, \quad \gamma_i = \frac{\theta_0^2}{\tau^2} + \frac{y_i^2}{\sigma_i^2}. \qquad \text{(A8)}$$

Now find the marginal density of $\underset{\sim}{y}$ by integrating (A5) with respect to $\underset{\sim}{\theta}$. Then:

$$p(\underset{\sim}{y} \mid \theta_0, \tau^2, \underset{\sim}{\sigma}^2) \propto J(\theta_0) \exp\left\{\left(-\frac{1}{2}\sum_1^n \alpha_i \delta_i\right)\right\},$$

$$J(\theta_0) \equiv \int \exp\left\{\left(-\frac{1}{2}\right)\sum_1^n \alpha_i \left(\theta_i - \frac{\beta_i}{\alpha_i}\right)^2\right\} d\underset{\sim}{\theta},$$

$$\delta_i = \left(\frac{\gamma_i}{\alpha_i} - \frac{\beta_i^2}{\alpha_i^2}\right). \qquad \text{(A9)}$$

Rewriting (A9) in vector and matrix form, to simplify the integration, it is found that if

$$f \equiv \left(\frac{\beta_i}{\alpha_i}\right), \qquad K^{-1} \equiv diag(\alpha_1, \ldots, \alpha_n),$$

$$(\underset{\sim}{\theta} - \underset{\sim}{f})' K^{-1} (\underset{\sim}{\theta} - \underset{\sim}{f}) = \sum_1^n \alpha_i \left(\theta_i - \frac{\beta_i}{\alpha_i}\right)^2. \qquad \text{(A10)}$$

Carrying out the (normal) integration gives:

$$p(\underset{\sim}{y} \mid \theta_0, \tau^2, \underset{\sim}{\sigma}^2) \propto \frac{1}{\left|K^{-1}\right|^{1/2}} \exp\left\{\left(-\frac{1}{2}\sum_1^n \alpha_i \delta_i\right)\right\}. \qquad \text{(A11)}$$

Now note that $\left|K^{-1}\right| = \prod_1^n \alpha_i =$ constant and the constant can be absorbed into the proportionality constant, but δ_i depends on θ_0. So:

$$p(\underset{\sim}{y} \mid \theta_0, \tau^2, \underset{\sim}{\sigma}^2) \propto \exp\left\{\left(-\frac{1}{2}\sum_1^n \alpha_i \delta_i\right)\right\}. \qquad \text{(A12)}$$

Now apply Bayes' theorem to θ_0 in (A12).

$$p(\theta_0 \mid \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2) \propto p(\theta_0) \exp\left\{\left(-\frac{1}{2}\sum_1^n \alpha_i \delta_i\right)\right\}, \qquad \text{(A13)}$$

where $p(\theta_0)$ denotes a prior density for θ_0. Prior belief (prior to observing the point and bound estimates of the respondents) is that for the large sample sizes typically associated with sample surveys, the population mean, θ_0, might lie, with equal probability, anywhere in the interval (a_0, b_0), where a_0 denotes the smallest lower bound given by any respondent, and b_0 denotes the largest. So adopt a uniform prior distribution on (a_0, b_0). To be fully confident of covering all possibilities, however, adopt the (improper) prior density on the entire positive real line. Therefore adopt a prior density of the form:

$$p(\theta_0) \propto \text{constant}, \qquad \text{(A14)}$$

for all θ_0 on the positive half line. (In some survey situations the same survey is carried out repeatedly so that there is strong prior information available for providing a realistic finite range for θ_0; in such cases it is possible to improve on the estimator by using a proper prior distribution for θ_0 instead of the one given in eqn. (A14).) Inserting (A14) into (A13), and noting that $p(\theta_0) \propto$ constant, gives:

$$p(\theta_0 \left| \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2 \right.) \propto \exp\left\{ \left(-\frac{1}{2}\sum_1^n \alpha_i \delta_i \right) \right\}. \tag{A15}$$

Next substitute for δ_i and complete the square in θ_0 to get the final result:

$$p(\theta_0 \left| \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2 \right.) \propto \exp\left\{ \left(-\frac{u}{2} \right)\left(\theta_0 - \frac{v}{u} \right)^2 \right\}, \tag{A16}$$

$$u = \sum_1^n \left(\frac{1}{\tau^2} - \frac{1}{\alpha_i \tau^4} \right), \qquad v = \sum_1^n \left(\frac{y_i}{\alpha_i \sigma_i^2 \tau^2} \right). \tag{A17}$$

Thus, the conditional posterior density of θ_0 is seen to be expressible as:

$$(\theta_0 \left| \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2 \right.) \sim N(\tilde{\theta}, \omega^2), \tag{A18}$$

where: $\tilde{\theta} \equiv \dfrac{v}{u}$, and $\omega^2 \equiv \dfrac{1}{u}$. (A 19)

Conditional Posterior Mean Of θ_0 As A Convex Mixture Of Usages

The appropriate measure of location of the posterior distribution in Eq. (A18) to use in any given situation depends upon the loss function that is appropriate. For many cases of interest the quadratic loss function (mean squared error) is appropriate. For such situations, interest centers on the posterior mean (under the normality assumptions in the current model, the conditional posterior distribution of θ_0 is also normal, so the posterior mean, median, and mode are all the same). It can be readily found by simple algebra that if:

$$\lambda_i \equiv \frac{\left(\dfrac{1}{\sigma_i^2 + \tau^2} \right)}{\sum_1^n \left(\dfrac{1}{\sigma_i^2 + \tau^2} \right)}, \qquad \sum_1^n \lambda_i = 1, \tag{A20}$$

$$\tilde{\theta} = \sum_1^n \lambda_i y_i .$$

Thus, the mean of the conditional posterior density of the population mean is a convex combination of the respondents' point estimates, that is, their usage quantities. It is an unequally weighted average of the usage quantities, as compared with the sample estimator of the population mean, which is an equally weighted estimator, $\overline{y}$. Interpret $(\sigma_i^2 + \tau^2)^{-1}$ as the precision attributable to respondent i's response, and $\sum_1^n (\sigma_i^2 + \tau^2)^{-1}$ as the total precision attributable to all respondents; then, λ_i is interpretable as the proportion of total precision attributable to respondent i. Thus, the greater his/her precision proportion, the greater the weight that is automatically assigned to respondent i's usage response.

Assessing the Variance Parameters

Take: a) $k_1\sigma_i = (b_i - a_i)$, for all i = 1,…, n; for some k_1, such as k_1 = 4,5,6. Typically, take k = 6 (3 standard deviations on either side of the mean). Define, as above: b) $\overline{a} = \frac{1}{n}\sum_1^n a_i$, and $\overline{b} = \frac{1}{n}\sum_1^n b_i$. Then, take c) $k_2\tau = \overline{b} - \overline{a}$ for some pre-assigned k_2. τ is the same for all respondents. Use an interval of 3 standard deviations on either side of the (normal) mean of the individual recall distribution means for the respondents. It is required to have an assessment that will be reasonable for all respondents. Use the average respondent interval.

Different analysts might interpret the k's somewhat differently. Using these variance assessments, the weights become approximately:

$$\lambda_i \doteq \frac{\left(\dfrac{1}{\dfrac{(b_i - a_i)^2}{k_1^2} + \dfrac{r_0^2}{k_2^2}} \right)}{\sum_1^n \left(\dfrac{1}{\dfrac{(b_i - a_i)^2}{k_1^2} + \dfrac{r_0^2}{k_2^2}} \right)}, \qquad \sum_1^n \lambda_i = 1, \tag{A21}$$

where: $r_0 \equiv \overline{b} - \overline{a}$. Note that in the special case that $k_1 = k_2$, the k's cancel out in numerator and denominator, so that the weights do not depend upon the k's. Then, the weights become:

$$\lambda_i \doteq \frac{\left(\dfrac{1}{(b_i - a_i)^2 + r_0^2} \right)}{\sum_1^n \left(\dfrac{1}{(b_i - a_i)^2 + r_0^2} \right)}. \tag{A22}$$

Conditional Posterior Variance Of θ_0

It is straightforward to check that the conditional posterior variance of θ_0 is given by:

$$\omega^2 = \frac{1}{\sum_1^n \left(\dfrac{1}{\sigma_i^2 + \tau^2} \right)} \doteq \frac{1}{\sum_1^n \left(\dfrac{1}{\dfrac{(b_i - a_i)^2}{k_1^2} + \dfrac{r_0^2}{k_2^2}} \right)}, \tag{A23}$$

the reciprocal of the total precision for all respondents in the sample. For $k_1 = k_2 = k$,

$$\omega^2 \doteq \frac{1}{\sum_1^n \left(\dfrac{k^2}{(b_i - a_i)^2 + r_0^2} \right)}, \tag{A24}$$

so that in this case, while the conditional posterior mean does not depend upon k, the conditional posterior variance does. So the conditional posterior distribution of the population mean is given by:

$$(\theta_0 \mid \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2) \sim N(\tilde{\theta}, \omega^2), \tag{A25}$$

where $\tilde{\theta}$ and ω^2 are given, respectively, in (A19), (A20), and (A23) or (A24).

Credibility Intervals

Let z_γ denote the $\gamma / 2$-percentile of the standard normal distribution. Then, from (A25), a (100-γ)% credibility interval for the population mean, θ_0 is given by:

$$(\tilde{\theta} - z_\gamma \omega,\ \tilde{\theta} + z_\gamma \omega). \tag{A26}$$

That is,

$$P\{\tilde{\theta} - z_\gamma \omega \le \theta_0 \le \tilde{\theta} + z_\gamma \omega \mid \underset{\sim}{y}, \tau^2, \underset{\sim}{\sigma}^2\} = (100 - \gamma)\%.$$

(A27).