[illegible]

SURVEY OF FARM OPERATORS IN SELECTED IOWA COUNTIES

BY

G. E. BATTESE, W. A. FULLER AND R. D. HICKMAN

STATISTICAL LABORATORY
IOWA STATE UNIVERSITY
AMES, IOWA

PREPARED AS A REPORT TO THE
STATISTICAL REPORTING SERVICE

JULY, 1972

CONTENTS

PREFACE

This research was conducted under a cooperative agreement between Iowa State University and the USDA's Statistical Reporting Service. The Statistical Reporting Service funded the project. Technical personnel from the Standards and Research Division of the Statistical Reporting Service provided assistance in the planning of the study. The authors directed the collection and analysis of the data.

The helpful comments of Harold Huddleston, Statistical Reporting Service, and the programming assistance of Robert Mason, Iowa State University, are gratefully acknowledged.

SUMMARY

This report discusses a survey of Iowa farm operators in which two responses were obtained for each sample respondent. The farm operators in the survey were interviewed in person by two different interviewers one month apart. A model, which contains interviewer effects, trial effects, and response errors associated with the respondents, is presented for the survey responses. The variability of responses due to the interviewers and the respondents is calculated. Although the interviewer effects are not significant, the variance of respondent errors is a significant proportion of the total variance of responses for certain acreage, livestock and labor items in the survey.

INTERVIEWER EFFECTS AND RESPONSE ERRORS IN A REPLICATED SURVEY OF FARM OPERATORS IN SELECTED IOWA COUNTIES

by

G. E. Battese, W. A. Fuller and R. D. Hickman*

1. INTRODUCTION

In the last two decades, considerable attention has been given to errors of measurement in survey sampling. Cochran [2] cites some of the papers in this area in his review paper on measurement errors in statistics.[1/] The earlier papers define the "response error" involved in a survey response as the difference between the observed value and the "true" value for the respondent involved and analyze the survey data by use of analysis-of-variance procedures (e.g., [5], [6], [8], [10], [11]). In several more recent papers, survey responses are considered in a somewhat different manner. In these papers, attempts are made to develop general response models by defining a response deviation as the difference between the actual response and the expected response for the given respondent involved (e.g., [1], [3], [4], [12]). The survey responses are considered as the sum of the response deviation, the sampling deviation (defined as the difference between the expected response for the given respondent and the average of the expected re-

*G. E. Battese is an associate in the Statistical Laboratory, W. A. Fuller is a professor of statistics and economics, and R. D. Hickman is an associate professor of statistics, all of Iowa State University.

1/ The numbers in the brackets refer to papers listed in the References at the end of the report.

sponses of the respondents in the sampled population), and the average of the expected responses. Different correlations between the response deviations and sampling deviations are defined. In the more general formulations, it is usually impossible to estimate all the parameters defined in the model (e.g., see [3]).

In this report the analysis of the data from a survey of farm operators in selected Iowa counties, is considered. The survey was designed and conducted to investigate the variance of response errors for selected items of the 1970 June acreage, livestock and labor enumeration survey questionnaire. The variability of responses due to the interviewers in the survey is of particular interest for the estimation of the percentage by which interviewer effects increase the variance of average responses for different average interviewer workloads. The survey design and data obtained are discussed in section 2; the model for the survey responses is presented in section 3; the analyses of the data are discussed in sections 4-6; and section 7 contains concluding remarks.

2. DESCRIPTION OF THE SURVEY

An area sample of farm operators was selected within each of three geographic areas in Iowa. Each of these areas consisted of two adjoining counties. Four interviewers were assigned to each of the three areas, and interviews were obtained with eligible farm operators early in September 1970 and again about one month later.

Interviewing was restricted to farm operators who satisfied two criteria: (i) operate forty or more acres in the farm, and (ii) have a major cattle enterprise (ten or more head of cattle or calves) or a major hog enterprise (twenty five or more hogs and pigs). Sixty area segments were constructed within each area to yield approximately two farm operators per segment, or approximately eighty four eligible farm operators per area. Within each area, segments were randomly assigned to the four interviewers for the given area, and the first trial of interviewing was performed. The respondents (farm operators) for a given interviewer were then randomly partitioned into three groups of as nearly equal size as possible and assigned to the other three interviewers within that area for the second-trial interviews.

The twelve interviewers selected for the survey participated in a two-day training school immediately before the first-trial interviews. The interviewers were told that the basic objective of the survey was to estimate the size of farm operations and changes in livestock inventories for three areas in Iowa during September and October. They also were told that the estimation of the variability of farm operators' responses would aid in the design of improved questionnaires for agricultural surveys. Intensive instruction was given on the survey questionnaires and the interviewers participated in practice interviews with farm operators in the vicinity of Ames. Half of the interviewers in the survey had previous experience in several of the Iowa State University Statistical Laboratory surveys. The interviewers were paid by the hour, the rate depending on the experience and efficiency of the particular interviewer in previous surveys.

Most of the items in the survey questionnaires were extracted from the 1970 June acreage, livestock and labor enumeration survey questionnaire. The sample operators were told by the interviewers at trial 1 that they would be cooperating in a panel study of livestock movements and inventories and that they would be interviewed again the next month. The items on the trial-2 questionnaires were constructed so that the exact question was either (i) repeated with reference to the data of the first interview or (ii) tied to the date of the second interview with additional items included to obtain changes in inventory of the variable in question. Thus, two responses on each variable were obtained by two different interviewers.

In trial 1 of the survey, interviews were obtained with 92.0 percent of the eligible farm operators identified in the three areas. Of the farm operators assigned for trial 2, 91.8 percent were interviewed. Completed interviews were obtained from 262 farm operators in both trials of the survey. The respondents interviewed by the same interviewer for trial 1 and by the same interviewer for trial 2 (but different from the trial-1 interviewer) are said to belong to the same "respondent group." In each area, there were thus twelve different respondent groups. The number of respondents in each of the thirty six respondent groups is presented in table 1 of Appendix B. In trial 2, one interviewer was unable to participate and was replaced by an interviewer who had previous experience in sample surveys conducted by Iowa State University.

The responses obtained in the two trials of the survey for twenty one variates were analyzed. The particular acreage, livestock, labor and income items included in the analysis are shown in table 2.

3. MODEL FOR SURVEY RESPONSES

For a given survey variate, the true value for the j-th sample respondent in the i-th area is denoted by y_{ij} , and we assume the model representation

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad (1)$$

where μ is the overall mean to be estimated;

α_i is the fixed effect of the i-th area, where $\sum_{i=1}^{3} \alpha_i = 0$; and

e_{ij} is the sampling deviation associated with the j-th sample respondent in area i . The mean of the e_{ij} (for each i) is zero, and the variance, denoted σ_e^2 , is referred to as the sampling variance.

The response obtained from the j-th respondent by the k-th interviewer within the i-th area at trial t is denoted by Y_{ijkt} , and we assume that the model representation for these responses is

$$Y_{ijkt} = y_{ij} + \beta_{ik} + \gamma_t + \epsilon_{ijkt} \qquad (2)$$

where β_{ik} is the effect associated with the k-th interviewer within the i-th area;

γ_t is the fixed effect of the t-th trial, where $\sum_{t=1}^{2} \gamma_t = 0$; and

ϵ_{ijkt} is the response error associated with the j-th sample respondent who is interviewed by the k-th interviewer in the i-th area at trial t .

We assume that the β_{ik} and ϵ_{ijkt} are independently distributed with zero means and variances σ_β^2 and σ_ϵ^2 , respectively, and that these errors are uncorrelated with the true responses y_{ij} in equation 2. In the linear model (2), the interviewer effects, β_{ik}, are assumed additive, with zero mean and no interviewer-by-respondent interaction effect. If there is an overall bias associated with the interviewer effects and respondent response errors, it is confounded with the true value and y_{ij} is then interpreted as the "operationally true value" (see [2, p. 641]).

These assumptions may represent some simplification of the true situation. For some survey variates, the response errors and interviewer effects may be correlated with the true responses. For example, for a point binomial random variable, the response errors are negatively correlated with the true values (see [2, p. 643]). Further, the respondent response errors for different trials may be correlated, and their variance depend on the size of the farm operation. Research is required to develop models for satisfactory analysis of survey data having correlated and heteroscedastic response errors.

In the model representation, we refer to $(\beta_{ik} + \epsilon_{ijkt})$ as the total response error associated with the survey response Y_{ijkt}. The variance of the survey responses, denoted σ^2, is $(\sigma_e^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$, and the covariance between the responses from any two different respondents interviewed by the same interviewer, denoted $\rho_\beta \sigma^2$, is σ_β^2. It is easily verified that, under the model given by equation 2, the variance of the average of the responses obtained by the k-th sample interviewer in area i at trial t is $\sigma^2[1 + (n_{ik} - 1)\rho_\beta]/n_{ik}$, where n_{ik} denotes the number of respondents interviewed by interviewer k in area i. If interviewers have equal workloads (i.e., $n_{ik} = n$ for all i and k), then the factor by which the usual variance of the overall mean is increased is $[1 + (n-1)\rho_\beta]$. The estimation of ρ_β, the intra-interviewer correlation coefficient, is thus of interest. The estimation of the variance components σ_β^2, σ_ϵ^2 and σ_e^2 permits the estimation of the relative importance of interviewer effects, respondent response errors, and sampling deviations in the total variance of individual survey responses.

In our replicated survey, the interviewers in trial 1 participated in trial 2 a month later. In the analysis of the survey responses, we introduce a trial-by-interviewer interaction in our model. This term may not be negligible if the effect of "experience" is different for different interviewers. Thus, for the analysis of the observations obtained for each survey variate, we consider the mixed model

$$Y_{ijkt} = \mu + \alpha_i + \beta_{ik} + \gamma_t + (\beta\gamma)_{ikt} + e_{ij} + \epsilon_{ijkt} \ . \qquad (3)$$

If this linear model is expressed as a full-rank linear regression model, there are 282 independent variables. Because one interviewer in the first area dropped out after trial 1 and a new interviewer was obtained for trial 2, there are ten independent interviewer effects. There are nine independent trial-by-interviewer interaction effects and 259 independent respondent-within-area effects (e_{ij}) associated with the 262 survey respondents. For each survey variate, 524 survey responses are involved in the analysis.

4. ESTIMATION OF THE VARIANCE COMPONENTS

We present estimators for the variance components, σ_ϵ^2 , σ_β^2 and σ_e^2 , defined in section 3.

4.1 Variance Associated with Respondents

We estimate the respondent response error variance, σ_ϵ^2 , from thirty six analyses of variance on the data from the thirty six respondent groups. The analysis of variance for a respondent group with seven respondents estimates σ_ϵ^2 with six degrees of freedom (table 3). In this analysis of variance, the trial and interviewer effects are confounded because the trial-2 response is obtained by a different interivewer in the area. The weighted average of the thirty six analyses-of-variance error mean squares estimates σ_ϵ^2 with 226 degrees of freedom. These estimates for σ_ϵ^2 for each of the twenty one survey variates are given in column 2 of table 4.

4.2 Variance of Interviewer Effects

The averages of the responses for each respondent group for each trial, as the averages of the linear model in equation 3, form a linear model having fifty six independent variables,

$$\bar{Y}_{ipkt} = \mu + \alpha_i + \beta_{ik} + \gamma_t + (\beta\gamma)_{ikt} + \bar{e}_{ip} + \bar{\epsilon}_{ipkt} , \qquad (4)$$

where $\bar{Y}_{ipkt}$ is the average response of the p-th respondent group (p = 1, 2, ..., 12) assigned to interviewer k in the i-th area at trial t;

$\bar{e}_{ip}$ is the average of the sampling deviations, e_{ij} , for the respondents in the p-th respondent group in area i; and

$\bar{\epsilon}_{ipkt}$ is the average response error associated with $\bar{Y}_{ipkt}$.

To estimate the parameters in this linear model, we construct the full-rank, linear-regression model

$$\bar{Y}_{ipkt} = X_0 \mu + X_1 \alpha_1 + X_2 \alpha_2 + \sum_{s=1}^{4} Z_{1s} \beta_{1s} + \sum_{r=2}^{3} \sum_{s=1}^{3} Z_{rs} \beta_{rs} + X_3 \gamma_1 + \sum_{r=1}^{3} \sum_{s=1}^{3} Z_{rs} X_3 (\beta\gamma)_{rs1} + \sum_{r=1}^{3} \sum_{s=1}^{11} X_{rs} \bar{e}_{rs} + \bar{\epsilon}_{ipkt} , \qquad (5)$$

where $X_0 = 1$ for all i, p, k, t;

$X_1 = 1$ for $i = 1$

$= 0$ for $i = 2$

$= -1$ for $i = 3$;

$X_2 = 0$ for $i = 1$

$= 1$ for $i = 2$

$= -1$ for $i = 3$;

$Z_{1s} = 1$ if $i = 1, k = s$

$= -1$ if $i = 1, t = 2, k = 5$

$= 0$ otherwise, for $s = 1, 2, 3, 4$;[2/]

$Z_{rs} = 1$ if $i = r, k = s$

$= -1$ if $i = r, k = 4$

$= 0$ otherwise, for $s = 1, 2, 3$ and $r = 2, 3$;

$X_3 = 1$ for $t = 1$

$= -1$ for $t = 2$

[2/] There are four dummy variables associated with interviewer effects for the first area because of the replacement of one of the interviewers after the first-trial interviews were completed.

$$X_{rs} = 1 \text{ if } i = r, p = s$$
$$= -1 \text{ if } i = r, p = 12$$
$$= 0 \text{ otherwise, for } r = 1, 2, 3 \text{ and } s = 1, 2, \ldots, 11 .$$

The errors, $\bar{\epsilon}_{ipkt}$, in this regression model have different variances. That is, the variance of $\bar{\epsilon}_{ipkt}$ is σ_ϵ^2/n_{ip} , where n_{ip} denotes the number of respondents in the p-th respondent group within the i-th area. For estimation of the parameters in the linear model in equation 5, all variables are multiplied by the square root of the number of respondents in the given respondent group before performing the ordinary-least-squares regressions. We obtain the linear model

$$Y = Z\,\beta + X\,\alpha + \epsilon \tag{6}$$

where Y is the vector obtained by multiplying the mean responses $\bar{Y}_{ipkt}$ by the square root of the number of respondents in the given respondent group;

β is the column vector of ten independent interviewer effects;

α is the column vector of the remaining forty six parameters in the linear model in equation 5;

Z is the (72 x 10) matrix of transformed observations for the dummy variables corresponding to the interviewer effects;

X is the (72 x 46) matrix of transformed observations for the remaining dummy variables.

Under our model assumptions, the residual mean square for the linear model in equation 6 estimates σ_ϵ^2 with sixteen degrees of freedom. The ratios of these residual mean squares to the analysis-of-variance estimators for σ_ϵ^2 are in the "Test 1" column of table 7. The critical values for these test statistics are obtained from the $F(16, 226)$ distribution. Under the null hypothesis, the expectation of the test statistic is $226/224 \doteq 1.01$ if the respondent response errors are normally distributed.[3/] Since the average of the twenty one test ratios is 1.10, we conclude that the data and the model are consistent for the twenty one variates involved.

The regression sum of squares for the ordinary-least-squares regression of Y on Z and X is denoted $R(\beta, \alpha)$, and the regression sum of squares for the ordinary-least-squares regression of Y on X is denoted $R(\alpha)$. The difference in these regression sums of squares estimates a linear combination of the variance components σ_β^2 and σ_ϵ^2. It can be shown that (see [9, p. 54])

$$E\,[R(\beta, \alpha) - R(\alpha)] = \text{trace}\,\{[Z'\,MZ]\; E(\beta\beta')\} + \text{rank}\,(Z)\,\sigma_\epsilon^2\,, \tag{7}$$

where $M = I - X(X'X)^{-1}X'$.

For our full-rank reparameterization of model (4), the elements of the vector β in model (6) are differences between the original interviewer effects and the average of the interviewer effects for the given area. That is, the first four elements of β are $\beta_{1j} - \sum_{k=1}^{5} \beta_{1k}/5$,

---

[3/] The test statistic is the ratio of two independent mean squares which have the same expectations under the model assumptions. Given that the observations are normally distributed, the expectation of the ratio is thus $d/(d-2)$ where d denotes the degrees of freedom of the denominator mean square.

$j = 1, 2, 3, 4$; the remaining elements are $\beta_{ij} - \sum_{k=1}^{4} \beta_{ik}/4$, $i = 2, 3$ and $j = 1, 2, 3$. Thus, $E(\beta\, \beta') = \sigma_\beta^2 (I - J^*)$, where J^* is the block-diagonal matrix

$$J^* = \begin{pmatrix} J_4/5 & 0 & 0 \\ 0 & J_3/4 & 0 \\ 0 & 0 & J_3/4 \end{pmatrix}$$

where J_3 and J_4 are square matrices of order three and four, respectively, with all elements equal to one.

It follows that an unbiased estimator for the variance of the interviewer effects is

$$\hat{\sigma}_\beta^2 = \{[R(\beta, \alpha) - R(\alpha)] - 10\, \hat{\sigma}_\epsilon^2\}/tr[Z'\, M\, Z(I - J^*)] \ , \tag{8}$$

where $\hat{\sigma}_\epsilon^2$ is the analysis-of-variance estimator for the variance of the respondent response errors.

The matrix $Z'\, M\, Z$ is the sum of squares and products matrix of the residuals obtained by regressing each column of Z on the columns of X . The trace of $[Z'\, M Z(I - J^*)]$ was calculated to be $459.6 - 207.5 = 252.1$.

Under the assumption of zero interviewer effects and normality of the respondent response errors, the variance of the estimated variance of interviewer effects is

$$\begin{aligned} Var(\hat{\sigma}_\beta^2) &= [Var(\epsilon'\, A\, \epsilon) + 100\, Var(\hat{\sigma}_\epsilon^2)]/(252.1)^2 \\ &= [20\, \sigma_\epsilon^4 + 100(2\, \sigma_\epsilon^4/226)]/(252.1)^2 \\ &= \sigma_\epsilon^4/3{,}042.35 \ , \end{aligned} \tag{9}$$

where $A = M\, Z(Z'\, M\, Z)^{-1}\, Z'\, M$.

The values obtained for the estimator for σ_β^2 for the twenty one survey variates are presented in column 3 of table 4. Of the twenty one estimates, fourteen are positive.

4.3 Variance of Sampling Deviations

Unbiased estimators for σ_e^2 are obtained from two-fold-nested analyses of variance on the responses for the separate trials. By dropping the trial subscript, t, from the model representation (2) we obtain

$$Y_{ijk} = \mu + \alpha_i + \beta_{ik} + (e_{ij} + \epsilon_{ijk}) \tag{10}$$

where $j = 1, 2, \ldots, n_{ik}$; $k = 1, 2, 3, 4$; $i = 1, 2, 3$; and n_{ik} denotes the number of respondents in the assignment of the k-th interviewer within the i-th area. The error in this two-fold-nested-error model is the sum of the sampling deviation and the respondent response error. The two-fold-nested analysis of variance of table 6 is obtained from the responses in each trial of the survey.

In table 6, the factor K in the expectation of the interviewers-within-areas mean square is defined by

$$K = [\sum_{i=1}^{3} \sum_{k=1}^{4} n_{ik} - \sum_{i=1}^{3} (\sum_{k=1}^{4} n_{ik}^2 / \sum_{k=1}^{4} n_{ik})]/9 .$$

The respondents-within-interviewers mean square estimates the sum of the sampling variance and respondent response error variance. An unbiased estimator for the sampling variance is

$$\hat{\sigma}_e^2 = \sum_{t=1}^{2} M_{2t}/2 - \hat{\sigma}_\epsilon^2 \qquad (11)$$

where M_{2t} denotes the respondents-within-interviewers mean square for the analysis of variance of trial-t responses, t = 1, 2 . The estimates for σ_e^2 obtained from the estimator in equation 11 are given in column 4 of table 4.

The relative importance of respondent response errors, interviewer effects and sampling deviations is estimated by the ratios of the estimated variances to the sum of the estimated variances for each survey item. These ratios are expressed as percentages in table 5. Negative variance component estimates are reported and the average relative proportions over the twenty one survey variates obtained. These results indicate that, on the average, sampling variability accounts for approximately eighty percent of the total variance of survey responses, and that respondent response errors account for approximately twenty percent of the total variance. On the average, the interviewer effects account for about 0.04 of one percent of the total variance.

5. UNIVARIATE TESTS FOR INTERVIEWER EFFECTS

The estimate for the variance of interviewer effects is negative for seven of the twenty one survey variates considered (table 4). Of the estimates that are positive, the estimated variance of interviewer effects does not exceed 1.2 percent of the estimated total variance of the survey responses. To obtain a test for interviewer effects, we consider the linear model given by equation 6 and the regression model obtained by dropping the interviewer effects. For zero interviewer effects in the linear model in equation 6, we obtain the linear model

$$Y = X \alpha + \epsilon \quad . \tag{12}$$

The difference between the regression sums of squares for these two models, divided by ten, is divided by the analysis-of-variance estimator for the variance of the respondent response errors to obtain the test statistic for interviewer effects. This test statistic is calculated for each of the twenty one survey variates, and the results are presented in the "Test 2" column of table 7. The critical values for the test statistic are obtained from the $F(10, 226)$ distribution. The upper twenty five, ten, five and one percentage points for the $F(10, 226)$ distribution are approximately 1.27, 1.63, 1.87 and 2.41, respectively. Thus, for nine of the survey variates the interviewer effects are significant at the twenty-five-percent level, four are significant at the ten-percent level and one is significant at the one-percent level. At the one-percent level for the test, interviewer effects seem significant for the variate, the number of chickens on the farm. If the twenty-five-percent significance level for the test is considered appropriate, the variates for which there seem significant interviewer effects are acres operated, acres rented, acres of corn, idle acres, breeding hogs, expected farrowings during October-December, sheep and number of chickens on the farm.

To test the time and trial-by-interviewer-interaction effects for significance, we consider the linear model given by equation 6 and that obtained by excluding the columns in X which correspond to the trial and trial-by-interviewer interactions, γ_1 and $(\beta\gamma)_{rsl}$, in equation 5. The difference between the regression sums of squares for these two models, divided by ten, is divided by the analysis-of-variance estimator for the variance of the respondent response errors to obtain the test statistic for time and trial-by-interviewer-interaction effects. The values of this test statistic for the twenty one variates are presented in the "Test 3" column of table 7. The critical values for the test statistic are obtained from the $F(10, 226)$ distribution. The only survey variate having significant time and trial-by-interviewer interactions at the five-percent level is the number of acres of permanent pasture.

6. MULTIVARIATE TEST FOR INTERVIEWER EFFECTS

Because response errors associated with interviewers are of considerable interest, a multivariate statistic is obtained to test for the presence of significant interviewer effects in the twenty one survey variates considered. We obtain the Wilks' statistic, which is defined

$$U = |\hat{\Sigma}_{\Omega}|/|\hat{\Sigma}_{\omega}| \tag{13}$$

where $\hat{\Sigma}_{\Omega}$ is the estimated covariance matrix for the respondent response errors; and

$\hat{\Sigma}_{\omega}$ is the estimated covariance matrix for the response errors under the assumption that the interviewer effects are zero.

The covariance matrix $\hat{\Sigma}_{\Omega}$ was obtained from the residuals in the thirty six analyses of variance for the respondent groups. For each of the twenty one survey variates, 524 residuals were obtained from the analyses of variance used to estimate the variance of the respondent response errors. The sums-of-squares and products matrix obtained with the residuals is divided by 226, the degrees of freedom for the respondent response error. The estimated correlations between the respondent response errors are presented in table 8. The covariance matrix $\hat{\Sigma}_{\Omega}$ is obtained from table 8 and the estimates of the variances of the respondent response errors (table 4).

To obtain the covariance matrix $\hat{\Sigma}_{\omega}$, the linear models in equations 6 and 9 are used. The sums-of-squares and products matrix obtained with the residuals from the linear model in equation 6 is denoted by S_1 . The sums-of-squares and products matrix obtained with the residuals from the linear model in equation 12 is denoted by S_2 . The covariance matrix for the response errors under the assumption of zero interviewer effects is thus estimated by

$$\hat{\Sigma}_{\omega} = [(S_2 - S_1) + 226 \ \hat{\Sigma}_{\Omega}]/236 \ . \qquad (14)$$

By a result stated in Kramer and Jensen [7, p. 268], the distribution of $-220 \log_e U$ is approximately chi-square with 210 degrees of freedom. The multiple of $-\log_e U$ is $226 - (21 - 10 + 1)/2 = 220$, where 226 is the degrees of freedom for estimation of Σ_{Ω} , 21 is the number of variates involved in the multivariate analysis, and 10 is the number of independent interviewer effects. The degrees of freedom, 210, is the product of the number of variates and the number of independent interviewer effects. The value of U for the survey data is 0.7622 . The test statistic thus has the value $-220 \log_e 0.7622 = 59.75$. This value is not significant in relation to upper critical values for the chi-square distribution with 210 degrees of freedom. The Wilks' test does not indicate that the interviewer effects were significant in the survey.

7. CONCLUSIONS

For the variates analyzed from this survey, interviewer effects do not have an important influence on the total variability of the survey responses. The average of the estimated intra-interviewer correlation coefficients for the twenty one survey variates is 0.0004. Thus, for workloads of twenty five and fifty respondents per interviewer, the traditional variance of average responses is increased by only 0.96 percent and 1.96 percent, respectively.[4/]

For zero interviewer effects, the variance of average responses from the survey is thus the total variance of individual responses divided by the appropriate factor. The variance of respondent response errors, however, is a significant proportion of the total variance of individual responses for most of the survey variates. On the average, the proportion of the total variance due to respondent response errors is estimated at twenty percent.

[4/] If there are c acres and each area has k interviewers in the sample such that each interviewer has n respondents, then the variance of the average response in the sample is $\sigma^2(1 + (n-1)\rho)/c\,k\,n$, where ρ is the intra-interviewer correlation coefficient and σ^2 is the total variance of the observation (see section 3). For $n = 25$ and $\rho = 0.0004$, the factor $1 + (n-1)\rho$ is equal to $1 + (25-1)\,0.0004 = 1.0096$.

For the land items, respondent response errors seem important in such items as number of acres rented from others, number of acres of permanent pasture, number of acres of hay and number of acres of idle land. Imprecise definition of these items contribute to the response errors. For example, the distinction often is unclear between acres of permanent pasture and acres for hay.

The respondent errors are a particularly important source of variability for sow farrowings in the livestock items. As expected, the respondent response errors are more prominent in the variates number of sows expected to farrow before October 1 and number of sows expected to farrow during October-December. These variates involve responses to expected number to farrow. Trial 2 of the survey was after October 1, however, and the response corresponding to the trial-1 variate, number of sows expected to farrow before October 1, was number of sows which actually farrowed between the trial-1 interview and October 1. Further, the response for the variate, number of sows expected to farrow during October-December, was negatively correlated with the number of sows farrowing before October 1.

The two labor items in the questionnaires clearly indicate that respondent response errors make significant contributions to the variance of survey responses. Poor respondent recall and the ambiguity of the questions may be reasons for these large variances for the respondent response error.

Although the data in table 5 agree, for the most part, with prior expectations, they suggest areas which deserve attention in questionnaire construction for agricultural surveys.

REFERENCES

[1] Bailar, Barbara A., "Recent Research in Reinterview Procedures," Journal of the American Statistical Association 63 (March, 1968), 41-63.

[2] Cochran, W. G., "Errors of Measurement in Statistics," Technometrics 10 (November, 1968), 637-666.

[3] Fellegi, I. P., "Response Variance and its Estimation," Journal of the American Statistical Association 59 (December, 1964), 1016-1041.

[4] Hansen, M. H., Hurwitz, W. N., and Bershad, M. A., "Measurement Errors in Censuses and Surveys," Bulletin of the International Statistical Institute 38 (1961), 359-374.

[5] Hansen, M. H., Hurwitz, W. N., Marks, E. S., and Mauldin, W. P., "Response Errors in Surveys," Journal of the American Statistical Association 46 (June, 1951), 147-190.

[6] Hanson, R. H., and Marks, E. S., "Influence of the Interviewers on the Accuracy of Survey Results," Journal of the American Statistical Association 53 (September, 1958), 635-655.

[7] Kramer, C. Y., and Jensen, D. R., "Fundamentals of Multivariate Analysis, Part III. Analysis of Variance for One-Way Classifications," Journal of Quality Technology 1 (October, 1969), 264-276.

[8] Kish, L., "Studies of Interviewer Variance for Attitudinal Variables," Journal of the American Statistical Association 57 (March, 1962), 92-115.

[9] Searle, S. R., "Topics in Variance Component Estimation," Biometrics 27 (March, 1971), 1-76.

[10] Stock, J. S., and Hochstim, J. R., "A Method of Measuring Interviewer Variability," Public Opinion Quarterly 15 (Summer, 1951), 322-334.

[11] Sukhatme, P. V., and Seth, G. R., "Non-Sampling Errors in Surveys," Journal of the Indian Society of Agricultural Statistics 4 (1952), 5-41.

[12] U.S. Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960, Effects of Interviewers and Crew Leaders, Series E R 60 No. 7, Washington, D. C. (February, 1968).

APPENDIX A. COMPARISON WITH A BUREAU OF THE CENSUS STUDY

In a U.S. Bureau of the Census study [12], estimates for the variance of interviewer effects were obtained from an interpenetrating sample survey. For each cluster in the survey, there were two interviewers, each being assigned a random sample from the population. Only one response was obtained from each respondent in the sample. The model for that study [12, pp. 13-14] can be written as the components-of-variance model

$$Y_{ijk} = \mu + \alpha_i + \beta_{ik} + e_{ij} + \epsilon_{ijk} \tag{A.1}$$

where the notation here is that of section 3, except that the subscript denoting trial, t, is dropped. The random variables β_{ik}, e_{ij} and ϵ_{ijk} are assumed uncorrelated. The correlations introduced by sampling without replacement from a finite population are ignored. In the notation of the Bureau of the Census, the response error for the j-th respondent in the i-th area, $d_{ijk} \equiv Y_{ijk} - E(Y_{ijk} \mid i, j)$ is $(\beta_{ik} + \epsilon_{ijk})$. The total response variance σ_d^2 is thus $\sigma_\beta^2 + \sigma_\epsilon^2$, and the covariance between the response deviations for a given interviewer $\rho_d \sigma_d^2$ is σ_β^2. The sampling deviation associated with Y_{ijk},

$\Delta_{ij} \equiv E(Y_{ijk} \mid i, j) - E(Y_{ijk} \mid i)$, is e_{ij}.

The variance of interviewer effects is estimated from the two-fold-nested analysis of variance of the survey responses (table 6). The weighted average of the cluster estimators, $(C_i - D_i)$, obtained by the Bureau of the Census [12, p. 14] is the estimator

$$\tilde{\sigma}_\beta^{\;2} = (M_1 - M_2)/K \qquad (A.2)$$

where M_1 is the interviewers-within-areas mean square;

M_2 is the respondent-within-interviewers mean square; and

$$K = \sum_{i=1}^{3} \left\{ \sum_{k=1}^{4} n_{ik} - \sum_{k=1}^{4} n_{ik}^{\;2} / \sum_{k=1}^{4} n_{ik} \right\} / 9 \;.$$

From the trial-1 and trial-2 responses in our survey, the estimates obtained for the variance of interviewer effects obtained from $\tilde{\sigma}_\beta^{\;2}$ are presented in table 9.

To compare the variance of the $\hat{\sigma}_\beta^{\;2}$, defined by equation 9, with that of $\tilde{\sigma}_\beta^{\;2}$, defined by equation (A.2), we assume that the estimator $\tilde{\sigma}_\beta^{\;2}$ is obtained from a survey in which each of twelve interviewers interviewed forty four respondents. The total number of responses from such a survey is 528, which exceeds the total number from our survey by four.

Under the assumptions that the errors $(e_{ij} + \epsilon_{ijk})$ are normally distributed and that $\sigma_\beta^{\;2} = 0$, the variance of the estimator $\tilde{\sigma}_\beta^{\;2}$ is

$$\begin{aligned} Var(\tilde{\sigma}_\beta^{\;2}) &= Var\,[(M_1 - M_2)/\, n] \\ &= [2(\sigma_e^{\;2} + \sigma_\epsilon^{\;2})^2\;(12\,n - 3)]/\;108\;n^2(n-1) \\ &= (\sigma_e^{\;2} + \sigma_\epsilon^{\;2})^2/\;8{,}562.65 \qquad (16) \end{aligned}$$

for $n = 44$.

If we denote $\sigma_e^{\;2} + \sigma_\epsilon^{\;2} = \sigma^2$ and $\sigma_\epsilon^{\;2} = p\,\sigma^2$, then it is easily verified that

$$\mathrm{Var}(\hat{\sigma}_\beta^{\,2}) \leq \mathrm{Var}(\tilde{\sigma}_\beta^{\,2}) \quad \text{if} \quad p \leq 0.60 \quad .$$

That is, if the respondent errors account for not more than sixty percent of the total variance, assuming interviewer effects are zero, the variance of the estimator for the variance of interviewer effects based on replicated responses is less than that for unreplicated responses under our model assumptions and sample size. The results in table 5 indicate that this condition is satisfied for all twenty one survey variates analyzed.

APPENDIX B. TABLES 1-9 OF THE REPORT

Table 1. Numbers of respondents in the respondent groups

Respondent group within area	Area 1	Area 2	Area 3	
1	8	7	7	
2	7	6	8	
3	8	8	6	
4	7	7	7	
5	8	7	7	
6	8	7	8	
7	7	6	9	
8	7	7	8	
9	7	6	8	
10	6	8	7	
11	6	8	8	
12	7	8	8	
Totals	86	85	91	262

Table 2. Description of the Survey Variates Analyzed

Variate Description
Total number of acres operated
Number of acres rented from or worked on shares for others
Number of acres of corn
Number of acres of soybeans
Number of acres of permanent pasture
Number of acres of hay
Number of idle acres
Number of cattle and calves on the farm
Number of breeding hogs on the farm
Number of sows farrowed on the farm during March-May, 1970
Number of sows farrowed on the farm during June-August, 1970
Number of sows and gilts expected to farrow on the farm before October 1, 1970
Number of sows and gilts expected to farrow on the farm during October-December, 1970
Number of sheep on the farm
Are there chickens on the farm? (Yes = 1, No = 0)
Number of chickens on the farm
Number of non-family hired workers on the farm during the week of August 23-29, 1970
Number of hired workers on the farm during 1969
Total value of sales in 1969 (eleven classes with responses coded as 0 through 10)
The most important agricultural product, in terms of percent of total sales, sold from the farm in 1969 (eight enterprises coded as 1 through 8)
Percentage of total sales from the most important agricultural product in 1969

Table 3. Analysis of variance for a respondent group having 7 respondents

Source of variation	d.f.	E(M.S.)
Trial (+ Interviewer)	2 - 1	
Respondents	7 - 1	
Error	6	σ_{ϵ}^2
Total	(2 x 7) - 1	

Table 4. Estimates for response error variances ($\hat{\sigma}_\epsilon^2$), variances of interviewer effects ($\hat{\sigma}_\beta^2$), and variance of sampling deviations ($\hat{\sigma}_e^2$).

Variate	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_\beta^2$	$\hat{\sigma}_e^2$
Acres operated	496.75	11.73	32,685.39
Acres rented	1,364.47	14.71	25,505.03
Acres of corn	131.04	3.10	11,334.54
Acres of soybeans	89.30	0.45	3,133.29
Acres of permanent pasture	439.92	0.09	2,088.73
Acres of hay	295.15	2.63	312.59
Idle acres	263.51	6.55	322.04
Cattle and calves	1,079.20	-10.26	50,687.18
Breeding hogs	216.76	4.20	851.23
March-May farrowings	60.42	-1.37	279.91
June-August farrowings	33.73	-0.44	168.59
Expected farrowings before October 1	27.94	0.27	50.29
Expected October-December farrowings	40.85	0.93	110.51
Sheep	23.48	0.59	1,049.23
Chickens (Yes, No)	0.03	0.00	0.21
Number of chickens	2,859.68	216.73	272,481.55
Non-family workers, August 23-29	0.99	-0.01	0.88
Non-family workers, 1969	5.82	-0.02	7.87
Total sales	0.28	0.01	3.15
Most important product	0.68	-0.02	1.19
Percentage of sales from most important product	101.64	-0.96	273.33

Table 6. Analysis of variance for the responses in a given trial of the survey

Source of Variation	d.f.	S.S.	M.S.	E(M.S.)
Areas	3 - 1	$\sum_{i=1}^{3} \sum_{k=1}^{4} n_{ik} (\overline{Y}_{i..} - \overline{Y}_{...})^2$		
Interviewers/Areas	3(4 - 1)	$\sum_{i=1}^{3} \sum_{k=1}^{4} n_{ik} (\overline{Y}_{i.k} - \overline{Y}_{i..})^2$	M_1	$\sigma_e^2 + \sigma_\epsilon^2 + K \sigma_\beta^2$
Respondents/Interviewers	262 - 12	$\sum_{i=1}^{3} \sum_{k=1}^{4} \sum_{j=1}^{n_{ik}} (Y_{ijk} - \overline{Y}_{i.k})^2$	M_2	$\sigma_e^2 + \sigma_\epsilon^2$
Total	262 - 1	$\sum_{i=1}^{3} \sum_{k=1}^{4} \sum_{j=1}^{n_{ik}} (Y_{ijk} - \overline{Y}_{...})^2$		

Table 7. Univariate tests of significance based on the analysis-of-variance estimates for respondent response error variances.

Variate	Test 1[a]	Test 2[b]	Test 3[c]
Acres operated	0.58	1.60	0.84
Acres rented	2.10	1.27	1.16
Acres of corn	1.24	1.60	0.80
Acres of soybeans	0.43	1.13	1.01
Acres of permanent pasture	0.96	1.01	2.02
Acres of hay	1.02	1.22	1.35
Idle acres	1.41	1.63	1.35
Cattle and calves	0.94	0.76	1.60
Breeding hogs	0.74	1.49	1.05
March-May farrowings	1.13	0.43	1.41
June-August farrowings	0.79	0.67	0.84
Expected farrowings before October 1	0.86	1.25	1.70
Expected October-December farrowings	0.81	1.58	1.28
Sheep	1.26	1.63	0.64
Chickens (Yes, No)	1.41	1.14	0.79
Number of chickens	1.59	2.91	1.21
Non-family workers, August 23-29	1.27	0.78	0.79
Non-family workers, 1969	1.15	0.92	0.34
Total sales	1.39	1.84	1.36
Most important product	1.05	0.27	0.57
Percentage of sales from most important product	0.90	0.76	1.26
Average	1.10	1.23	1.11
F-distribution for tests	F(16, 226)	F(10, 226)	F(10, 226)

[a] Test-statistics for testing residual mean square for the model based on mean responses.

[b] Test-statistics for significance of interviewer effects.

[c] Test-statistics for significance of time and trial-by-interviewer interaction effects.

Table 8. Estimated correlation matrix of respondent response errors for the 21 survey variates.[a]

	1	2	3	4	5	6	7	8	9	10	11
1	1.00										
2	0.29	1.00									
3	0.41	-0.10	1.00								
4	0.25	0.16	0.22	1.00							
5	0.42	0.21	-0.11	-0.05	1.00						
6	-0.02	-0.00	0.21	-0.19	-0.17	1.00					
7	0.24	0.08	-0.13	0.00	0.04	-0.69	1.00				
8	0.15	0.12	0.03	0.09	0.18	-0.03	-0.01	1.00			
9	-0.16	-0.05	-0.23	0.04	-0.09	-0.06	0.05	-0.05	1.00		
10	0.01	0.01	0.07	0.07	-0.06	0.01	0.01	0.02	0.18	1.00	
11	-0.16	-0.01	0.03	-0.03	-0.06	0.06	-0.09	-0.06	0.07	-0.27	1.00
12	0.07	0.01	0.01	0.22	0.09	-0.09	-0.07	0.17	0.01	0.12	-0.27
13	-0.06	0.03	-0.04	-0.03	-0.09	-0.04	0.00	-0.04	-0.07	-0.01	0.16
14	0.05	0.02	0.02	-0.01	-0.02	0.01	0.00	-0.02	-0.04	-0.01	0.09
15	0.15	0.13	0.11	0.02	0.03	0.02	0.04	-0.03	-0.03	-0.05	0.09
16	0.05	0.07	-0.05	0.02	0.05	-0.04	0.05	0.01	0.03	0.03	-0.00
17	-0.10	-0.14	0.05	-0.06	-0.02	0.23	-0.31	0.16	-0.00	0.00	0.08
18	-0.01	0.03	-0.02	-0.15	0.07	-0.13	0.08	-0.04	0.16	-0.00	0.08
19	-0.02	-0.02	0.17	-0.02	-0.10	0.10	-0.03	-0.06	-0.30	-0.08	0.06
20	0.07	-0.01	0.04	-0.00	0.03	-0.04	0.00	0.01	0.02	-0.02	-0.02
21	-0.07	-0.02	-0.15	-0.10	0.06	-0.04	0.03	0.01	0.04	-0.01	0.04

[a] The 21 survey variates are listed in the same order as given in Table 2.

Table 8. (cont.)

	12	13	14	15	16	17	18	19	20	21
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12	1.00									
13	-0.45	1.00								
14	-0.03	0.02	1.00							
15	-0.09	0.06	0.01	1.00						
16	-0.08	0.05	-0.01	0.08	1.00					
17	-0.02	0.04	-0.03	-0.09	0.04	1.00				
18	-0.01	0.11	0.04	0.08	0.13	0.03	1.00			
19	-0.01	0.08	0.09	0.14	0.13	0.04	-0.10	1.00		
20	0.07	-0.11	0.06	-0.08	-0.06	0.03	0.06	0.03	1.00	
21	0.05	-0.04	-0.06	0.02	0.03	0.03	0.01	0.04	0.16	1.00

Table 9. Analysis-of-variance estimates $(\tilde{\sigma}_{\beta}^{2})$ for the variance of the interviewer effects from the trial 1 and trial 2 responses

Variate	Trial 1	Trial 2
Acres operated	-3.11	349.49
Acres rented	616.54	-302.77
Acres of corn	-66.92	161.57
Acres of soybeans	66.68	3.30
Acres of permanent pasture	59.81	-56.77
Acres of hay	-14.22	-4.36
Idle acres	11.20	11.79
Cattle and calves	-1,000.50	283.58
Breeding hogs	-6.83	21.01
March-May farrowings	-13.32	-2.92
June-August farrowings	1.05	-0.75
Expected farrowings before October 1	-2.67	0.72
Expected October-December farrowings	0.29	-0.79
Sheep	-15.05	-5.96
Chickens (Yes, No)	-0.00	0.01
Number of chickens	-2,672.72	-2,164.06
Non-family workers, August 23-29	-0.02	0.05
Non-family workers, 1969	-0.43	-0.03
Total sales	0.03	0.04
Most important product	-0.04	-0.02
Percentage of sales from most important product	3.55	-14.29