Estimating Sampling Error for Cluster Sample Travel Surveys by Replicated Subsampling

DON L. OCHOA AND GEORGE M. RAMSEY

The California Department of Transportation conducted in-person home interview travel surveys in six counties of the state before converting to the telephone survey technique in 1979. The surveys updated existing data bases that support the development of regional travel forecasting models. During the survey period cluster sampling was employed to minimize travel time for survey interviewers and facilitate call-back procedures. Because cluster sampling was used, the simple random sample model often cited [$s/(n^{1/2})$] was not appropriate for estimating sampling error because that formula tends to underestimate actual standard errors. Estimates of sampling error for the surveys were thus made using the method of "replicated subsampling," which takes sample clustering into account and yields a higher total standard error than does the conventional method. This paper is intended to illustrate application of replicated subsampling in estimating sampling error for cluster sample travel surveys. Comparisons of standard errors derived using the method of replicated subsampling are made with standard errors derived by the conventional formula, which assumes a simple random sampling design. Replicated subsampling provides an unbiased, reliable, and generally applicable framework for estimating sampling error.

The California Department of Transportation (Caltrans) conducted in-person home interview travel surveys in the counties of Fresno, Kern, Sacramento, San Diego, San Joaquin, and Stanislaus before converting to the telephone survey technique in 1979. The six regional travel surveys, conducted in 1977 and 1978, updated data bases that support the development of regional travel forecasting models and augmented the data base of California's more extensive 1976–1980 Statewide Travel Survey. Travel survey findings that were previously reported *(1)* will not be discussed; rather, application of W. Edwards Deming's method (*2*, pp.87–101) of "replicated subsampling" for estimating sampling error, particularly for cluster sample travel surveys, will be demonstrated. The conventional standard error formula [$s/(n^{1/2})$] is not appropriate to use on cluster samples because it assumes a simple random sampling design and usually underestimates actual standard errors.

SAMPLE SELECTION

Because of budget and time constraints for conducting the surveys, only 500 households were sampled in each of the survey regions, except in the San Diego area where 1,000 households were sampled. The larger sample size in the San Diego area was in recognition of the complexity of transportation problems and the need for more highly stratified information in that region. A brief discussion of the sample selection process follows.

Sample selection for each of the surveys involved a three-stage process. In the first stage, 25 census tracts were systematically selected (except in the San Diego area where 50 tracts were selected). From a random start, the census tract containing every *n*th housing unit was selected. [Because the skip interval *(n)* was based on the number of housing units in a region, the skip interval varied by region.]

In the second stage, five census blocks within each selected census tract were systematically chosen. From a random start, the block with every *k*th housing unit was selected. This skip interval was based on the number of housing units in a census tract and varied by census tract.

Finally, at the third stage, 16 housing units within each of the blocks selected in the second stage were enumerated in the field by employing a uniform enumeration and systematic sample selection procedure. From a random start, every fourth housing unit among the 16 listed in the block was systematically selected to be interviewed.

The uniform enumeration procedure involved starting at the northwestern corner of the block selected, proceeding in a clockwise direction around the block, and listing housing units on the right side of the street in the direction of travel until 16 housing units were listed (Figures 1 and 2). Note that the housing units

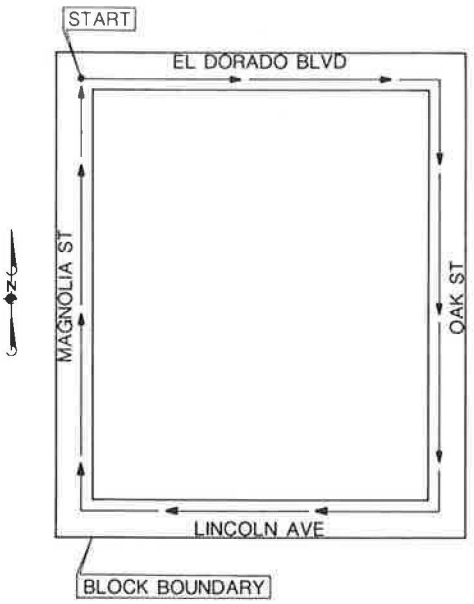

FIGURE 1 Field enumeration procedure showing both sides of the street and indicating the lister's starting place and direction of travel within a selected rectangular census block. The lister starts at the northeastern corner of the block and proceeds clockwise tallying housing units on the right side.

Division of Transportation Planning, California Department of Transportation, 1120 N Street, Sacramento, Calif. 95814.

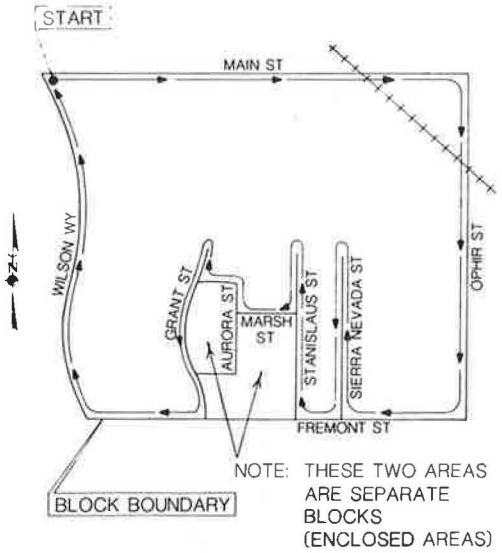

FIGURE 2 Field enumeration procedure showing right side of the street only and indicating the lister's starting place and direction of travel within a nonrectangular census block.

selected do not represent an equal proportion of the units by block; they do, however, meet the requirements of attaining a minimum of 500 samples for each survey region. (Statistical weights were applied to compensate for nonproportional samples when needed for survey data summaries.)

CLUSTER SAMPLING

During the survey period, cluster sampling was employed to minimize travel time for survey interviewers and facilitate calling back. The conventional standard error formula $[s/(n^{1/2})]$ is not appropriate in cluster sampling situations because variances of estimates derived from cluster samples tend to be greater than those derived from simple random samples (or systematic random samples) of the same size. As pointed out by Hubert M. Blalock, "[For cluster sampling] . . . the simple random sample formula will underestimate the true error" (*3*, p.527).

Guidelines for Designing Travel Surveys for Statewide Transportation Planning (*4*, p.5.12) suggests the use of the "design effect" for estimating standard errors of statistics acquired through cluster sample travel surveys. Leslie Kish who initially described the design effect as a means of accounting for the effects of clustering is quoted: "[T]he ratio of the actual variance [of a cluster or other complex sample] to the variance of a simple random sample of the same number of elements" (*4*, p.5.12). However, the design effect factors for the 1977–1978 surveys could not be determined because of lack of comparable data from simple random sampling.

Because multistage cluster sampling was employed in the six regional surveys, estimates of sampling error for those surveys were made using Deming's method of replicated subsamples. This method takes sample clustering into consideration and usually yields a higher (more conservative or safer) total standard error than does the conventional method.

Herbert Arkin and Raymond Colton define "standard error" as follows: "The standard deviation of a sampling distribution of means, or any other statistical measure computed from samples, is termed the standard error of the mean . . . or the standard error of the other statistical measure" (*5*, p.144).

Deming points out (*2*, p.87), "The distinguishing feature of the [replicated subsampling] design is . . . subsamples, drawn and processed completely independent of each other. The chief advantage of replication is ease in the estimation of the standard errors."

STATISTICAL RELIABILITY OF KEY SURVEY ESTIMATES

It should be noted that the particular variables presented in this paper are not intended to be all-inclusive. They are provided simply to illustrate application of the method of replicated subsampling to determine standard errors from cluster sample travel surveys.

Reliability estimates for the cluster sample surveys are presented for three variables by survey region and type of housing unit—persons per household, vehicles per household, and weekday person trips per household. The confidence intervals given in Tables 1–6 represent ranges of estimated sampling error at both the 90 percent and 95 percent confidence levels. (It should be kept in mind that errors occur whether a sample or a complete enumeration is used and that nonsampling errors are not taken into account when presenting statistical reliability estimates. Strict quality control procedures, of course, are required to minimize errors.)

TABLE 1 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, FRESNO REGION.

	SINGLE HOUSING UNITS					
	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	3.12		1.91		9.73	
Standard Error of the Mean[a]	0.088		0.062		0.481	
Confidence Interval	90%[a] ±0.145	95%[b] ±0.172	90%[a] ±0.102	95%[b] ±0.122	90%[a] ±0.794	95%[b] ±0.943

TABLE 1 *continued*

MULTIPLE HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.09		0.92		5.30	
Standard Error of the Mean[a]	0.143		0.121		0.784	
Confidence Interval	90%[a] ±0.236	95%[b] ±0.280	90%[a] ±0.200	95%[b] ±0.237	90%[a] ±1.294	95%[b] ±1.537

TOTAL HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.83		1.63		8.49	
Standard Error of the Mean[a]	0.126		0.102		0.553	
Confidence Interval	90%[a] ±0.208	95%[b] ±0.247	90%[a] ±0.168	95%[b] ±0.200	90%[a] ±0.912	95%[b] ±1.084

Note: Derived by replicated subsamples.

[a] ± 1.65 times the standard error of the mean.

[b] ± 1.96 times the standard error of the mean.

TABLE 2 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, KERN REGION

SINGLE HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	3.11		1.85		10.33	
Standard Error of the Mean[a]	0.242		0.045		0.484	
Confidence Interval	90%[a] ±0.399	95%[b] ±0.474	90%[a] ±0.074	95%[b] ±0.088	90%[a] ±0.799	95%[b] ±0.949

MULTIPLE HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.46		1.32		8.12	
Standard Error of the Mean[a]	0.119		0.118		0.657	
Confidence Interval	90%[a] ±0.196	95%[b] ±0.233	90%[a] ±0.195	95%[b] ±0.231	90%[a] ±1.084	95%[b] ±1.288

TOTAL HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.94		1.72		9.77	
Standard Error of the Mean[a]	0.077		0.053		0.525	
Confidence Interval	90%[a] ±0.127	95%[b] ±0.151	90%[a] ±0.087	95%[b] ±0.104	90%[a] ±0.866	95%[b] ±1.029

Note: Derived by replicated subsamples.

[a] ± 1.65 times the standard error of the mean.

[b] ± 1.96 times the standard error of the mean.

TABLE 3 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SACRAMENTO REGION

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
SINGLE HOUSING UNITS						
Mean	3.02		1.86		11.40	
Standard Error of the Mean[a]	0.106		0.095		0.739	
Confidence Interval	90%[a] ±0.175	95%[b] ±0.208	90%[a] ±0.157	95%[b] ±0.186	90%[a] ±1.219	95%[b] ±1.448
MULTIPLE HOUSING UNITS						
Mean	2.01		1.05		5.82	
Standard Error of the Mean[a]	0.176		0.090		0.834	
Confidence Interval	90%[a] ±0.290	95%[b] ±0.345	90%[a] ±0.148	95%[b] ±0.176	90%[a] ±1.376	95%[b] ±1.635
TOTAL HOUSING UNITS						
Mean	2.65		1.56		9.34	
Standard Error of the Mean[a]	0.071		0.063		0.648	
Confidence Interval	90%[a] ±0.117	95%[b] ±0.139	90%[a] ±0.104	95%[b] ±0.123	90%[a] ±0.069	95%[b] ±1.270

Note: Derived by replicated subsamples.

[a] ± 1.65 times the standard error of the mean.

[b] ± 1.96 times the standard error of the mean.

TABLE 4 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SAN DIEGO REGION

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
SINGLE HOUSING UNITS						
Mean	3.29		2.00		11.88	
Standard Error of the Mean[a]	0.041		0.037		0.267	
Confidence Interval	90%[a] ±0.068	95%[b] ±0.080	90%[a] ±0.061	95%[b] ±0.073	90%[a] ±0.441	95%[b] ±0.523
MULTIPLE HOUSING UNITS						
Mean	2.10		1.19		6.87	
Standard Error of the Mean[a]	0.092		0.053		0.274	
Confidence Interval	90%[a] ±0.152	95%[b] ±0.180	90%[a] ±0.087	95%[b] ±0.104	90%[a] ±0.452	95%[b] ±0.537

TABLE 4 *continued*

TOTAL HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.84		1.69		9.97	
Standard Error of the Mean[a]	0.049		0.038		0.226	
Confidence Interval	90%[a] ±0.081	95%[b] ±0.096	90%[a] ±0.063	95%[b] ±0.074	90%[a] ±0.373	95%[b] ±0.443

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

METHOD OF REPLICATED SUBSAMPLING

Briefly, the method of replicated subsampling (henceforth subsampling) is applied by examining estimates of a particular statistic derived from subsamples designed into the original survey sample. To find the standard error of a statistic derived from cluster samples, the lowest subsample mean value is subtracted from the highest mean value and divided by the number of subsamples compared (Table 7). The resulting number is an unbiased and reliable estimate of the standard error of the sample. To find its 90 percent or 95 percent confidence interval, the standard error is multiplied by the confidence factor 1.65 or 1.96, respectively.

TABLE 5 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SAN JOAQUIN REGION

SINGLE HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.96		1.88		9.82	
Standard Error of the Mean[a]	0.056		0.076		0.187	
Confidence Interval	90%[a] ±0.092	95%[b] ±0.110	90%[a] ±0.125	95%[b] ±0.149	90%[a] ±0.309	95%[b] ±0.367

MULTIPLE HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.25		1.08		6.02	
Standard Error of the Mean[a]	0.275		0.116		1.400	
Confidence Interval	90%[a] ±0.454	95%[b] ±0.539	90%[a] ±0.191	95%[b] ±0.227	90%[a] ±2.310	95%[b] ±2.744

TOTAL HOUSING UNITS

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.75		1.64		8.70	
Standard Error of the Mean[a]	0.092		0.101		0.500	
Confidence Interval	90%[a] ±0.152	95%[b] ±0.180	90%[a] ±0.167	95%[b] ±0.198	90%[a] ±0.825	95%[b] ±0.980

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

TABLE 6 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, STANISLAUS REGION

	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
SINGLE HOUSING UNITS						
Mean	3.11		1.81		9.40	
Standard Error of the Mean[a]	0.140		0.068		0.406	
Confidence Interval	90%[a] ±0.231	95%[b] ±0.274	90%[a] ±0.112	95%[b] ±0.133	90%[a] ±0.670	95%[b] ±0.796
MULTIPLE HOUSING UNITS						
Mean	2.03		1.07		5.79	
Standard Error of the Mean[a]	0.211		0.111		0.981	
Confidence Interval	90%[a] ±0.348	95%[b] ±0.414	90%[a] ±0.183	95%[b] ±0.218	90%[a] ±1.619	95%[b] ±1.923
TOTAL HOUSING UNITS						
Mean	2.85		1.63		8.53	
Standard Error of the Mean[a]	0.106		0.071		0.346	
Confidence Interval	90%[a] ±0.175	95%[b] ±0.208	90%[a] ±0.117	95%[b] ±0.139	90%[a] ±0.571	95%[b] ±0.678

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

TABLE 7 WEEKDAY MEAN PERSON TRIPS PER HOUSEHOLD BY CENSUS TRACT IN SACRAMENTO REGION

Census Tract Subsample	Weekday Person Trips per Household	Census Tract Subsample	Weekday Person Trips per Household
1	7.60	14	9.58
2	2.41	15	5.59
3	8.25	16	10.60
4	14.73	17	18.61
5	6.30	18	8.41
6	8.10	19	12.01
7	8.15	20	9.32
8	9.17	21	11.99
9	6.96	22	16.31
10	9.84	23	11.10
11	9.87	24	10.96
12	7.95	25	9.62
13	10.20		

Note: Scored numbers are lowest and highest mean values.

For these surveys, the subsamples to be considered are the census tracts selected for sampling. So, to estimate the standard error of a survey statistic, the mean value of the statistic was computed for each census tract subsample. Examination of the census tract means yields the range of the statistic.

For each survey region, except the San Diego region, 20 households (five blocks per census tract and four housing units per block) were sampled in each of the 25 census tracts. In the case of San Diego, 20 households were sampled in each of 50 census tracts.

To estimate the standard error of weekday person trips per household, means were obtained for each of the census tracts in a region. For example, census tract means for the Sacramento region were as given in Table 7. The range of means was found to be 18.61 − 2.41 = 16.20. therefore dividing the range by the number of subsamples yields the estimate of the standard error of person trips per household (16.20/25 = 0.648).

Subsample means were computed for each census tract within each of the six surveys. Table 8 gives a comparison of the standard

TABLE 8 COMPARISON OF STANDARD ERRORS DERIVED BY REPLICATED SUBSAMPLING AND BY THE CONVENTIONAL STANDARD ERROR FORMULA [$s/(n^{1/2})$]

Region	Method	Standard Errors for: Persons per Household	Vehicles per Household	Person Trips per Household
Fresno	Replicated subsampling	0.126	0.102	0.553
	Conventional	0.076	0.050	0.417
Kern	Replicated subsampling	0.077	0.053	0.525
	Conventional	0.069	0.044	0.400
Sacramento	Replicated subsampling	0.071	0.063	0.648
	Conventional	0.069	0.046	0.525
San Diego	Replicated subsampling	0.049	0.038	0.266
	Conventional	0.047	0.032	0.273
San Joaquin	Replicated subsampling	0.092	0.101	0.500
	Conventional	0.071	0.050	0.375
Stanislaus	Replicated subsampling	0.106	0.071	0.346
	Conventional	0.079	0.048	0.342

errors obtained by subsampling with those derived by the conventional standard error formula [$s/(n^{1/2})$].

As the data in Table 8 indicate, subsampling almost always provided higher estimates of standard errors than did the conventional method for the variables measured. In only one case (for the variable "Person Trips per Household" in the San Diego region) did the standard error acquired from the conventional formula exceed that derived from subsampling. This was a rare situation in which variances within sample clusters were greater than the variance of cluster means.

SUMMARY AND CONCLUSIONS

Because multistage cluster sampling was employed for six regional home interview travel surveys conducted in California, the conventional standard error formula [$s/(n^{1/2})$] underestimated actual standard errors in the survey regions of concern. It was possible, however, to estimate standard errors for the regions using Deming's method of replicated subsampling, which takes into account sample clustering.

Application of replicated subsampling yielded higher and more defensible estimates of total sample error than did the conventional standard error formula, which assumes a simple random sampling design. Leslie Kish's method for calculating standard errors for statistics obtained by cluster sampling is another available technique, but it does not have general application because appropriate design effect factors are not always determinable. Replicated subsampling for large data sets can now, of course, be done quite easily and expeditiously with the use of modern high-speed computers. In brief, replicated subsampling provides an appropriate, unbiased, reliable, and generally applicable framework for estimating sampling error for cluster sample travel surveys.

REFERENCES

1. D. L. Ochoa and G. M. Ramsey. *The 1976–1980 Statewide Travel Survey.* Division of Transportation Planning, California Department of Transportation, Sacramento, Dec. 1981.
2. W. E. Deming. *Sampling Designs in Business Research.* John Wiley and Sons Inc., New York, 1960.
3. H. M. Blalock. *Social Statistics,* 2nd ed. McGraw Hill Book Company, New York, 1978.
4. Peat, Marwick, Mitchell and Co. *Guidelines for Designing Travel Surveys for Statewide Transportation Planning.* FHWA, U.S. Department of Transportation, May 1976.
5. H. Arkin and R. R. Colton. *Statistical Methods.* Barnes and Noble Books, New York, 1970.

California's regional travel surveys were financed through the FHWA, U.S. Department of Transportation. The authors are responsible, however, for the facts and accuracy of the information provided herein. The contents reflect the views of the authors and not necessarily those of the California Department of Transportation or the U.S. Department of Transportation. This paper does not constitute a standard, specification, or regulation.