A procedure is examined for adjusting observed values of a sample of entities toward the overall sample mean. This procedure provides estimators with lower expected mean square errors than the ordinary least-squares estimators, at the cost of accepting a slight bias. The procedure was applied to a variety of analysis problems, using data from two sample surveys, and the performance of the adjusted estimators was compared to that of maximum likelihood estimators. While generally the adjusted estimator was somewhat superior to the ordinary estimator, the improvement was too slight to warrant recommending widespread use of the technique.

An Application of James-Stein Estimation to Survey Data

WILLARD RODGERS
University of Michigan

The phenomenon of regression to the mean is a familiar one to social scientists (e.g., Coleman, 1968; Lord, 1967). Less familiar, however, are strategies for dealing with this statistical phenomenon in the analysis of data. It is the purpose of this article to describe some types of problems in which regression to the mean is a relevant concern, to describe an approach to such problems that has been described in the statistical literature but has not been widely used by social scientists, and to illustrate this approach with two survey data sets for which it is relevant.

Measures of social and psychological variables are generally assumed to include error, or stochastic, terms as well as reflecting the underlying variable. It is commonly assumed that such error terms are normally distributed with expected values of zero, so that the expected observed value for any unit (or set of units)

AUTHOR'S NOTE: *This research was supported by the National Institute of Mental Health Grant 1 R01MH29278-01, "Quality of Life Indicators: Analysis of Detroit Data."*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 10 No. 4, May 1982 463-483

is equal to the true value for that unit. When units are identified by their observed values, however, this implication does not hold. The *conditional* mean of the error term for a unit, given that the observed value is, say, above the overall mean for the sample of units, is *not* zero but is a positive value, and therefore the mean *true* score for such units is less than the mean *observed* score. This leads directly to the phenomenon of regression to the mean; if a second measurement is made of a variable, those individuals with above average observed scores on the first measure will have observed scores on the second measure that tend to be lower, whereas those with below average scores on the first measure will tend to have scores on the second measure that are *higher*—that is, in both types of cases, the scores on the second measurement will tend to be closer to the overall mean.

In general, this phenomenon has been noted only as a warning against possible misinterpretations of data. If participants in an experimental treatment are selected on the basis of high or low scores on a screening test, for example, one should be careful to avoid interpreting a fall or rise in scores on a replication of that test, or on any other measure that is correlated with those test scores, as a consequence of the treatment, when no more may be involved than the statistical artifact of regression to the mean.

For many types of analysis, however, the regression-to-the-mean phenomenon is apparently irrelevant. Consider the relationship between attitude toward a particular issue and education, as observed in a cross-sectional survey data set. The fact that the true scores of individuals with above average levels of reported education tend to be somewhat less than the reported levels, and vice versa for those with below average reported education, does not alter the fact that, in the absence of any overall reporting bias, the reported education of any individual (i.e., one not selected on the basis of reported education or any variable correlated with education) is an unbiased measure of the education of that individual. Moreover, any adjustments that might be considered to take account of the regression to the mean would not change the relationship between the two variables. Shrinking all measured values of either attitude or education (or both) toward the overall means would not alter the ordering of the two variables, and the product-moment correlation would be unaffected. It is perhaps for this reason that the regression-to-the-mean phenomenon has generally been ignored by analysts of survey data.

Consider the following situations for each of which a case can be made that taking account of regression to the mean may improve the quality of the data analysis. First is a study in which multiple reports are obtained about each of a set of entities. This may, for example, be a practical situation, in which recommendation letters have been obtained for each of a set of candidates for admission to a university, and in which the goal is to predict the success or failure of each candidate were he or she accepted. We can conceive of the judges as a sample of the population from which reports could have been obtained, and the rating made by each as a sample of what would have been reported on different days or in response to variations on the solicitation letter and questions. Thus one might assume that the ratings obtained for each candidate reflect, in part, the true capacities of that person for the school and, in part, random variation about that true score. If all judges agree in their ratings of an individual, we would have more confidence that the average rating would have been about the same if we repeated the process than if the judges are widely discrepant in their ratings. The implication is that better predictions of success could be achieved if one were to regress the mean rating of each individual toward the overall mean for *all* individuals, but to make a bigger adjustment for those in which we are less confident—that is, the ones for whom the judges show considerable disagreement.

Next, consider analysis of survey data collected from a design such that samples of individuals are interviewed in each of a set of sites. The aggregated data obtained from each such sample are to be used to characterize the site (in terms, for example, of the racial composition, the attained education, or satisfaction with city services). Our confidence in the characterization of each site will depend on the number of persons interviewed at that site, which may vary in proportion to the population of the site or because of differential response rates, and on the variability of the answers given by those interviewed. It will be argued that the quality of these characterizations of sites can be improved by regressing the aggregate data to the overall mean and making greater adjustments for those sites from which fewer individual reports were obtained or where individual reports show more variability.

STATISTICAL FORMULATION

The type of problem that I have been describing is one in which we wish to estimate each of a sample of parameters from a set of independent observations about each of those parameters. The standard procedure is to use the maximum likelihood estimator (MLE) for each of the parameters, but Stein (1956) showed that when the number of parameters to be estimated is three or more, the MLE is inadmissible under the least-squares criterion. James and Stein (1961) developed an estimator which, although also inadmissible, does have a lower loss function than the MLE. Their estimator was subsequently improved by suggestions in articles by Lindley (1962) and Efron and Morris (1973, 1975). It is assumed that the parameters are normally distributed with variance τ^2 and that the observations themselves are also normally and independently distributed about their parameter values:

$$x_{ij} \mid \theta_i \sim \mathrm{NID}(\theta_i, D_i), \qquad i = 1, 2, \ldots, k, \tag{1}$$

where x_{ij} is the j^{th} observation about θ_i, the i^{th} parameter; and D_i is the variance of the observations about θ_i. The estimator of θ_i that is used in the analyses reported in this article is the following:

$$\hat{\delta}_i = \bar{x} + (1 - \hat{B}_i)(\bar{x}_i - \bar{x}), \tag{2}$$

where $\bar{x}$ is the mean of all observations and $\bar{x}_i$ is the mean of the observations about θ_i. B_i is equal to:

$$B_i = D_i/(\tau^2 + D_i), \quad [3]$$

and is estimated by the following expression:

$$\hat{B}_i = \hat{D}_i/(\hat{A} + \hat{D}_i), \quad [4]$$

where $\hat{D}_i$ is the observed variance of the x_{ij} observations about θ_i and $\hat{A}$ is the MLE for τ^2 which is found through an iterative process, as given by Efron and Morris (1975: expression 3.6):

$$\hat{A} = \sum_{i=1}^{k} (S_i - D_i) I_i(\hat{A}) / \sum_{i=1}^{k} I_i(\hat{A}), \quad [5]$$

where

$$S_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \sim (D_i + \tau^2)\chi_1^2, \quad [6]$$

and $I_i(\tau^2)$ is the Fisher information for τ^2 in S_i:

$$I_i(\tau^2) = 1/\mathrm{Var}(S_i) = 1/[2(D_i + \tau^2)^2]. \quad [7]$$

These procedures were illustrated by Efron and Morris (1975) using data from sports: The batting averages of 36 baseball players after their first 45 at bats and the James-Stein transformations of those averages were compared to their batting averages for the remainder of the season. The transformation reduced the mean square error by a factor of 3.5. In a more substantive analysis involving epidemiological data, these authors adjusted the prevalence rates for taxoplasmosis in each of 36 El Salvador cities, using data from samples of residents of each city. The James-Stein estimates did better than the MLE estimates by a factor of 2.

Finally, we note that the specific problem addressed in this article—the estimation of a sample of subpopulation means—is a particular case of a general problem: that of estimating several parameters all of the same type, and all assumed to be drawn from a normal population of possible values. In this broader context, the problem can be stated as one of finding estimators that improve on the least-squares estimators of the unknown parameters of the general linear model. An alternative method that has been developed in this broader context is that of ridge regression (cf. Hoerl and Kennard, 1970). Ridge regression, similarly to the James-Stein estimation procedure, seeks to find a set of estimators for the regression coefficients in a linear regression model in which a slight amount of bias is accepted in order to reduce the mean square error. No attempt is made in this article to discuss this broader context; there are numerous reviews available in the literature (e.g., Draper and van Nostrand, 1979; Darlington, 1978).

THE DETROIT STUDY

The first study involved drawing a probability sample of persons living in the Detroit metropolitan area (specifically, the Detroit standard metropolitan statistical area as defined in 1970, which included all of Wayne, Oakland, and Macomb counties in Michigan). The sample design involved stratification: Two strata consisted of census tracts in Detroit (and in two "enclave suburbs," Hamtramck and Highland Park), differentiated by racial composition; while the third stratum consisted of the remainder of the standard metropolitan statistical area. A higher sampling fraction was used in the two Detroit strata than in the suburban stratum so as to provide approximately equal numbers of respondents from Detroit as from the suburbs.[1] Blocks in urban areas and chunks of rural areas were sampled with probability proportionate to the estimated number of housing units (HU's). Sampling within blocks and chunks was inversely proportional to size, to yield an expected 16 HU's from each primary selection. Generally the HU's were dispersed throughout a block rather than being compact clusters.

This sample design is unusual, among studies conducted by the Survey Research Center, in the rather large number of households (16) expected from each primary selection. This design was used because of a heavy emphasis in the study objectives on collection of objective data about each such primary selection. This design also permits characterizations of these small local areas in terms of the sample of households sampled from each such area. As will be seen, this aspect of the design is central to the type of analysis reported in this article.

It should be noted that, although the *expected* number of HU's to be selected from each primary unit was 16, the actual number of completed interviews obtained ranged from 0 to 31 for the different units. This variation reflects both nonresponse and deviation of the actual number of HU's found in a selected primary unit from the expected number, due to vacant housing, new construction, and so on.

FINDINGS FROM THE DETROIT DATA

Three types of analyses of the Detroit study data will be reported in this article. The first is concerned with assessing the validity of a characteristic of a neighborhood by aggregating the characteristics of a sample of persons or households in that neighborhood. The second is concerned with the extent to which subjective indicators (i.e., perceptions or evaluations of environmental characteristics as reported by residents of an area) are related to objective indicators of those characteristics. Finally, the third type of analysis is concerned with the accuracy with which respondents are able to report characteristics of their neighbors, either in absolute terms or relative to their own situation. In each case, I will examine the usefulness of making adjustments for regression to the mean by calculating James-Stein statistics.

AGGREGATE CLUSTER CHARACTERISTICS

A primary reason for incorporating relatively large cluster sizes into the design of the Detroit sample was to permit the aggregated characteristics of the sampled households or individuals to be used as descriptions of the neighborhoods in which the individual respondents lived. There are some obvious defects in this plan; for example, there was no attempt to define the clusters in terms of the definitions of the local neighborhoods which were relevant to the respondents. There is little that can be done about this problem. Another problem, though, is that the entire population of households and individuals living in a particular cluster, however that was defined, was not interviewed, only a sample was drawn. As such, an aggregated characteristic of the respondents is a random variable subject to sampling error. Moreover, the amount of sampling error is likely to vary across clusters; if the households in a cluster are homogeneous with respect to a particular characteristic (e.g., their race or age), the aggregate data for the sampled households will generally be closer to the overall cluster characteristic than if the cluster is more heterogeneous. Furthermore, we have noted that the number of households from which data were actually collected ranged considerably (from 0 to 31) across the clusters, and we would expect less sampling error for clusters with larger numbers of respondents than for clusters with few respondents. (We ignore problems that may exist because of nonresponse, which is one of the contributors to variation in the number of respondents in different clusters. The implicit assumption is that nonrespondents are similar to respondents on the characteristics of interest.)

Since aggregate data from respondents in a cluster are subject to sampling variability, it is reasonable to use James-Stein estimators instead of maximum likelihood estimators. Moreover, I follow Lindley's (1962) suggestion of adjusting the estimates toward the observed mean for the sample; specifically, since there were three strata in the sample design, I adjust toward the observed mean for the stratum instead of to the overall mean. The variance of the parameters (i.e., the characteristics of clusters) is estimated using Equation 5. It was found that if we started with an initial estimate for A of 0.0, our estimates of $\hat{A}$ converged (to three or more significant digits) within six or seven iterations.

The first characteristic I examined was the racial composition of the clusters. Across the 115 clusters from which any households were actually interviewed, there was a complete range from 100% white to 100% non-white, with an unweighted mean of about 70% white. Since race is treated here as a dichotomous variable, the variability of the sample mean is a simple function of the proportion and the sample size. The clusters with no observed variance were assigned minimal values greater than zero. Once a stable estimate was obtained for A, the adjusted values were calculated for each cluster using Equation 2.[2]

To evaluate the validity of the aggregate cluster data concerning racial composition, I examined the correlation of this variable with 1970 census data concerning the proportion of the residents of the census tract in which the cluster was located who were non-white. One would not expect perfect agreement between the two variables; the survey data were collected almost four years after the 1970 census, and census tracts are considerably larger than the sampling units selected in this sample design. Nevertheless, the segregated housing patterns of residential areas and the manner in which census tracts are defined guarantee considerable homogeneity of the racial composition across census tracts and over time.

Overall, about 25% of the residents of the census tracts were non-white, compared to an average of 30% non-whites in the selected segments. (The average of the adjusted scores for the segments is almost identical to the average of the sample means.) There is, then, an overall discrepancy which could reflect a trend over time toward a higher proportion of non-whites in the city of Detroit. The correlation coefficient, however, is not sensitive to overall differences of this type.

The correlation between the 1970 census tract data and the simple proportion of non-whites in each segment is .931. The correlation with the *adjusted* proportion is .934, which is barely higher than the correlation with the unadjusted estimates.

A second example of this type of analysis concerns household income. Respondents were asked to report the total income received by their household during 1973. The mean income reported by respondents in a particular cluster ranged from less than $1000 to almost $50,000, with the mean across segments of about $14,000. This variable is compared with the median family income in 1969 for the census tracts within which the selected segments were located, of which the mean was about $12,000. The correlation of the census tract data with the simple mean for the segments was .760, compared to a correlation of .800 for the adjusted means. The proportion of explained variance is increased by a factor of about 1.11.

THE RELATIONSHIP OF SUBJECTIVE AND OBJECTIVE INDICATORS

The next type of analysis with which we are concerned is focused on the relationship of subjective and objective indicators of environmental characteristics. The hypothesis is that the manner in which persons describe or evaluate their environment is in part a reflection of objective characteristics of that environment (i.e., the manner in which it would be described by impersonal measurement or, perhaps, by trained observers who were not residents of that environment). The relationship is not expected to be a terribly strong one, since an individual's perceptions and evaluations probably depend on a wide range of additional factors, including environments experienced in the past and levels of expectation and aspiration. Moreover, environments are generally variable over time and from place to place, even within a restricted area such as that encompassed by a sample segment in the study. Nevertheless, one would expect at least some relationship to exist between the two types of measures, and, indeed, this has been found to be true in past studies. The next question is whether the quality of the subjective ratings can be improved by adjusting the mean rating of all persons sampled in a particular cluster toward the stratum mean, following the procedures outlined earlier.

Indicators of two environmental characteristics were examined: safety from crime and population density. An objective measure of public safety was obtained from the Uniform Crime Reports for 1973. The specific measure used was the FBI Index Crime Rate, which combines several serious (mostly violent) types of crimes and is expressed in incidents per 100,000 inhabitants. These statistics were generally available only for the entire community in which a segment was located,[3] and this lack of specificity with respect to geographic location renders it far from perfect as a measure of safety in the segments. Moreover, the quality of crime reports has been questioned; a large proportion of crimes are not reported to police, and so not included in such statistics.

As a subjective indicator of public safety, respondents were asked to rate the neighborhoods in which they lived on a seven-point semantic differential type scale, with the end-points defined as "safe" and "unsafe." The answers to this question given by respondents selected from a particular sample segment were aggregated to form a composite index, in a process analogous to the combining of the ratings of several judges to form a score on the suitability of an applicant for a job or a place in a university student body. The behavior of an index formed as a simple mean of the several respondents is to be compared with that of an index formed by adjusting those composite scores toward the mean for all respondents in the entire stratum, using the same procedures outlined above.

The criterion used in comparing the sample mean and the James-Stein estimator is the agreement between each of these indices of subjective ratings and the objective crime index. The correlation coefficient is once again used as the agreement coefficient. It was found that the correlation is somewhat higher for the James-Stein (adjusted) scores than for the simple means: .735 versus .707. In terms of "explained variance," the crime index explains slightly more variance in the adjusted scores than in the unadjusted scores, by a factor of 1.08.

The objective indicator of the second environmental characteristic, population density, was population per square mile for the census tract in which each sampled segment was located. This index was calculated from the population reported in the 1970 census, divided by the area of that census tract as measured from maps. The corresponding subjective measure was derived by aggregating responses to a semantic differential item on which each respondent was asked to describe his or her neighborhood on a scale running from "crowded" to "uncrowded." The gain from using the James-Stein adjustment to the cluster mean was, once again, only slight: The correlation coefficient increased, from .375 to .404, and the explained variance increased by a factor of 1.16.

RESPONDENT PERCEPTIONS OF THEIR NEIGHBORS

As part of the interview, each of the respondents was asked a series of questions about his or her close neighbors, defined as the 10 to 15 families living nearest to him or her. Included in this series were questions about age, education, and family income—each to be rated in comparison to the respondent's own situation—and the proportion of neighbors who were white and the proportion who had lived there for at least five years, both with seven-point answer scales.

In addition, each of these five characteristics was ascertained for each of the respondents. Race was noted by the interviewer; family income, the respondent's education and age, and the time lived in the dwelling unit were all asked of the respondents. The characteristics of the respondents living in each sample segment were aggregated, and these sample statistics were compared to the ratings of neighbors given by each respondent.

In Table 1 are shown three ratings of neighbors in relation to the respondent: whether close neighbors were seen as having more or less education than the respondent, as having higher or lower income, and as being younger or older. For all respondents who gave a particular rating, the table shows the difference between the average level for the segment and the level of the respondent; for example, the first row of Table 1 is for respondents who said that their neighbors had less education than they

TABLE 1
Ratings vs. Sample Estimates of Neighbors Relative to Respondent*

Rating (Relative to Respondent)	N	Segment Mean	Adjusted Mean
Education			
1. Less	211	-2.15	-2.12
2. About the same	622	0.02	-0.00
3. More	177	+1.99	+2.11
Correlation, r		.502	.513
Explained variance, r^2		.252	.264
Family Income			
1. Lower	131	-5042	-5619
2. About the same	426	-1335	-2004
3. Higher	243	+3906	+3715
Correlation, r		.372	.378
Explained variance, r^2		.138	.143
Age			
1. Younger	213	-15.49	-16.16
2. About the same	456	-2.05	-2.27
3. Older	410	+10.49	+10.39
Correlation, r		.668	.676
Explained variance, r^2		.447	.457

*The entries are the simple or adjusted mean levels for all respondents in the segment where a particular respondent was located, minus that respondent's level on that characteristic. For example, for the 211 respondents who thought their neighbors had less education than their own, other respondents in their segments reported 2.15 fewer years of education than did those respondents; after adjustment, this discrepancy is estimated to be 2.12 years in the population as a whole.

themselves had, and on the average these respondents reported 2.15 more years of education than the mean level for the respondents in their sample segment. The correlation of the rating of relative education and the observed discrepancy for the sample is .502. This increases very slightly, to .513, if the James-Stein estimation procedure is used to adjust the segment means toward the stratum means. Similar slight improvements in the correlation are observed for James-Stein estimates of segment income and age levels.

Respondents were also asked their perceptions of the racial composition of their neighborhoods and the proportion of their close neighbors who had lived in that neighborhood for at least five years. Use of the James-Stein estimation procedure for the average length of residence of sample respondents increases the correlation of this variable with the ratings by a noticeable amount, from .503 to .543. On the other hand, this procedure results in a *lower* correlation than using the simple segment mean in estimating the racial composition of the segments (the correlation with perceived racial composition drops slightly, from .936 to .924).

THE "MONITORING THE FUTURE" STUDY

The second study from which data will be utilized is a continuing study of high school seniors (see Johnston et al., 1980). Each year since 1975, approximately 17,000 seniors in a set of 130 private and public schools, drawn so as to constitute a national sample of all seniors, have been administered questionnaires dealing primarily with behaviors and attitudes concerning a variety of legal and illegal drugs. Generally a given school participates in the study for two consecutive years, so that about half of the sampled students each year are from schools that participated the previous year. For the larger schools only a sample of the seniors in a participating school is drawn into the study. Sampling weights are applied to each respondent to compensate for any differences in the probability of selection at the two sampling stages.

In this study, the seniors from a given school who participate in the study provide information that can be used to describe that school. Just as in the case of the Detroit study, the accuracy of such sample data as indicators of school characteristics can be expected to vary across schools. In particular, aggregates based on larger numbers of students are more accurate than aggregates based on fewer students. This is an important consideration, since the number of students from a school who participated in the study ranged from fewer than 10 to almost 400.

Answers to several questions answered by each responding senior in a sampled school were aggregated to the school level. These aggregate variables were either simple proportions or proportions adjusted to the overall mean using the James-Stein procedure. These two types of estimators are compared with one another in two ways: (1) which provides a more *stable* estimate of a school's student body, in the sense of agreement across two consecutive years of data collection and (2) which provides a more *valid* estimate, as assessed in terms of correlations with other school characteristics?

STABILITY OF ESTIMATORS OF STUDENT BODY CHARACTERISTICS

It is expected that the seniors in a particular school in a given year will tend to resemble the seniors at the same school the following year. This similarity would arise from shared backgrounds, school characteristics, or student norms, but the separation of these sources of commonality is not relevant to the present analysis. The hypothesis to be tested here is that more stable estimates of student body characteristics can be obtained by adjusting the maximum likelihood estimators toward the overall mean, using the James-Stein procedure. Stability will be assessed by examining the correlation between estimates for two successive years, so that only schools which participated for at least two years are included in this analysis.

The first characteristic examined is the racial composition of the senior class. This is measured in terms of the proportion of respondents in a school who classify themselves as "Black or Afro-American" as opposed to any other race group. The range across schools was almost complete, from 0% to 99% black. For this analysis, only those schools that provided data for each of two consecutive years are examined; this provided a total of 300 data pairs. After obtaining the simple proportion of black students, these proportions were adjusted toward the overall average across all 300 schools (about 10% black). The correlation between the sample proportions for two consecutive years was .9786; the correlation of the adjusted proportions was .9785.

Clearly, no improvement in scores, as assessed in terms of stability from one year to the next, has resulted from the James-Stein adjustment procedure.

Next, I examined the proportion of seniors who reported having used marijuana. Overall, about half of the seniors reported having used marijuana in the past year, but within schools the range was from 0% to 97%. Moreover, the proportions of students who reported having used marijuana 20 or more times in the last 30 days (i.e., on almost a daily basis) was less than 10% overall, but the range within schools was from 0% to 61%. Adjusting these latter proportions by the James-Stein procedure has a dramatic effect on the range across schools; the highest adjusted proportion is only 24%, as compared to the highest unadjusted proportion of 61%. The usefulness of the adjustment, however, is not confirmed when we examine the stability across years. The correlation of the unadjusted proportions of seniors who have used marijuana in the last 12 months is .791, while the correlation of the adjusted proportions is .794. For the proportion of daily users, the correlation of the unadjusted proportions is .718, while the correlation of the adjusted proportions is only .631.

VALIDITY OF ESTIMATORS OF STUDENT BODY CHARACTERISTICS

No outside sources of information about drug use or racial composition are available for the schools in this study. We can, however, examine the *construct* validity by looking at the correlations of aggregated reports of the senior classes with other school characteristics that we expect to be related. The relevant school characteristics which were available were limited to the following: whether the school is public or private; the mean education of the parents of the responding seniors (as reported by the seniors); the proportion of the responding seniors in a college preparatory program (as reported by the seniors); and region of residence. The correlations of each of these school characteristics with the adjusted and unadjusted measures of

TABLE 2
Correlations of Unadjusted and Adjusted Proportions of Seniors Who are Black with School Characteristics

School Characteristic	Year 1 Unadjusted	Year 1 Adjusted	Year 2 Unadjusted	Year 2 Adjusted
Public vs. private	.077	.076	.088	.086
Mean education of parents	-.253	-.250	-.261	-.259
Proportion of seniors in college preparatory program	-.050	-.049	-.044	-.043
Region: Northwest	-.062	-.062	-.056	-.056
North central	-.146	-.145	-.153	-.151
South	.304	.301	.311	.307

racial composition of the senior class are shown in Table 2. The correlations are shown for both "Year 1" and "Year 2," which are essentially replicates of one another ("year" refers to the measurement of stability reported above). It is clear from Table 2 that the adjusted proportions provide no evidence of greater validity as compared to the unadjusted proportions.

Tables 3 and 4 provide comparable data for measures of marijuana usage. Again, there is little support for the hypothesis that the James-Stein adjustment procedure provides more valid measures of drug usage among students at different schools.

SUMMARY AND DISCUSSION

I have applied the James-Stein estimation procedure to a variety of analyses of data collected from a clustered sample of residents of a large metropolitan area and from samples of students in each of a sample of high schools. First, I adjusted the aggregated characteristics of respondents in a particular geographic cluster toward the mean values across their entire strata and compared the agreement of these adjusted and unadjusted

TABLE 3
Correlations of Unadjusted and Adjusted Proportions of Seniors Who have Used Marijuana in the Last 12 Months with School Characteristics

School Characteristic	Year 1		Year 2	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Public vs. private	.014	.014	.042	.050
Mean education of parents	.384	.377	.289	.280
Proportion of seniors in college preparatory program	.239	.233	.273	.267
Region: Northwest	.269	.270	.321	.328
North central	-.109	-.105	-.070	-.068
South	-.303	-.296	-.317	-.321

aggregates to census tract data (i.e., population data, albeit for a different year and larger area). Next, I aggregated subjective data from respondents in a cluster and compared the strengths of association between adjusted versus unadjusted aggregates and objective data. Finally, I asked whether the agreement between respondent perceptions of their neighbors and aggregated cluster scores would be improved by using the James-Stein procedure to adjust those cluster aggregates. In almost every case, the James-Stein estimators were better than the simple (MLE) estimators, but the improvement was slight at best.

Data from the samples of seniors were aggregated to provide indicators of student body characteristics. Adjusting such aggregate measures by the James-Stein procedure failed to produce consistent improvements in their quality as assessed in terms either of the stability of the measures from one year to the next or the construct validity of those measures as assessed by their correlations with school characteristics.

My conclusion from this series of analyses is that little has been gained by using the James-Stein estimators as opposed to maximum likelihood estimators. The analyses reported in this article cover a variety of types of problems from two data sets.

TABLE 4
Correlations of Unadjusted and Adjusted Proportions of Seniors Who have Used Marijuana Daily with School Characteristics

School Characteristic	Year 1		Year 2	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Public vs. private	.114	.143	.078	.126
Mean education of parents	.186	.192	.136	.148
Proportion of seniors in college preparatory program	-.042	.043	.020	.116
Region: Northwest	.167	.215	.227	.273
North central	-.060	-.023	-.035	.008
South	-.171	-.184	-.205	-.224

They offered a reasonable context in which to judge the James-Stein procedure, both in terms of the opportunity for making substantial adjustments (i.e., there was a wide range of variances and numbers of reporting units, so that some elements were adjusted considerably more than others) and for assessing the relative quality of the estimators. The absence of any meaningful improvements in the quality of the estimators after making the James-Stein adjustments leaves little basis for thinking more encouraging results would be obtained in other problems or with other data sets.

It is possible, of course, that the particular problems to which I have applied the adjustment procedure are not representative of the population of analysis problems to which such a procedure might be applied or that there are certain types of problems not represented here for which the James-Stein estimators are very useful. An alternative to the approach of this article, where I have applied the adjustment procedure to actual data, would be to carry out a simulation study in which the MLE and James-Stein estimators are compared to the known (true) values. This approach has been used in an extensive comparison of various alternatives to least squares by Dempster et al. (1977). Such simulation studies are also limited, however, by the specific characteristics of the simulated data. In conclusion, since Efron and Morris describe the James-Stein adjustment procedure as an "empirical Bayes" technique, perhaps it is fair to say that while my prior beliefs in the potential benefits from the application of this technique were sufficiently high to motivate the analyses described in this article, my posterior evaluation is that the average benefits would not justify the effort of further use of the method.

NOTES

1. In substantive analyses of these data, respondents are weighted to compensate for the two different probabilities of selection of housing units within the different strata. This has not been done, however, for the methodological analyses reported in this article.

2. In general, the cluster means were adjusted toward the overall mean for the sampling stratum from which each cluster was selected. In this particular case, the two Detroit strata were combined because the definitions of these strata were based on the racial composition of the census tracts. In an initial analysis, I kept the two strata separate, and this resulted in an artifactual improvement of the estimates.

3. Statistics were available for the individual police precincts, of which there are 13, for segments located in the city of Detroit.

REFERENCES

COLEMAN, J. S. (1968) "The mathematical study of change," pp. 428-478 in H. M. Blalock, Jr., and A. B. Blalock (eds.) Methodology in Social Research. New York: McGraw-Hill.

DARLINGTON, R. B. (1978) "Reduced-variance regression." Psych. Bull. 85: 1238-1255.

DEMPSTER, A. P., M. SCHATZOFF, and N. WERMUTH (1977) "A simulation study of alternatives to ordinary least squares." J. of Amer. Stat. Assn. 72: 77-91.

DRAPER, N. R. and R. C. van NOSTRAND (1979) "Ridge regression and James-Stein estimation: review and comments." Technometrics: 21: 451-466.

EFRON, B. and C. MORRIS (1975) "Data analysis using Stein's estimator and its generalizations." J. of Amer. Stat. Assn. 70: 311-319.

——— (1973) "Stein's estimation rule and its competitors—an empirical Bayes approach." J. of Amer. Stat. Assn. 68: 117-130.

HOERL, A. E. and R. W. KENNARD (1970) "Ridge regression: biased estimation for nonorthogonal problems." Technometrics 12: 55-67.

JAMES, W. and C. W. STEIN (1961) "Estimation with quadratic loss," pp. 361-379 in J. Neyman (ed.) Proceedings of Fourth Berkeley Symposium (Vol. 1). Berkeley: Univ. of California Press.

JOHNSTON, L. D., J. G. BACHMAN, and P. M. O'MALLEY (1980) Monitoring the Future: Questionnaire Responses from the Nation's High School Seniors, 1979. Ann Arbor, MI: Institute for Social Research.

LINDLEY, D. V. (1962) "Discussion." J. of Royal Stat. Society B-24: 285-287.

LORD, F. M. (1967) "Elementary models for measuring change," pp. 21-38 in C. W. Harris (ed.) Problems in Measuring Change. Madison: Univ. of Wisconsin Press.

STEIN, C. M. (1956) "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," pp. 197-206 in J. Neyman (ed.) Proceedings of Third Berkeley Symposium (Vol. 1). Berkeley: Univ. of California Press.

Willard Rodgers is Senior Study Director at the Survey Research Center at the University of Michigan. He is coauthor of The Quality of American Life *(Russell Sage Foundation, 1976) and directed the Detroit quality of life study mentioned in this article.*