AGRICULTURAL ECONOMICS RESEARCH

A Journal of Economic and Statistical Research in the
United States Department of Agriculture and Cooperating Agencies

Volume VII APRIL 1955 Number 2

Measurement of Sales of Apples in Retail Stores

By Earl E. Houseman

The relative merits of different methods of measuring volume of retail sales of particular commodities has been a debated subject in recent years. As a byproduct of an experiment in retail store merchandising, a direct comparison of some alternative methods is made in this paper, a matter of importance, we believe, to persons interested in the measurement of retail sales.

IN OCTOBER and November 1953, an experiment was conducted in 12 retail stores in Pittsburgh to ascertain the effects of four methods of merchandising on the sale of apples.[1] A report of information obtained in the study of merchandising methods will become available. In this article we are concerned with alternative statistical techniques, therefore we shall discuss the experiment only to the extent necessary to provide a description of the data upon which results here reported are based.

Plans for the experiment included provision for obtaining information weekly, for an 8-week period, on total sales of apples by adjusting each store's purchases for inventory changes and losses. Total sales were the variable to be used in evaluating the effects of treatment. But, in addition, a count of customer units and the pounds of apples bought by each were recorded every day in each store for two 45-minute periods. A customer unit was defined as one or more persons shopping together, irrespective of whether apples were bought.

[1] This study was conducted by the former Production and Marketing Administration of the United States Department of Agriculture in cooperation with Pennsylvania State University and Cornell University.

From this experiment it was possible to compare (1) lbs/45m (pounds sold per 45-minute period) and lbs/100c (pounds sold per 100 customer units) as derived from the observation of customer purchases, with (2) total quantities of apples sold by the 12 stores during each week. The method of estimating retail sales from a sample of stores wherein total sales are obtained by adjusting each store's purchases for inventory changes and losses will be referred to as the "audit method."

The method of observing customer purchases during specified hours will be referred to as the "observation method." The observation method involves the sampling of stores *and* time (sample of hours), in contrast to the audit method, which requires only the sampling of stores.

With the exception of a relatively few observation periods with missing data, as just indicated, data were available for 12 observation periods (two each day for 6 days) for each week and for each store over an 8-week period. For convenience in the analysis, substitutions were made for the periods with missing observation by visually inspecting the data, deciding upon a range of values within which the actual value for the missing period probably would have been, and substituting a number selected at random within that range. For various reasons, approximately 2 per-

336409—55——1

cent of the observations were missing, but in many of these cases notes made by the field staff were available. These were helpful for assigning substitute values for missing observations. It seemed clear that the assignment of values for missing data could have only a negligible effect on the results to be presented.

The 12 Pittsburgh stores in the experiment, members of the same chain, were distributed over the city. They encompassed a variety of conditions. None had a gross business of less than roughly $20,000 a week.

Responsibility for maintaining the experiment was assigned to six men, with two stores to each man. They helped to prepare displays, kept an appropriate supply of apples on hand, and collected necessary information. Working arrangements between field staff and test stores were such that data on total weekly sales are believed to be virtually without any measurement error. It was not possible for the field supervisor to check the extent to which customers who did not buy apples may have been missed in the count of customers, either because of crowded conditions during rush periods or through observers' inattention at other times. But the data provide a useful source of information for comparing alternative methods of measuring retail sales.

As the measurement of sales or changes in sales through time and the measurement of effects of merchandising practices are different problems in terms of study design, they are discussed separately.

Measurement of Changes in Volume of Retail Sales

To evaluate alternatives, it is necessary to specify a true value for the population sampled and to consider the departures from the true value of estimates based upon the alternative methods. It is assumed that we are trying to estimate

$$R_i = \frac{T_i}{T_b}$$

where R_i is the ratio, for all stores in the defined universe, of the true total sales, T_i, for the ith week to the true total sales, T_b, during some base week. It is possible to design samples for estimating T_i but for this study the ratio R_i is considered as the true value to be estimated instead of the universe total, T_i, because of the problems involved in any attempt to expand either pounds per hour or per 100 customers to a total, particularly in the practical setting under which observations during specified hours have been made in past surveys or experiments. Let us consider lbs/45m versus lbs/100c as estimates of R_i before contrasting the audit method with the observation method.

Pounds per 45-minute period vs. pounds per 100 customers.—To compare these two measures of rate of sales, the data from the Pittsburgh apple experiment were divided into 12 subsamples. One observation period a week for each store was used in each subsample. With the aid of Latin-square arrangements it was possible to have each subsample consist of an equal number of large and small stores by 2-day time-periods within a week. *The underlying idea in the subsampling was to obtain 12 subsamples such that any one of the 12 might have been the sample of hours chosen if the project had been planned originally so that only one observation per week was taken in each store.*

The schedule of hours for observation in each store changed from one week to the next but the same schedule was used for weeks 1, 3, 5 and 7, and another for weeks 2, 4, 6 and 8. Treatments (merchandising practices) were changed every 2 weeks. Hence, between weeks 3 and 4, there was no change of treatments but the schedule of observation hours for a given store was different for each of the two weeks. From week 4 to week 5, both the treatments and schedule of observation hours changed.

The field staff adhered closely to the schedule of hours for observing customer purchases. Observations for any given store and day were identified as first and second observations. As the set of 12 subsamples was held constant from week to week, the store-hours in any given subsample are matched between odd numbered weeks and between even numbered weeks. This also means, for example, that the first observation in store A on Tuesday in an odd numbered week must be in the same subsample as the first observation in store A on Tuesday in an even numbered week.

For each subsample, pounds of apples bought per 100 customers and pounds bought per 45-minute period were computed for each week. As a means of getting the two measures on a common basis for comparison, ratios were then computed for each subsample for each week to the preceding week, for each even numbered week to the preceding even numbered week, and for each odd numbered week to the preceding odd numbered week.

TABLE 1.—*Summary of comparisons of 12 subsamples*

Line	Item	Weeks compared [1]								
		4/3	5/3	5/4	6/4	6/5	7/5	7/6	8/6	8/7
1	Ratios of total sales	0.42	0.44	1.05	1.04	1.00	0.98	0.99	1.66	1.68
	Pounds per 100 customers									
2	Ratios for the combined subsamples [2]	.59	.52	.89	.85	.96	1.02	1.06	1.46	1.38
3	Averages of the subsample ratios	.60	.53	.91	.88	.98	1.06	1.10	1.54	1.43
4	Root mean square errors [3]	.20	.16	.28	.31	.19	.29	.27	.47	.43
	Range of values of subsample ratios									
5	Lowest	.50	.37	.61	.57	.80	.57	.66	1.06	.99
6	Highest	.79	.75	1.48	1.48	1.41	1.58	1.43	2.56	2.04
	Pounds per 45 minute period									
7	Ratios for the combined subsamples [4]	.53	.49	.92	.82	.89	.93	1.04	1.64	1.57
8	Averages of the subsample ratios	.54	.51	.93	.85	.89	.96	1.12	1.74	1.60
9	Root mean square errors [3]	.15	.16	.29	.35	.24	.26	.36	.61	.38
	Range of values of subsample ratios									
10	Lowest	.41	.31	.58	.53	.64	.52	.71	1.06	1.02
11	Highest	.68	.69	1.36	1.65	1.41	1.40	1.78	3.19	2.25

[1] For example 4/3 refers to ratios of week 4 to week 3.

[2] The values recorded in this line are $\left(\frac{t_i}{c_i}\right)\left(\frac{c_j}{t_j}\right)$ where t and c are total pounds sold and total customers counted for all 12 subsamples and the subscripts refer to weeks.

[3] For any column (pair of weeks) the root mean square error is $\sqrt{\frac{\Sigma(r_i-R)^2}{12}}$ where r_i is the ratio for the ith subsample and R is the true ratio (line 1 of this table).

[4] The values recorded in this line are $\frac{t_i}{t_j}$ where t is total pounds sold for all 12 subsamples and the subscripts refer to weeks.

The data for each subsample were also converted to an index using the eighth week as a base, but that analysis is not presented here as the results were essentially the same.

The "true" ratios, of which the subsample ratios should be estimates, are given in the first line of table 1. These ratios are based upon the total sales of apples by the 12 stores as derived from the audit method. It is clear from the table that neither lbs/100c nor lbs/45m, for the size of sample involved, reflected change with a satisfactory degree of accuracy. From a comparison of lines 2 and 7 with line 1, it is clear that subsampling time even to the extent of taking two observations each day did not provide good estimates of the ratios for all pairs of weeks. Part of the data underlying table 1 are presented graphically in figure 1.

Because of a sharp increase in price at the end of the third week, the quantity of apples sold during weeks 4, 5, 6 and 7 was comparatively low until week 8, when the price dropped. It is interesting that on the two occasions (weeks 3 to 4 and 7 to 8) when the greatest change occurred, as shown by the "true" ratios, neither lbs/100c nor lbs/45m came close to reflecting the full extent of the true change but lbs/45m was appreciably closer in both cases. Incidentally, week 7 included Thanksgiving Day. Under the concept of pounds per hour that was used for this analysis, a regularly scheduled observation period that falls on a holiday is counted as an observation period with zero sales.

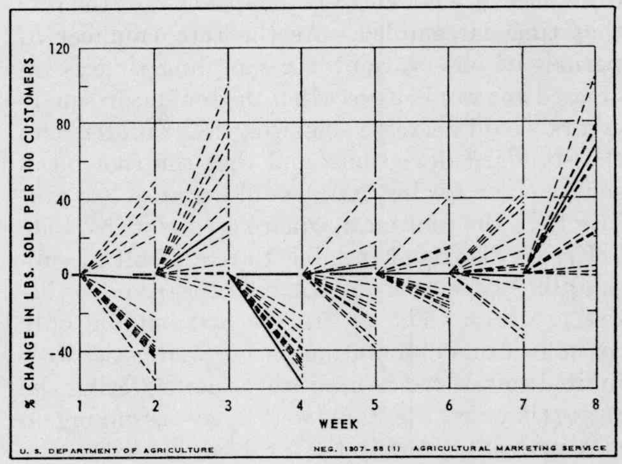

FIGURE 1.—Week to week changes in pounds of apples sold per 100 customers for each of 12 subsamples (dotted lines) compared with changes in actual sales by same stores (solid lines).

Note that the ordinary standard deviation, which would be a measure of the variation among the 12 subsamples, was not used in table 1 as a measure of precision. Instead, the root mean square error was used, which is the square root of the average of the squares of the deviations of the 12 subsample ratios from the "true" ratio. The two principal reasons for using the root mean square error instead of the standard error were: (1) There might be some biases in the time subsamples, and (2) even if the subsamples were unbiased (or random) samples of time, lbs/100c, technically speaking, gives estimates having bias (although the bias might be small) in addition to the usually negligible bias that exists in ordinary ratio estimates.

Because of the possibility of a bias, as just indicated, and because holidays, weather, and other factors influence the number of customers during a specified hour of the week, it was anticipated that the standard deviation for lbs/100c might be considerably less than the standard deviation for lbs/45m. For lbs/100c and lbs/45m the simple average of the 9 root mean square errors in lines 4 and 9 of table 1 and the simple average of the corresponding standard deviations are as follows:

	lbs/100c	*lbs/45m*
Root mean square error	.29	.31
Standard deviation	.25	.29

This shows that the standard deviations were not much less than the root mean square errors and that on the average lbs/100c and lbs/45m were about equally accurate. However, the relative differences are not entirely independent of the rate that time is sampled. As the rate (number of periods of observation) for sampling time is increased one would expect that the root mean square errors would decrease somewhat less rapidly than the standard deviations, and that the root mean square error for lbs/100c would decrease less rapidly than the root mean square error for lbs/45m.

Errors displayed in table 1 are a result of subsampling time, and do not reflect variability between stores. The errors are attributable only to the fact that observations on customer sales were limited to selected hours rather than covering the entire time that the stores were open—assuming no measurement errors in the total sales based on the audit method. It is not clear, *a priori*, that the comparative precision of lbs/45m and lbs/100c must be about the same when the between-store component of error is brought into consideration. But as we are concerned with estimating "change" rather than "level" from the same sample of stores through time, one might expect, intuitively, that the accuracy of the two estimators would remain about the same when between-store variation is added.

To make a comparison, taking store differences into account, it was necessary to decide whether to do this by use of variance formulas or by drawing a number of subsamples of stores and examining the differences between the subsamples. The latter method was chosen primarily because of the small number of stores involved and because the appropriate variance formulas, which are approximations, are of doubtful validity under the present circumstances. Note, when lbs/100c for one week is divided by lbs/100c for another, that the variance of a quantity like $\frac{X_1}{X_2} \cdot \frac{X_3}{X_4}$ is involved, where X_1, X_2, X_3, and X_4 are all variables.

Thirty subsamples of 4 stores were selected from the 12 stores by use of a table of random numbers. Totals of the 12 observation periods for each store-week were used in the computation of ratios for each subsample of stores to be compared with the true ratios. Thus, the analysis followed the same pattern as that for results given in table 1. In this case we are dealing with subsamples of stores whereas in the previous case we were dealing with subsamples of time. Results from subsampling of stores are shown in table 2. The section of the table headed "total sales" does not involve sampling of time, whereas the sections on lbs/100c and lbs/45m involve sampling time to the extent of 12 45-minute observation periods per week in each store.

Again, the two measures, lbs/100c and lbs/45m, appear to be about equally accurate.

Audit method vs. observation method.—It is clear from theoretical considerations that, given a probability sample of n stores, the standard error of $r_i = \frac{t_i}{t_b}$ is less than the standard error of $r'_i = \frac{t'_i}{t'_b}$ where t_i and t_b are sample totals from the audit method for the ith week and a base week, respectively, and t'_i and t'_b are sample totals of pounds sold during specified hours from the observation method. In fact, if only one hour is observed each week in each of the n stores, the sample on which r'_i is based is in a sense only roughly 2 to 3 percent as large as the sample on which r_i is based. If any

TABLE 2.—*Summary of differences among 30 subsamples of 4 stores each*

Line	Item	Weeks compared			
		5/3	7/5	6/4	8/6
1	Ratio of total sales ("true" ratios)	0.44	0.98	1.04	1.66
	Total sales (audit method):				
2	Average of subsample ratios	.44	.98	1.06	1.67
3	Standard error	.08	.17	.11	.31
	Range of values of subsample ratios:				
4	Lowest	.32	.71	.85	1.06
5	Highest	.60	1.25	1.25	2.26
	Pounds per 100 customers:				
6	Average of subsample ratios	.52	1.06	.85	1.56
7	Standard error	.14	.32	.11	.41
8	Root mean square error	.17	.33	.22	.42
	Range of values of subsample ratios:				
9	Lowest	.27	.69	.64	1.13
10	Highest	.79	1.74	1.12	2.21
	Pounds per 45 minute period:				
11	Average of subsample ratios	.50	.96	.83	1.75
12	Standard error	.12	.26	.14	.43
13	Root mean square error	.13	.26	.26	.44
	Range of values of subsample ratios:				
14	Lowest	.30	.68	.61	1.07
15	Highest	.73	1.49	1.19	2.45

measurement errors for the audit method are under control, it is clear from theory and experience that the error in r_i as an estimator of R_i must, on the average, be much smaller than the error in r'_i. Even with 12 different 45 minute periods stratified by time, the difference as indicated in table 2 (line 3 compared with line 13) is substantial. This is nothing more than a reflection of the fact that if one has a sample of n sampling units (stores) the error will be greater when the n sampling units are subsampled than when they are enumerated completely.

Next, let us consider $r''_i = \frac{t'_i}{t'_b} \frac{c_b}{c_i}$ where c_i and c_b are the sample total numbers of customers counted during the observation periods for the ith week and base week, respectively. In other words, r''_i is the rate one would compute when using lbs/100c. It is not axiomatic, as was the case with r'_i, that the sampling error of r''_i as an estimator of R_i must be less than the sampling error of r_i. This partially explains the interest manifest in this article in comparing lbs/45m with lbs/100c. If the sampling error for r''_i is not appreciably less than the sampling error for r'_i, there is no hope for r''_i being better than r_i as an estimator of R_i. The results in tables 1 and 2 and the results of at least one other experience not reported here do not indicate any appreciable superiority of r''_i over r'_i.

We have been considering r_i, r'_i and r''_i as estimators of the parameter R_i; but if $R'_i = \frac{T_i}{T_b} \frac{C_b}{C_i}$, where C_i and C_b are total customer counts in all N stores in the population, is a satisfactory parameter to be estimating, then r''_i probably should be evaluated as an estimator of R'_i instead of R_i. The precision of r''_i as an estimator of R'_i does not, however, appear to be much different from the precision of r'_i as an estimator of R_i.

As estimators of R_i it is clear that r_i must be much better than r'_i. Therefore, if r'_i and r''_i are about equally good, the conclusion to be drawn is that estimates provided by the audit method must be more precise than estimates based upon lbs/100c from the observation method.

Two assumptions underlying the preceding statement are: (1) That the same number of stores would be used in the audit and observation samples and (2) that the sample does not change from week to week. The statement also overlooks the question of certain biases that might be associated with each method.

Measurement of Effects of Merchandising Practices

Pounds sold per 100 customers could also be used for measuring differences in merchandising practices under either the survey or the experi-

TABLE 3.—*Analysis of variance of data on apple sales*

Source of variation	Degrees of freedom	Mean squares		
		Pounds per 45 minute period [1]	Pounds per 100 customers [1]	Total sales
Replications	2	16. 5	178	1, 861, 000
Periods	3	68. 8	840	2, 594, 000
Replications × periods	6	9. 0	76	30, 000
Stores within replications	9	54. 5	962	680, 000
Treatments	3	66. 9	1, 022	1, 997, 000
Error	24	5. 8	89	134, 000

[1] Combined data from all observation periods.

mental type of study design. For the Pittsburgh experiment, a 4 x 4 Latin-square design replicated three times over stores was used—each treatment remaining in each store for a 2-week period. Using the Pittsburgh data, let us examine the comparative "power" of lbs/45m and lbs/100c from the observation method with total sales from the audit method for measuring differences between treatments (merchandising practices). This will be done without reference to the question of which measure is conceptually the more useful.

From a standard analysis of variance procedure, the results presented in table 3 were obtained for each of the three ways of measuring sales. As indicators of the power for discriminating between treatments, two criteria are appropriate, the "F" ratios for the treatment mean squares, and the coefficients of variation (square roots of the error mean squares divided by the general means):

	"F" ratios	*Coefficients of variation*
Total sales (audit method)	14. 8	0. 19
Full time sample (observation method):		
lbs/45m	11. 6	. 24
lbs/100c	11. 5	. 23
Average over the 12 subsamples:		
lbs/45m	[1] 3. 1	[2] . 57

[1] Average of the 12 treatment mean squares, from separate analysis of variance for each subsample, divided by the average error mean square.

[2] Square root of the average error mean square divided by the mean.

The full sample of time refers to aggregates of the 24 observations on a treatment in a store during a 2-week period. For this rate of sampling time, there is some loss in precision as compared with use of total sales, but a large loss in precision occurs when time is sampled only to the extent of one observation period per week.

Table 4 shows sales for each treatment as a percent of the total for all four treatments for each method of measuring sales. The three measures give about the same results when the full sample of time is used, namely, 12 45-minute periods of observation in each store each week. Percentages corresponding to those in line 2 of table 4 were computed for each of the 12 subsamples of time. There were large differences between the subsamples; for example, three of the 12 subsamples showed slightly greater sales for treatment A than for treatment B, and two out of the 12 showed sales for B at slightly more than twice the sales for A.

Discussion

The large variation among stores because of the range in size of stores and other factors is apparent. Likewise, for any given store, amount sold during an hour varies widely with a number of factors, including time of day, day of the week, and weather. Therefore, intuitively, one might regard lbs/100c as a good basis for measuring sales because of an expectation that such variations would not influence it as much as pounds per hour, or, for example, pounds per store in the case of the audit method. Incidentally, with the audit method lbs/100c can also be used if arrangements are made with the stores for getting the total number of cash-register "ring-ups."

A point that is sometimes overlooked is that different measures of rate of sales, such as pounds per 100 customers, pounds per store, and dollars worth of apples sold per $100 of sales of all com-

TABLE 4.—*Percentage distribution of sales of apples, by merchandising practices*

Item	Percentage of total sales by practice				
	A	B	C	D	Total
	Percent	*Percent*	*Percent*	*Percent*	*Percent*
Total sales	17. 7	24. 5	29. 6	28. 2	100
Pounds per 45 minute period [1]	17. 5	23. 3	30. 3	28. 9	100
Pounds per 100 customers [1]	16. 9	22. 8	31. 8	28. 5	100

[1] Based upon the full sample of time.

modities or of produce, involve different concepts. These are conceptual differences that are more than just a question of whether distance, for example, should be measured in terms of inches or centimeters. Thus, the criteria for choosing a measure of rate of sales should include the utility of the different measures assuming no sampling error as well as sampling variability and biases.

Let us examine the coefficients of variation among the 12 stores in the Pittsburgh experiment, comparing lbs/45m and lbs/100c from the observation method and pounds per store from the audit method, by using aggregates for the eight weeks. For the observation method two levels of sampling time will be considered: (1) All 96 observation periods in each store over the 8 week period, and (2) 8 observation periods, one each week in each store.

Method	*Coefficient of variation among stores*
Audit, lbs/store	0.25
Customer, lbs/45m:	
96 observation periods	.34
8 observation periods	[1] .63
Customer, lbs/100c:	
96 observation periods	.38
8 observation periods	[1] .57

[1] Mean squares between stores were computed for each of the 12 subsamples. The square root of the average mean square for lbs/45m and for lbs/100c was divided by lbs/45m and by lbs/100c, respectively.

Remember, the present setting differs from that represented in table 2. We are now considering relative variation in estimates for a given time period, not relative variation in ratios of one time period to another. Under the present setting, presumably, a customer count has a greater potential for reducing variability because estimates of "level" are involved rather than "change" from a matched sample.

The coefficients of variation are estimates and, in this case, subject to rather large sampling errors. But a translation of these coefficients of variation into requirements for equal precision indicates for the observation method that roughly twice as many stores sampled at the rate of 12 observation periods (one each half day) per week would be required to provide the same precision as that provided by the audit method. If one observation instead of 12 is taken each week in each store, about 5 or 6 times as many stores would be required to obtain the same precision as that provided by the audit method. These statements about sample size should not be interpreted as applying generally.

It was clear, before making the above computations, that the coefficient of variation for lbs/45m must be larger than the coefficient of variation for the audit method. One might have expected the coefficient of variation for lbs/100c to be less than the coefficient of variation for lbs/45m. The failure of this to happen is evidently attributable, to a considerable degree, to one or more of three factors:

(1) The range of variation from store to store in total quantities of apples sold was rather limited because of the choice of stores for the experiment; hence the potentiality for a customer count being effective in reducing variation was rather limited.

(2) Customer buying patterns or habits differed considerably among some of the stores. This is believed to be the principal reason why, for example, customer counts for the two stores that sold the most apples were in the proportions of 2 to 1, whereas the difference in total quantities sold was only 10 percent. One of these two stores was in an outlying area. Most of its trade was of the "drive-in" type. The other store was in an

area in which population density was much higher. Most of its trade was of the "walk-in" type.

(3) Some differences exist from store to store in a "customer unit" because of differences in the physical arrangements within stores, differences between enumerators, difficulties of counting all non-apple-buying customers under crowded conditions. Possibly there are other sources of differences. In fact, one might think in terms of a different definition being associated with each store. For example, if a customer count involves counting customer units passing through the produce department, would one say that the definition of a customer unit is the same for two stores if the floor plan of one of the stores is such that 90 percent of the persons entering the store go through the produce department, whereas in the other only 70 percent pass through the produce department?

It is important to keep these three factors in mind when attempting to judge whether a customer count will be helpful for purposes of reducing sampling variation as compared to the lbs/45m estimator. They are the most likely reasons why the customer counts and pounds sold, during the observation hours, were not sufficiently correlated to more than offset the added variability through the introduction of customer counts. That is, since customer counts are subject to sampling error, when total pounds sold during the observation periods is reduced to lbs/100c, a component of variation is added. If the correlation between customer counts and pounds sold is not great enough to more than offset this added component of variation, lbs/100c must have a greater coefficient of variation than lbs/45m.

There is one additional point on which some comment seems warranted. Variations due to differences in size of store, for example, can be "controlled" in various ways. After variation attributable to one source has been effectively controlled by one means, little if anything is gained—in fact a loss might occur—by superimposing another control on the same type of variation. To be more specific, suppose, for example, that the observation method and the same sample of stores and observation hours through time are used to estimate changes in sales. Variation due to size of store and time of day or week is fairly well taken care of by the design. Hence, it is reasonable that lbs/45m and lbs/100c appeared about equally accurate in tables 1 and 2 and in the measurement of differences between merchandising practices. Other examples could be cited in both experimental and survey types of studies.

There are many practical aspects of the problem of choosing a method for measuring retail sales, a discussion of which is beyond the scope of this article.[2]

Summary

Sampling Error or Efficiency.—Regardless of whether a survey or a controlled experiment is involved, for a given sample of stores, the relative sampling error for either lbs/hour or lbs/100c from the observation method must be appreciably greater than the relative sampling error for either lbs/store or lbs/100c from the audit method, respectively. It is theoretically possible for lbs/100c from the observation method to have a smaller relative sampling error than lbs/store from the audit method. But this did not happen in the Pittsburgh study, and is almost certain not to occur when variation associated with size of store is "controlled" in the survey or experimental design. The results presented in this report indicate that a large loss in statistical efficiency occurs when, instead of total sales, the purchases of apples during a 45-minute period once a week in each store are used to measure rate of sales.

When the observation method has been applied, the statistical population has usually been restricted to the larger stores, at least partly, to avoid sending an enumerator to a store in which he

[2] In reviewing the preliminary draft of this article, M. E. Brunk of Cornell University made the following observation: "Your analysis clearly indicates the relative advantage of measuring movement rate by use of the audit method. But the movement rate alone does not provide the trade with the essential information needed. The trade needs to know *why* movement rates change, and this can be accomplished only by including data on retail prices and practices which affect such movement. Adding this information complicates the problem. It suggests the possibility of combining the audit method of determining movement rate with a probability sample of stores for the purposes of determining associated trade practices. From customer observations both types of information are obtained, but your analysis suggests that this may possibly be an inefficient approach to the problem. Certainly additional research is needed in order to determine the most practical method of such market reporting."

would frequently have no sales, or very few, to observe during an hour. For a survey type of operation, and with the audit method, one could use a probability sample of stores with allocation of the sample by size groups in a statistically optimum manner. A cut-off point might be used to eliminate stores below a certain size. But the statistical population would be less restricted, so the results, as estimates of rate of sales or change in rate of sales for a city, would be less subject to the potential biases attributable to any limitations placed on the kind of stores that are permitted in the sample. We should not overlook the fact that differences in sampling variation resulting from changes in specifications of the population are not of the same nature as differences in sampling error associated with alternative methods of sampling and estimation for a given population.

Because it is more precise, the writer believes the audit method should be generally used unless it fails because appropriate records are not available, because of noncooperation, or for similar reasons.

Differences in Concepts.—It has been pointed out that lbs/100c and pounds per store (or simply total sales of a particular commodity) involve a difference in concepts. We need to recognize that reduction of data to lbs/100c gives results with a particular meaning depending upon the meaning of the customer count. A difference of 10 percent in lbs/100c does not necessarily mean a difference of 10 percent in per capita purchases or a difference of 10 percent in retail sales. To illustrate, suppose the actual per capita purchases of a commodity were the same in January and April and that the family buyers went to the store 10 percent more frequently in April than in January. Pounds bought per 100 customers would be less for April than for January even though the rate of purchases per capita remained unchanged.

If a "treatment" is tried in a sample of stores and evaluated by comparison of data for the test period with data from the same stores for a base period—should the comparison be in terms of the relative change in actual sales, the relative change in lbs/100c, the relative change in the proportion of sales of the particular commodity to the sales of a group of commodities, or on some other basis? This question of concepts should not be confused with questions of experimental or sampling technique. The merits of each concept should be considered, assuming no errors of any kind, and then a choice should be made on the basis of joint consideration of the utility of the concepts and the experimental or survey problems associated with each.