R E P O R T   R E S U M E S

ED 019 075 JC 680 154

SAMPLING AND COMMUNITY RELATIONS.

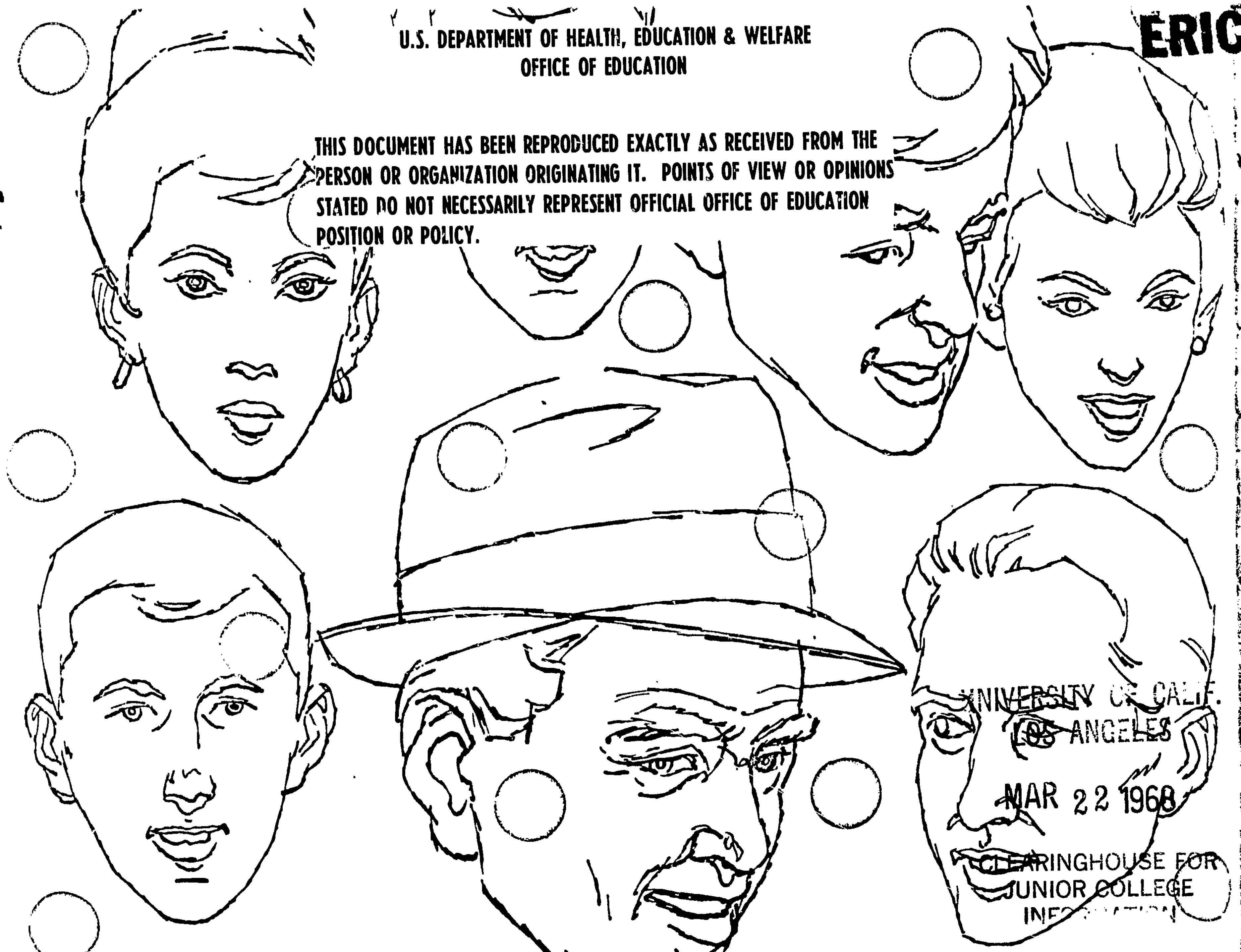
BY- WELCH, TIMOTHY

PUB DATE NOV 65

EDRS PRICE MF-$0.25 HC-$0.28 5P.

DESCRIPTORS- *JUNIOR COLLEGES, *COMMUNITY ATTITUDES, COMMUNITY STUDY, COMMUNITY SURVEYS, *STATISTICAL SURVEYS, *SAMPLING,

ACCURATE ASSESSMENT OF COMMUNITY OPINION AND REACTION DEPENDS UPON ADEQUATE SAMPLING TECHNIQUES. VOTER REGISTRATION LISTS PROVIDE ONE SOURCE FROM WHICH TO DRAW SAMPLES. RESEARCHERS MUST CONSIDER THE COSTS OF LARGE SAMPLES IN RELATION TO TWO QUESTIONS--(1) THE CLOSENESS WITH WHICH THE TOTAL POPULATION IS TO BE MATCHED BY THE SAMPLE, AND (2) THE DESIRED DEGREE OF CERTAINTY OF THIS MATCHING. IN ORDER TO DETERMINE DESIRABLE SAMPLE SIZE, A FORMULA AND TABLE OF "CERTAINTY FIGURES" HAS BEEN DEVELOPED, AND EXAMPLES OF THEIR USE ARE PRESENTED. IF A SAMPLE OF ANY SIZE IS TO BE VALID, RANDOM SELECTION IS ESSENTIAL. VALIDITY OF DATA ARE ALSO DEPENDENT ON THE CARE USED IN DEVISING A SURVEY INSTRUMENT AND UPON THE METHOD OF HANDLING THE PROBLEM OF NONRESPONSE. THIS ARTICLE IS PUBLISHED IN "JUNIOR COLLEGE JOURNAL," VOLUME 36, NUMBER 3, NOVEMBER 1965. (WO)



SAMPLING AND COMMUNITY RELATIONS

Interdependence of Institution and Public Calls for Valid Lines of Communication

By Timothy Welch

"Feedback" is elementary. When talking to a second party, the first party looks for feedback signs of approval or disapproval. When talking to an audience, a speaker looks for feedback indicators of the crowd reaction.

If there are only two people whose feedback must be attended, sampling obviously isn't necessary. When the universe becomes 20, or 200, or 200,000, however, feedback gets lost. Yet the reaction of these individuals, whether as voting taxpayers or consumers who vote through their purchases, is vital.

But the "drop-in" vote is badly skewed. That is, telephone calls and letters received in anger can't be regarded as a "normal" sample. At best, some indication is being given about the flash point of certain members of the population: what it takes to ignite certain people to the point of complaining.

A related objection can be made to the sampling of advisory committees set up by the college. Although the committees are invaluable, they are by nature specialized, and could constitute a good sample only if they were accurately matched to the population strata. This, of course, isn't the reason for advisory committees, and they are not so matched.

The temptation is to give too much weight to these, in a sense, "free samples."

All this is only to restate the obvious: that sampling is a continual process, basic to existence, which goes on every waking hour. The choice is whether the sampling should be disciplined or haphazard. Certainly for public information, or public relations, or community service, or community relations, a valid sampling program is basic. The voters—political or economic—are out there, and their votes must be counted.

The needs for a simple sampling procedure are various. Periodic, general-impression polls, or "image" polls, with continuing questions, are valuable for long-run estimates of how the sampling agency is seen. Postcard polls can gauge opinion on fast-rising issues. A rotating mailing list for newsletters, drawn as a sample, will achieve equal distribution among the voting population.

Sampling procedures are necessarily statistical, and for a measure of any complexity should be turned over to professional sampling organizations. But for samples of a modest, day-to-day scope, it isn't practicable to pay a sampling organization.

The purpose of this article, then, is to suggest sampling methods for local agencies with limited budgets.

The best definition of a school district's population, it is assumed, is the most current list of registered voters. This is an opinion-forming group, and one which acts on its opinions. But even more to the point, it is the "employer group" whose opinion is crucial. For a public body dependent upon taxes, registered voters are, for practical purposes, the true population.

Note that a moot ethical point is couched here. Shouldn't *all* taxpayers who contribute to the support of a school district be considered? In a sense, anyone who pays any taxes for the support of a school is an "employer." Further, shouldn't *all* citizens over 21 years be included?

Perhaps. But the objection can be answered by noting that, as a practical matter, a high correlation exists between voting registration and property-tax payers who primarily support public schools. As for the ethics of excluding nonvoters, it may be argued that they are obviously not so interested in making decisions. A democracy is the product of those who take part in its operation. In any population, it is the active decision makers whose opinions should be sought, for their opinions are likely to be dominant, informed, and directive.

The most current available list of registered voters can be the frame, then, from which the sample is drawn. (A "frame" is simply a list of the sample units in the population.) Frequency of the list's changing should be noted to keep its information accurate.

County clerks in many instances sell such lists of registered voters. In Santa Clara County, California, for example, lists of the county's 300,000 registered voters may be purchased for 60 cents per thousand names. In Fresno County, California, lists cost 50 cents per thousand names, broken into some 760 precincts with between 175 to 200 names for each precinct. Some counties, such as Fresno, restrict the sale of this list to people using it for a political purpose, such as information about a bond issue. It may be successfully argued by a public taxing agency that communication with its employers is always political. Purchase of the entire list can still be expensive, however. In this event, a sub-sample of the population (by precincts) can be drawn so that an entire voting list needn't be purchased.

To help keep the sample list current, the names should be replaced (for instance, "Local taxpayer," or "Junior college taxpayer," or "Friend of Blank College") to lessen the problem of nonresponse occasioned by people moving.

Choosing the Sample

Given a voter registration list of 300,000 names, how is a sample drawn?

How many need to be reached to comprise an adequate sample? A balance must be struck between squandering money on too large a sample, or getting too small a sample for any confidence. (It should be noted, by the way, that the accuracy of an estimate of a proportion does not depend upon the size of the population—only upon the size of the sample.[1])

Stated one way: how high a probability of being accurate, within stated percentages, is desired?

Stated another way: how precise a sampling, separate from the question of your confidence in the sampling, is required?

A guide is needed to allow the sampler to decide how much accuracy he wants to purchase for his sample, and to define just what confidence he can place in a sample of a given size.

Once systematic bias has been eliminated by a well-drawn random sample,[2] the question of sample size and confidence interval becomes mathematical. So the mathematics of sampling must be briefly touched upon.

To determine the size of the sample to be drawn, we have to know two things: (1) How precise do we want to be? That is, how closely do we want our sample characteristics to match the real population characteristics? (2) And, how certain do we want to be that our sample falls within the above limits?

If, for example, there were a 60-40 split in the party registration, suppose we wanted a .95 probability that the error of estimate would not exceed .02 units. In other words, that our sample would draw 58 to 62 per cent Democrats, and, conversely, 42 to 38 per cent Republicans—and that we could be 95 per cent sure that our sample would do this. (The other 5 per cent of the time, we could miss badly, and draw heavy proportions from one party or the other, and misrepresent the political opinions of the population.)

A simple formula gives our sample size, as soon as we make the policy decision of how much accuracy and how much certainty we want. The formula is:

$$n - \frac{{}^{\bullet}0^2 pq}{e^2}$$

Here '0 is the certainty figure (95 per cent),[3] p and q are the corresponding proportions of Democrats (60 per cent) and Republicans (40 per cent) in the registered voting list, and e is the accuracy figure.[4]

A further table yields the "certainty figures" to insert in the formula:

95 per cent	1.96
97.5 per cent	2.24
99.8 per cent	3.09

So, then, to be 95 per cent certain that we get in our sample $\pm$ 2 per cent of the actual number of Democrats and Republicans, the formula for the sample size is

$$n = \frac{(1.96)^2 \quad (.60)\ (.40)}{(.02)^2} = 2{,}294 \text{ in sample}$$

If we wanted to be 97.5 per cent certain that we were $\pm$ 2 per cent accurate:

$$n = \frac{(2.24)^2 \quad (.60)\ (.40)}{(.02)^2} = 3{,}011 \text{ in sample}$$

If we wanted to be 99.8 per cent certain that we were $\pm$ 2 per cent accurate.

$$n = \frac{(3.09^2\ (.60)\ (.40)}{(.02)^2} = 5{,}729 \text{ in sample}$$

It should be obvious from the above examples that the cost of the sample rises as we increase both certainty and accuracy.[5]

The point should be made by now that two things may be "purchased" in a sample: precision or accuracy, and degree of certainty. The extent to which we desire each determines the size of the sample, and its cost.[6]

A Typical Survey

As an illustration, let us run a postcard survey which might be made to find out the voters' feelings about a bond proposal to build a combination football-baseball stadium.

Although party registration is a potent variable (along with age, sex, education level, income, religion, race), let's say that we're interested in another variable: voting behavior in the last school election.

Since turnout for these elections rarely reaches the figures of a national election, let's say that 11 per cent of the registered voters of the county, or school district, voted.[7] How large a sample would we need to be 99.74 per cent certain that our sample proportion would represent $\pm$ 3 percent of the *real* 11 per cent of the voting list who actually voted? The sample size is given by

$$n = \frac{{}^{\bullet}0^2 pq}{e^2} - \frac{(3)^2\ (.11)\ (.89)}{(.03)^2} = 979$$

which determines that 979 voters must be taken from the frame for this sample.

An absolutely strict requirement is that they be chosen randomly: the validity of the entire operation rests on this point. How does one choose these 979 voters randomly?

A table of random numbers, available in most statistics books, is the easiest manner of choosing the 979 voters. Assuming a frame of 760 precincts and 150,000 registered voters, we go through the random number table in a systematic manner for three-digit numbers from 001 to 760 until there are 979 precinct numbers listed.

Then, with each precinct list (in which the random number of voters appears on the front), we go down the random number table for the first three-digit number appearing within that interval. For example, if there are 182 voters in the 546th precinct, and our precinct draw has told us that we must have two voters from the 546th precinct, we look for the first two three-digit numbers between 001 and 182.

These names and addresses are the result.

When the 979 addresses have been assembled, the mailing list is complete. If one person is pulled twice in the same sample, his response must be given uble weight. Random sampling demands that each ember have an equal opportunity of being drawn each draw, and not eliminated from the list.

The problem of nonresponse must be touched on iefly. It is almost certain, of the 979 postcards we nd, that there will not be 979 returned.

From a statistical point of view, this must be ndled carefully, for nonrespondents may be a speal type of person whose voting behavior may be lated to our study. It will not suffice to simply raw additional names. That is, these nonrespondnts are saying something to us by not responding, nd this message is as important to us as that on e returned postcards, if we are to keep our sample andom and representational.

In the example at hand, nonresponse would probbly best be handled by setting up such a category. t would be simplest to treat the message of their onresponse as indifference to the issue, which could ndicate that they are a part of the 89 per cent who id not vote in the last school election and/or that hey don't care enough about the issue to vote in an lection which concerns it.

The response will almost certainly be greater than 1 per cent of the 979 sampled, indicating that more eople will have opinions and reply than voted in the ast election.

If greater rigor is desired, the problem of nonesponse may be met with telephone and field visit ollow-up. In this event, the sub-sample is weighted, nd the results projected. For instance, if there are 00 nonrespondents, and a random sub-sample of wenty is contacted, their replies will be multiplied y five. For such a sub-sample, a very simple samling method, such as taking every fifth nonrespondnt, is adequate.[8]

The Means of Asking

A further difficulty lies in the area of questionaire construction. It is assumed that the postcard urvey used here, or likely to be used, will be short nd brief, with check boxes for simple, nonambiguus answers. If, however, a long form is sent, then he matter of good questionnaire construction should e examined to obtain sharp, focused results.[9]

The techniques of handling interviews is also a eld to itself, which requires detailed consideration o avoid methodological problems for accurate data eduction and analysis.[10] Both questionnaire and nterview design are beyond the scope of this article.

Statistics and sampling are the study of certainty, elling us, in an uncertain world, how sure we may e of such-and-such.

In the future, the interdependence of an institution nd its various publics can only grow, creating the need for valid lines of communication. Certainly it is sound psychology to ask people how they feel about something before doing it to them or for them. Because of the large populations involved, disciplined sampling procedures are the necessary means of asking.

---

[1] Hoel, Paul G. *Elementary Statistics.* New York: John Wiley & Sons, Inc., 1960, p. 98.

[2] The two important points which enable the binomial distribution to be applied to practical sampling problems best define "random": that each sample be independent of the other, and that the probability of any unit being chosen from the universe be exactly the same. See Hoel, p. 70.

[3] Hoel, *op. cit.*, pp. 96-97. This equation is derived from the more familiar

$$\delta p = \frac{pq}{n} \quad \text{or} \quad n = \frac{pq}{\delta p^2}$$

[4] Note that although estimates of p and q as 60 per cent and 40 per cent are possible here (since we know the gross figures from the registered voting list), it would be possible to draw a sample on the assumption that both Democrats and Republicans appeared in equal numbers in the population. In other words, $p = .50$ and $q = .50$, which would reduce the sample size necessary to 1,225, since there would then be an equal chance of drawing either a Democrat or a Republican in a random draw. However, since we know from the voting list that there are *not* equal numbers of each party in the population, the sample size has to increase to be 95 per cent certain that we approximate the real 60-40 percentage split within the actual population. Drawing as though there were a 50-50 chance of drawing a Republican would result in a bias favoring Democrat representation in the sample.

[5] It costs more to increase the accuracy in the error of estimate in a sample than it does to increase the certainty. Taking 95 per cent certainty at ± 4 per cent accuracy as a base, it would require a sample of 2,305 to get ± 2 per cent precision at 95 per cent certainty, but would require only 706 to 97 per cent certainty at ± 4 per cent accuracy.

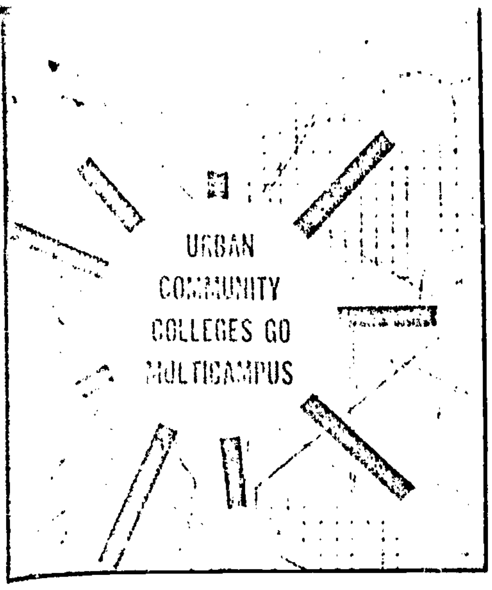
[6] Two assumptions are crucial to this method of estimating p in this formula: that a large sample is involved, and that the normal curve approximation to the binomial distribution is satisfactory. For justification of both assumptions, see Hoel, *op. cit.*, pp. 77-78, 97. A good background may be obtained from the first six chapters of Hoel.

[7] Since district and county lines are not always coterminous, a suitable frame must be constructed by purchasing appropriate precinct lists which approximate the school district.

[8] Yates, Frank. *Sampling Methods for Censuses and Surveys.* 3rd ed. New York: Hafner Publishing Company, 1960, p. 108.

[9] A good source is Selltiz *et. al., Research Methods in Social Relations* (rev. ed.). New York: Holt, Rinehart, and Winston, 1962, chapters 7 and 10.

[10] Máccoby, Eleanor E., and Maccoby, Nathan, in Lindsey (ed.), *Handbook of Social Psychology.* Vol. I. Reading, Massachusetts: Addison-Wesley Publishing Company, 1954, pp. 449-487.



Volume 36 ■ November, 1965 ■ Number 3

junior college journal

THE MAGAZINE FOR JUNIOR COLLEGE EDUCATION ■ AMERICAN ASSOCIATION OF JUNIOR COLLEGES

Roger Yarrington: Editor ■ Louise Scott Wrigley: Editorial Assistant ■ Richard T. Chinn: Business Manager ■ Thomas T. Gladden: Art Consultant

Junior College Journal is published monthly from September through May (December-January issue combined). Subscription: $4.00 per year, 50 cents per copy. Group subscriptions available to junior colleges at $3.00 a year each. Advertising rates on request. Second-class postage paid at Washington, D. C. by the American Association of Junior Colleges, 1315 Sixteenth Street, N.W., Washington, D. C. 20036. Average circulation for preceding twelve months: Total copies printed, 12,000; paid circulation 11,475, sales through agents, none; free distribution 240; total copies distributed, 11,715. Circulation of this issue: Total copies printed, 14,000; paid circulation, 13,270; sales through agents, none; free distribution, 240; total copies distributed 13,510.