AN EVALUATION OF TERMINAL DIGIT SOCIAL SECURITY NUMBER SAMPLING IN THE VETERANS ADMINISTRATION ANNUAL PATIENT SURVEY

William F. Page and Glenda E. Wright, Veterans Administration

INTRODUCTION

The Veterans Administration (VA) conducts annual surveys of the more than 100,000 persons in its facilities. The majority of the survey involves the sampling of patients by the terminal digit of their social security number (SSN). In this paper we present an analysis of the data collected in 1975, 1976, and 1977. The primary purpose of this analysis was to determine whether terminal digit SSN sampling was an adequate procedure for the VA Annual Patient Survey.

METHODOLOGY

Even though this study of the sampling has been limited to VA hospitals only, this still provides a sizeable population -- VA hospitals had about 75,000 patients remaining on each of the survey days in 1975, 1976, and 1977. The sampling procedure selects patients with terminal SSN digits of 1 or 5, so that about 15,000 patients were sampled in each of the three years. The sampling of both 1's and 5's allows us to compare 1's versus 5's as two half-samples. However, for this presentation half-samples will not be discussed; half-samples are presented in [1].

For survey reporting purposes the 161 hospitals have been divided into about 300 independent entities which we have termed "sampling units." Each "sampling unit" within a hospital independently selects the sample patients and reports the results. For brevity, we will sometimes refer to a sampling unit as a "hospital."

Within each sampling unit one can easily compute the expected number of patients to be sampled. In the combined sample it is simply 20% of the total number of bed occupants in a sampling unit. We can then compare the observed and expected number of patients sampled. In any one hospital we should not (and did not) expect exact agreement between the two figures since the sampling process allows for random variations. However, across the whole system the differences between the observed and expected number of patients sampled should average out.

We need to measure the discrepancy between the observed and expected sample in a way which allows us to compare hospitals of different sizes. This is a classical statistical problem and is solved here by computing standardized scores or "Z-scores".

One computes a standardized score by subtracting the expected value from the observed value and dividing by the standard deviation. This expresses the differences between the theoretical expected value and the actual observed value in "standard deviation units." The formula for the expected value is p times N, where p is the proportion expected, i.e. p = 0.20 for the total sample, and N = number of patients in the sampling unit. The formula for the standard deviation is: $\sqrt{Np(1 - p)}$. Hence,

$$Z = \frac{S - (.2*N)}{SQRT\ (N*.2*.8)}$$

where S = number of patients sampled and
N = total number of patients in sampling unit.

When the observed values follow a normal distribution the standardized scores or Z-scores are also normal, but they have been automatically rescaled to have a mean of 0 and a standard deviation of 1. Since the binomial distribution (which is the mathematical model for the sampling results) is approximately normal for large samples, our Z-scores should have a normal distribution.

RESULTS

For the remainder of this paper we will assume that terminal SSN digits are uniformly distributed. Data from the Social Security Administration [2] indicate that this is the case, and our study of hospital discharge records [1] indicates the same thing. The rest of the results section concentrates on the actual distribution of sampling results in terms of the Z-score distribution.

Figure 1 shows the observed distribution of Z-scores (the histogram) versus the expected distribution (the smooth standard normal distribution). We see that the discrepancy between the observed and expected distributions is greatest near the center of the distribution. In particular, there is an excess of observed Z-scores in the -1.0 to 0.0 range. This will be treated in the discussion section.

Table 1 shows the mean and variance of the observed z-scores for 1975, 1976, and 1977. The most striking result is that all of the mean z-scores are negative (their expected value is zero). The variances are in the range of 1.0.

Table 1

Mean and Variance of Z-Scores

Year	Mean	Variance
1977	-0.102	0.798
1976	-0.159*	1.072
1975	-0.085	1.118

*p< .05

FIGURE 1

DISTRIBUTION OF ACTUAL VS EXPECTED Z-SCORES

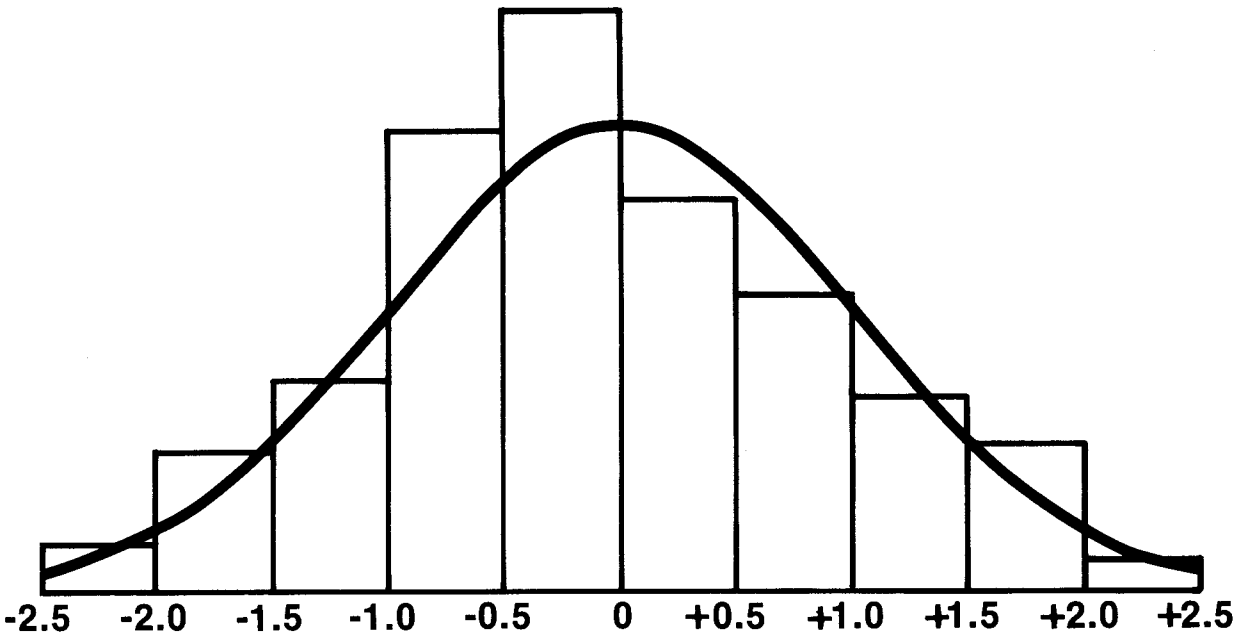
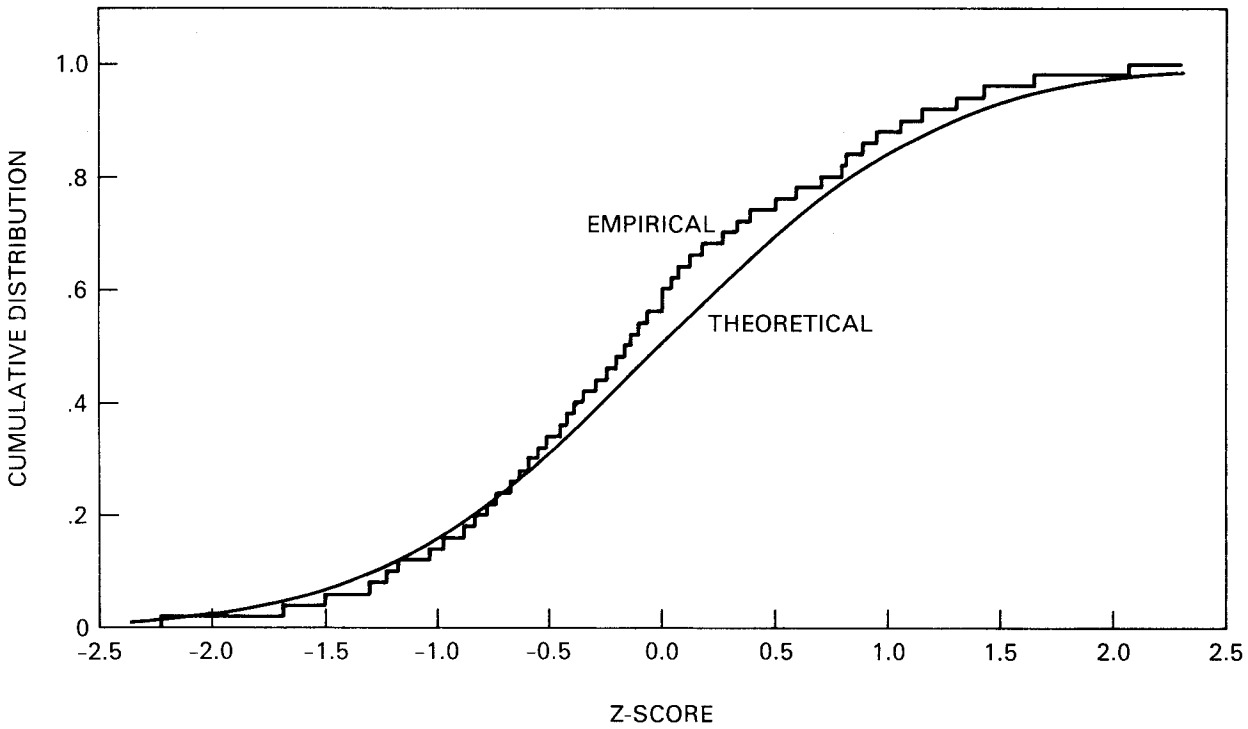


FIGURE 2
COMPARISON OF THEORETICAL NORMAL (0, 1) DISTRIBUTION AND EMPIRICAL Z-SCORE DISTRIBUTION: 1977 CENSUS



Although we undertook this analysis in an exploratory, rather than confirmatory, spirit, we did perform one-sample Kolmogorov-Smirnov goodness-of-fit tests. These tests have the advantage that they test the entire shape of the distribution (figure 2). We also performed the Lilliefors test for normality which simply compares the observed distribution to a normal curve with the observed mean and variance rather than a standard normal curve. We found significant differences ($p<0.05$) between the observed z-score distribution and the standard normal curve in all 3 years; however, we found no significant differences between the observed z-score distribution and the normal curve with observed mean and variance (Table 2).

Table 2

Kolmogorov-Smirnov and Lilliefors Tests of Normality of Z-Scores by Terminal Digit, 1975-1977

Year	Kolmogorov-Smirnov	Lilliefors
1977	0.1081**	0.0592
1976	0.0924*	0.0301
1975	0.0852*	0.0476

*$p < 0.05$
**$p < 0.01$

From these results we conclude that the observed z-score distribution has roughly a normal shape, but its mean is too negative. This negative mean z-score is evidence that, on the average, hospital patients are undersampled, i.e. negative z-scores occur when fewer patients are sampled than expected. In the next section we attempt to explain how this undersampling might have taken place.

DISCUSSION

In this section we develop an explanatory model for the undersampling observed. The model is a simple one, and depends on just three quantities: the actual number of 1's and 5's sampled, the expected number sampled, and the "true" number of 1's and 5's in the hospital.

We postulated that one of two situations occurred during the hospital sampling. Each will be discussed separately. First, suppose that the true percentage of 1's and 5's is low. For example, if we have a 100 bed hospital we expect 20 sampled patients, even though there may be only 15 patients with terminal SSN digits 1 or 5. We hypothesize that the hospital, which is expecting to sample 20 patients, will almost surely sample all 15 patients with SSN 1 or 5. Because the true number of 1's and 5's is low, we assert that there will be very few instances where the full 15 are not found due to a diligent search for an expected 20 patients.

On the other hand, if the true percentage of 1's and 5's is about 20 % or higher, we assert that the sampling could be different. Let us suppose that in the hypothetical 100 bed hospital 25 patients have SSN ending in 1 or 5. Then, if they sample only 23 we would not expect a careful search for the "missing" 2 patients since the expectation was to find only 20 patients. That is to say, undersampling should be more prevalent in the situation where the true number of 1's and 5's is high.

We can relate this hypothesis back to the Z-score distribution of figure 1. According to our hypothesis, in the first situation (true percentage of 1's and 5's is low) we should see little undersampling. On the graph this is the situation where z-scores are quite negative (say -1.0 or less). Here the observed and expected distributions agree quite well and we do not see undersampling.

In the second situation (true percentage of 1's and 5's about 20% or higher) we expected undersampling. Thus to show up we would expect this undersampling in the portion of the graph with z-scores -1.0 to +1.0. Furthermore, if there were such undersampling it would mean too many negative scores by our z-score definition. This is precisely what we observed. There are too many slightly negative z-scores in the middle of this graph, a fact that we attribute to selective undersampling.

CONCLUSION

Overall the SSN terminal digit sampling procedure has performed well, producing only an estimated 1.5% undersampling. We have hypothesized that the undersampling was selective, and was a function of the true percentage of 1's and 5's.

Based on the conclusions we have two suggestions to improve the sampling procedure, particularly the undersampling. The first suggestion is to require that the selection process be checked by an independent staff member. Secondly, we would also recommend that the patient selection be cross-checked against a centralized roster.

REFERENCES

1. Page, William F. and Wright, Glenda. Controller Monograph Technical Series No. 1, Sampling in the Veterans Administration Annual Patient Census, Veterans Administration, Office of Controller, 1979.

2. Actuarial Note #62, Social Security Administration.