COUNTING COUPONS

C. H. Proctor, North Carolina State University

1. Introduction

The following account of a sampling experiment offers an illustration of statistical methods applied to a workplace setting. We will first describe the purpose of the study and then discuss the preliminary results and some particularly interesting applications of a viewpoint on adjacency correlation through the use of Smith's b. There then follows some mathematical formulations of cost and variance functions involving the cost coefficients and variance factors that the experiment permitted us to estimate. We conclude with a few overall findings.

2. Background

For those who plan and execute coupon promotions it is most helpful to know how many coupons of each kind were redeemed. There is a code number on the coupon and the coupons, once they arrive at the final collection point, need to be sorted by this code and then counted. This operation requires careful work with attentive eyes and nimble fingers. For purposes of paying the retailers for the coupon, they need only be sorted and counted by face value and product, while the sorting by coupon code is an additional operation. The purpose of the research reported here was to design and analyze an experiment to compare various methods of sampling to provide estimates of proportions for each of the coupon codes.

3. Experimental Design

In actuality we carried out two separate but complementary experiments. One was done to compare: (I) complete sorting followed by weigh counting and (II) systematic sample removal by hand followed by sorting and hand counting. The other experiment compared three mechanically assisted sample removal methods. The methods are called subdivisions, splittings and yardstick. These were all designed to operate on a, so called, batch of about 2000 coupons in order of their processing.

The subdivisions method called for distributing by handfulls the coupons of a batch into 50 pots -- plastic containers fastened to a board and color coded in two Latin square arrangements with 5 colors. A color was randomly chosen and the contents of those 10 pots became the sample. The splittings methods made four groups with coupons still in processing order. A coin flip chose either groups 1 and 3 or, if tails, groups 2 and 4. The chosen groups were split and the new four groups were treated by a coin flip as the first four were. Two more coin flips gave a 1/16 sample which was then sorted and counted. The yardstick method also was applied to the coupons in their processing order. A stick with 10 prongs was randomly situated in one of eight positions then dropped on the stack of 2000 coupons. The coupons hit by a prong and the next four down the stream were left but the next 10 coupons were removed for the sample.

The experiment began by defining on the coupon processing stream twelve, so called, sectors of coupons as experimental units. There were from 100,000 to 250,000 coupons in a sector. All coupons were from two client-companies. Half of the sectors were already sorted by face value while half were not and for every sector collected from small mail (such as arriving directly from the retailer) there were two sectors of large mail. The twelve sectors are thus the 12 = 2x2x3 combinations of these conditions.

In each sector three batches were extracted without any more elaborate instruction than to be spread over the sector and not to pick some unusual coupons. The resulting 36 batches were first listed coupon by coupon as to code and also by a physical characterization of each coupon. Then the yardstick method was applied, followed by splittings and finally the subdivisions method. Since we had knowledge of the batch proportions the sample deviations were computed from the known population mean and estimates were easily constructed for the biases and variances of the three methods. The times required for sample removal were also recorded.

The coupons in the sector as a whole were subjected to systematic removal of every fifth coupon and the removed coupons were further distributed systematically into 10 systematic 1-in-50 subsamples. These were sorted and hand counted separately -- the even-numbered ones by one worker and the odd-numbered ones by another. The remaining four-fifths of the coupons in the sector were then sorted and weigh-counted. A number of packs were also randomly selected for a hand-count audit -- counting until agreement on two counts.

4. Preliminary Results

The experiment required roughly six months to plan, six months to run and record and a year to analyze. The following resume of results thus fails to depict the circuitous path we took to get them, but perhaps we could briefly indicate here some of the guiding principles.

One statistical problem concerned the departure from randomness of kinds of coupons along the actual processing stream. One might expect coupons of a given code to be adjacent to one another in the stream more so than chance might allow. The problem was how to represent this grouping tendency. In addition to this local grouping, it also might be expected that over the whole sector there would be trends in the relative prominence of various codes. Again the question was how to represent this.

These tendencies, local grouping with long term drifts, would affect the sampling variance. That is, systematic sample removal should give better results than random removal. Two nonsampling sources of error were the sorting operation and the counting operation. Since the weigh-counting to hand counting comparison was very much of concern we focussed our design on it. In the weighing method a few coupons are first put on the scales to establish a per-coupon average weight and then the whole pack is weighed to give a total count. On the other hand, our knowledge of the variability in the sorting operation was obtained by noting more variability between the even-numbered subsamples and odd-numbered ones than within either set. This, we decided, had to be due to different workers using slightly different categories for their sorting. It seemed likely that minor codes could be classed with codes having the same overall characteristics by some workers but kept separate by others. The data are available on about 500 codes per sector and could be used to study this situation in more detail and we leave it for the future.

In addition to the study of variability introduced by sample removal, by sorting and by counting we also had to keep track of the time spent in each operation. In particular we needed to estimate the cost coefficients required to show how much of a reduction in variance could be secured by expenditures of various kinds -- on a higher rate of sample removal, on a more expensive counting method or on a more careful kind of sample removal. Finally, we needed to judge how well such coefficients were being estimated. That is, do they seem to be the same from sector to sector?

5. Model Formulation

The quantities to be estimated are the code proportions in a sector, the π_c for $c=1,2,\ldots,C$, where C is from 300 to 500. The data are counts from a sample of n coupons, selected from the N in a sector. The sample proportion $p_c = n_c/n$ is used to estimate π_c. If the sample is a simple random one then the n_c will have a multi-hypergeometric distribution. For a particular code, c, the distribution of p_c is hypergeometric. When dealing with removal rates of one-in-ten or one-in-five the finiteness of the population is relevant. However, since most of the values of π_c are below .03 the quantity $\pi_c(1-\pi_c)$ may be replaced by just π_c. Compare .03(.97) = .0291 to .0300 or .003(.997) = .002991 to .003000.

Further notation on parameters estimated in the experiment can be given in summary form as:

C_1 = Systematic subsample removal cost coefficient

C_2 = Sorting and weigh-counting cost coefficient

C_3 = Sorting and hand counting cost coefficient

C_4 = Batch sampling cost coefficient

C_m = Coupon subsampling cost coefficient for method m (m=subdivisions, splittings, yardstick)

C_T = Total cost (excluding fixed costs) of survey

V_2 = Variance factor for hand counting coupons

V_3 = Variance factor for weigh-counting coupons

V_m = Batch subsampling variance factor for method m

σ_B^2 = Between batch variance component

b = Smith's b for batch variances

β = Smith's b for systematic subsampling

N = Sector size (number of coupons) or size of target population

π_c = Proportion of code c coupons in population

p_c = Sample proportion of code c

Our basic viewpoint on the observed estimate p_c can be expressed by the following model equation:

$$p_c = \pi_c + (P_c - \pi_c) + (p_c' - P_c) + (p_c - p_c')$$
$$= \pi_c + \delta_c + \varepsilon_c + \gamma_c \ , \qquad (5.1)$$

where P_c is the actual sample proportion of code c coupons, p_c' is based on the correct number, but of nominally code c coupons, while p_c is based on possibly incorrectly counted and incorrectly sorted code c coupons. That is, δ_c is sampling error, ε_c is sorting error and γ_c is counting error. We, basically, leave ε_c aside as mentioned above; γ_c is given two rather extreme possibilities: weigh counting and vigilant hand counting; while most of our attention is focussed on δ_c.

6. Variance and Cost Expressions

It was found that, over the more or less 300 codes in a sector, variation among systematic subsamples was very close to what would have been variation among simple random subsamples. We fit the following variance function to the subsample variances over the various codes:

$$V(\delta_c) \doteq \pi_c(n^{-\beta} - N^{-\beta}) \ , \qquad (6.1)$$

in which the case of $\beta = 1$ represents the random or hypergeometric case of $\pi_c(1-n/N)/n$, while $\beta>1$ is to be expected if coupons are grouped by code. We suspect that sorting error may have upset the finding of β clearly greater than 1, but variances among subsamples were close to those for random grouping.

A further attempt to estimate β was the following. From the coupon by coupon listing of the batches we constructed nested systematic samples. That is, we started with the 1-in-5 sample, then nested two 1-in-10 samples in it, five 1-in-50 in the 1-in-10's, and so forth. We then fit Smith's b to the mean squares from the resulting nested ANOVA. Values around 1.1 and 1.2 were common but so were 1.0 and even some .9's. Again, the evidence suggests that a β-value of greater than 1 is correct but that it is not much above 1.

The full variance expression for systematic subsample removal thus becomes

$$V(p_c) = \pi_c(V_i/n + n^{-\beta} - N^{-\beta}) \ , \qquad (6.2)$$

where V_i for i=2 or 3 is the counting variance appropriate to weigh counting (i=2) or hand counting (i=3) and n is sample size. The corresponding survey cost formula is:

$$C_T = C_1N + C_in, \ i=2,3 \qquad (6.3)$$

where C_1, C_2 and C_3 is expressed in minutes per coupon while C_T is in minutes.

For the batch subsampling methods we found variances of the observed sample proportions around the known total batch proportions. These variances were found for the physically defined categories of coupons rather than for coupon codes. That is, the code proportions were too small relative to the 2000 or so coupons in a batch so that we used a seven-category typology of coupons [(1) large, rectangular and thin; (2) medium size, irregular shape and thick; and so forth to (7) other]. These variances are likely larger than variances by code because the physical removal methods will be most affected by physical differences in the coupons.

We had some information on batch variability so we decided to allow for batch subsampling in the variance formula although it may not be used in practice. This is a more conventional situation for use of Smith's b in which our batches of 2000 can be broken into 10 of 200 and these in turn broken into 2 of 100 each and so forth. We estimated b to be around b = .5 to b = .7 for clusters of size around M_1 = 2000 coupons.

The variance expression for the mechanical removal methods thus becomes:

$$V(p_c) = \pi_c[(\sigma_B^2 + V_m)/n_1 + V_i/n_1n_2] \ , \qquad (6.4)$$

where n_1 equals the number of batches removed and n_2 is the number of coupons removed from each sampled batch. The quantity σ_B^2 is obtained as $\sigma_B^2 = M_1^{-b} - N^{-1}$ where M_1 is the size of a batch ($M_1 = 2000$ in our experiment) and b is Smith's b ($b \approx .5$ or .6 or .7). The factors V_m were obtained from the deviations of sample proportions from the known batch proportions.

An approximate cost function for these methods is

$$C_T = (C_4 + C_m)n_1 + C_i n_1 n_2 \,, \tag{6.5}$$

for i = 2, 3 and m = subdivisions, splittings and yardstick. It is possible to use these functions to derive an optimum size of subsample as:

$$n_{2,opt} = [V_i/C_i 0]^{\frac{1}{2}} [(C_4 + C_m)/(\sigma_B^2 + V_m)]^{\frac{1}{2}}. \tag{6.6}$$

This formula has a rather limited range of applicability in terms of values of n_2. The experimental condition had $n_2 = 400$ for the subdivisions method, $n_2 = 125$ for the splittings and $n_2 = 100$ for the yardstick method. With i=2 for weigh counting then $n_{2,opt} \doteq 50$, while i=3 for vigilant hand counting $n_{2,opt} \doteq 8$. This implies that the subdivisions method should perhaps use pots of 10 colors rather than 5, there should be one further splitting and the yardstick method should skip 5 coupons then remove five coupons at each prong hit. This also usually implies that almost all batches will be subsampled.

7. Final Results

Coupon handling is a competitive business and so knowledge of the actual values of the variance and cost coefficients can constitute somewhat of a business advantage. At least the client who sponsored this research had reason to hope for some such return on his investment. For this reason we will not furnish numerical values of the various cost coefficients and variance factors.

It may serve to illustrate the overall findings to show the effect of changing sample design on total variable cost in worker days and on the percent sampling coefficient of variation. Consider a total volume of around ten million coupons. When systematic removal is used at a rate of 1-in-10 these two quantities of interest are 583 days with a 0.64% sampling coefficient of variation (CV); with rate of 1-in-20 they are 500 days and 0.92% and with 1-in-50 removal they become 450 days and 1.48%. On the other hand the yardstick removal method applied to all batches (i.e., to the whole processing stream) costs 333 days and yields a 1.47% sampling CV and when applied to one in 10 batches cost 33 days and gives a 4.95% sampling CV. Choice of sampling method will thus depend on how important it is to know the success of a promotion accurately.