University of Illinois at Chicago
School of Public Health
Division of Epidemiology and Biostatistics

Technical report#:2006-005
May 2006

Title: **Cost-Effective Unbiased Estimation of Exponential Mean Life - Comparison of Ranked Set Sampling vis-a-vis Simple Random Sampling**

Author(s): Karabi Sinha

Affiliation(s): Division of Epidemiology and Biostatistics, University of Illinois at Chicago

Cost-Effective Unbiased Estimation of Exponential Mean Life - Comparison of Ranked Set Sampling vis-a-vis Simple Random Sampling

Karabi Sinha *

Division of Epidemiology and Biostatistics, University of Illinois at Chicago

Abstract: This paper studies the problem of unbiased estimation of the mean of exponential life distribution of items which are to be put on test. The performance of the mean based on Simple Random Sampling is compared against that based on Ranked Set Sampling - taking into account possible cost structures. Also taken into account is the "profit" accrued by the "remaining life" of items not completely destroyed by the tests. Statistical analysis reveals that most often one would gain by implementing the RSS procedure.

Key Words and Phrases: Exponential Life Distribution, BLUE of Mean Life, Order Statistics, Cost Function, Relative Efficiency

1 Introduction

The purpose is to estimate the mean of an exponentially distributed population. In particular, if we are examining the lifetime of electric bulbs, then given X = lifetime of a bulb $\sim$ Exp(θ), our objective is to estimate θ, the mean life. Traditionally, we draw a random sample of size n from the population and use the sample mean as the estimate of θ. Formally, let $X_1, X_2, ..., X_n$ be a random sample of size n from Exp(θ). Then random sampling suggests $\hat{\theta} = \sum_i X_i/n = \bar{X}$ where $\bar{X}$ is the sample mean and is known to be the UMVUE of θ. Further, $Var(\hat{\theta}) = Var(\bar{X}) = \theta^2/n$.

At this point, let us define a cost structure as follows:

(i) C_0 = Setup cost per unit, per unit lifetime;

(ii) C_m = manufacturing cost per unit, per unit lifetime;

(iii) C_1 = operating cost per unit time;

(vi) C_s = selling cost per unit, per unit lifetime;

(v) C_p = profit per unit, per unit lifetime.

*1603 W Taylor St. Rm 951, Chicago IL 60612; tel. no. 312-355-3611; email address: karabi@uic.edu

We set $C_s = C_m + C_p$. Then under random sampling of n items,

$$E[\text{Total Cost under SRS}] = E[C_T : SRS] = n(C_0 + C_m)\theta + C_1 E[X_{n:n}] \tag{1.1}$$

where $X_{n:n}$ is the largest order statistic in a sample of size n.. It is well-known that for an exponential population Exp(θ),

$$E[X_{n:n}] = \theta[1 + 1/2 + 1/3 + \ldots, +1/n]. \tag{1.2}$$

Now we will look at an alternative method of experimentation based on what is known as "Ranked Set Sampling" (RSS) and to keep similarity in terminology, we will refer to the former as a method based on Simple Random Sampling (SRS). Often times, measurement of experimental items can be quite expensive and time-consuming. And thus, a focal point of interest becomes devising a feasible and informative sampling strategy with as few actual measurements as possible. In particular, when the variable of interest from the experimental units can be more easily ranked than quantified, it turns out that for estimation of the population mean, a concept by McIntyre (1952), namely, RSS is highly beneficial and much superior to the standard SRS. The basic concept behind RSS can be briefly described in a very general context as follows: Draw n independent random samples of size n each. RSS then uses only one observation from each set; namely $X^1_{1:n}$, the lowest observation from the first set, then $X^2_{2:n}$, the second lowest observation from the second set and so on, finally, $X^n_{n:n}$, the largest observation from the last set of n observations. This can be shown in a table as follows:

Table1. Display of n^2 sampling units in n sets of n each

$X^1_{1:n}$	$X^1_{2:n}$	.	.	.	$X^1_{n:n}$
$X^2_{1:n}$	$X^2_{2:n}$	.	.	.	$X^2_{n:n}$
.	.	.	.	.	.
.	.	.	.	.	.
$X^n_{1:n}$	$X^n_{2:n}$	.	.	.	$X^n_{n:n}$

The important point to emphasize here is that although RSS requires as many as n^2 experimental or sampling units, only n of them, namely,$\{X^1_{1:n}, X^2_{2:n}, \ldots, X^n_{n:n}\}$ are actually measured, thus making a comparison of this sampling strategy with SRS of the same size meaningful. In our context, we will use a slight modification of the usual RSS method. Since we are referring to experiments on electric bulbs, to start with, we will assemble r^2 bulbs and divide these into r sets of r each. Then we will proceed and stop as in the case of RSS but collect data as follows:

$$[X^1_{1:r}];\ [X^2_{1:r},\ X^2_{2:r}]; \ldots, [X^r_{1:r},\ X^r_{2:r},\ \ldots, X^r_{r-1:r},\ X^r_{r:r}]. \tag{1.3}$$

We refer to the notations used in Table 1 and use r in place of n. In effect, therefore, using this method, we are in a position to obtain *actual* life times on $r(r+1)/2$ items. We now set $n = r(r+1)/2$ so that under both the methods, we have lifetime data on the *same* number of items, though under RSS, we start with $r^2[> n]$ items. The focus of this paper is to compare these two sampling strategies namely, SRS and RSS, in their efficiency in estimating θ and also in terms of cost minimization.

2 Analysis of RSS Data

For a sample of size r from an exponential lifetime distribution with mean θ, it is well-known that $U_1 = r[X_{1:r}], U_2 = (r-1)[X_{2:r} - X_{1:r}], U_3 = (r-2)[X_{3:r} - X_{2:r}], \ldots, U_r = [X_{r:r} - X_{r-1:r}] \overset{\text{iid}}{\sim} Exp(\theta)$ Note that $[X_{i:r};\ i = 1, 2, \ldots, r]$ refer to the order statistics in a sample of size r. We will now use the RSS sample in (1.3) to provide an estimate of θ.

From the properties of the spacings described in terms of $U_1, U_2, \ldots$ in the above, it turns out that (1.3) yields $r(r+1)/2$ iid Exp(θ) observations, namely, $rX^1_{1:r}; rX^2_{1:r}, (r-1)(X^2_{2:r} - X^2_{1:r}); \ldots; rX^r_{1:r}, (r-1)(X^r_{2:r} - X^r_{1:r}), \ldots, (X^r_{r:r} - X^r_{r-1:r})$.

Therefore, the average of these observations serves as the umvue for θ. Further, it has variance given by $2\theta^2/r(r+1) = \theta^2/n$, since our choice of r satisfies the relation $r(r+1) = 2n$.

Thus we find that both SRS and RSS give some kind of "mean" estimate and variances of these estimates are equal! This leads us to the obvious question, "Why RSS, then?". To evaluate this, we will be looking at two aspects in the next section:

(a) Expected duration of experiment under each strategy

(b) Cost aspect.

3 Expected Duration

In RSS, each set is an independent random sample and it may so happen that the second failure occurs in the second set even before the first failure occurs in the first set. In such a case we would stop the experiment in the second set to prevent any further failure but still keep it running in the first set waiting for the first failure there! This makes it clear that in RSS, the experiment continues till the maximum of $[X^1_{1:r},\ X^2_{2:r},\ \ldots, X^r_{r:r}]$ is observed. Define Z = maximum $[X^1_{1:r},\ X^2_{2:r},\ \ldots, X^r_{r:r}]$ so that E[Duration of Experiment in RSS] = $E(Z)$. Now, in view of independence of different sets of RSS samples, $F(z) = Prob[Z < z] = Prob.[X^1_{1:r} < z, X^2_{2:r} < z, \ldots, X^r_{r:r} < z] = F_1(z)F_2(z)\ldots F_r(z)$ where $F_i(z)$ refers to the cdf of the i^{th} order statistic in a sample of size r from Exp(θ) population. Since Z is non-negative, we can use the formula: $E(Z) = \int_0^\infty \bar{F}(z)dz$, where $\bar{F} = 1 - F$ and, moreover, we can set: $\bar{F}(z) = \sum_i \bar{F}_i - \sum_i \sum_j \bar{F}_i \bar{F}_j + \ldots$. On the other hand, we may use (1.2) to evaluate E(Duration of Experiment under SRS) = E($X_{n:n}$).

Numerical computations for expected duration have been carried out (using Mathematica) for values of r up to 10 and shown in Table 2. We see that the expected duration is consistently smaller in RSS than SRS based on $n = r(r+1)/2$ items.

Table 2: Computation of E[Z] and E[X_n] for Exp(1) population

r	2	3	4	5	6	7	8	9	10
$E[Z]$	1.58	1.96	2.23	2.45	2.63	2.78	2.91	3.02	3.13
E[X_n]	1.83	2.45	2.93	3.32	3.65	3.93	4.17	4.39	4.59

4 Cost Aspect and Comparison of Strategies

Once again, we refer to RSS and examine the cost structure for this method as against (1.1). Using the cost components introduced in Section 1, we may write

$$E[C_T : RSS] = r^2(C_0 + C_m)\theta + C_1 E[Z] - C_s E[RL], \tag{4.1}$$

where RL = Residual Life of all items under RSS and it contributes a *negative* cost of C_s per unit life time. It is clear that $RL = [X^1_{2:r} + X^1_{3:r} + \ldots + X^1_{r:r} - (r-1)X^1_{1:r}] + [X^2_{3:r} + X^2_{4:r} + \ldots + X^2_{r:r} - (r-2)X^2_{2:r}] + \ldots + [X^{r-1}_{r:r} - X^{r-1}_{r-1:r}]$ and it now follows, in view of the above, that

$$E[RL] = \theta[1 + 2 + \ldots + (r-1)] = \theta\{r(r-1)/2\}. \tag{4.2}$$

We are now in a position to compare the two strategies. It follows that RSS would be profitable whenever E[C_T:RSS] < E[C_T:SRS]. The condition is equivalent to:

$$r^2(C_0 + C_m)\theta + C_1 E(Z) - C_s\theta\{r(r-1)/2\} < n(C_0 + C_m)\theta + C_1 E[X_{n:n}]. \tag{4.3}$$

Note that $n = r(r+1)/2$. Hence, we can simplify the condition for superiority of RSS over SRS as

$$(C_s - C_m - C_0)\{r(r-1)/2\} + C_1[E(X_{n:n} - E(Z)] \ > \ 0. \tag{4.4}$$

We may now interpret (4.4) as follows: Whenever the profit per item, per unit lifetime $[C_p = C_s - C_m]$ is more than the set-up cost, irrespective of the value of r, RSS proves to be a better strategy than SRS. In situations where the profit is less than the set-up cost, we may solve for r from (4.4) and make our recommendations. For r = 5, for example, the above condition is equivalent to: $10(C_p - C_0) + 0.6896C_1 > 0$.

5 Conclusion

The purpose of this work is to make an attempt to promote the use of RSS as an alternative and better method than the traditional SRS. The so-called drawbacks of RSS may be over-ridden by using this approach. We have studied a simple problem with a structured cost function to derive the benefits of RSS as against SRS.

REFERENCES

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research.* **3**, 385-390.

Dell, T. R. and Clutter, J. L. (1972). Ranked set sampling with order statistics background. *Biometrics.* **28**, 545-553.

Kaur, A., Patil, G. P. , Shirk, S. J. and Taillie, C. (1996). Environmental sampling with a concomitant variable: A comparison between ranked set sampling and stratified simple random sampling. *Journal of Applied Statistics.* **23**, 231-255.

Sinha, Bimal K., Sinha, Bikas K. and Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions.* **14**, 223-240.