Ranked Set Sampling

Ranked Set Sampling

Edited by

Munir Ahmad
National College of Business Administration & Economics, Pakistan

Muhammad Hanif
Lahore University of Management Sciences, Pakistan

and

Hassen A. Muttlak
King Fahd University of Petroleum and Minerals, Saudi Arabia

CAMBRIDGE SCHOLARS
PUBLISHING

Ranked Set Sampling,
Edited by Munir Ahmad, Muhammad Hanif and Hassen A. Muttlak

This book first published 2010

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-2494-1, ISBN (13): 978-1-4438-2494-1

CONTENTS

FOREWORD

The Editorial Board of Pakistan Journal of Statistics, in its meeting held in September, 2009 to collect papers on one topic and publish in the form of a book. Each paper has been refereed by at least three experts actively engaged in “Ranked Set Sampling”. This book is the first in the series and we hope that in future, we shall be collecting papers and publishing in the form of books.

—Munir Ahmad
Editor-in-Chief PJS and
Rector, NCBA&E, Lahore, Pakistan

PREFACE

Recently attention is being paid to the basic concepts of "Ranked Set Sampling" and there are a number of papers available in the literature. New techniques and approaches are being studied recently but there is no collection of papers that provide recent developments in the area. The motivation of this book is the amount of recent papers published by various authors on the topic of "Ranked Set Sampling" in Pakistan Journal of Statistics.

Our main objective is to present before a wider audience on the work done on "Ranked Set Sampling" during the last decade and to motivate statisticians in this part of the world to work on some latest statistical technologies developed in various aspects of sampling. This book does not show any overlap with the current developments in the area, instead it has added new approaches to the area.

We are indebted to all the authors of the papers for their enormous hard work in preparation of the papers and their referees for the quality work they have done and to Mr. M. Imtiaz and Mr. M. Iftikhar for excellent job of reproduction / composition of papers and setting in the proper format.

—Munir Ahmad, Muhammad Hanif and Hassen A. Muttlak

CHAPTER ONE

STRATIFIED RANKED SET SAMPLE

HANI M. SAMAWI,
GEORGIA SOUTHERN UNIVERSITY

Abstract

Stratified simple random sampling (SSRS) is used in certain types of surveys because it combines the conceptual simplicity of simple random sample (SRS) with potentially significant gains in efficiency. It is a convenient technique to use whenever we wish to ensure that our sample is representative of the, population and also to obtain separate estimates for parameters of each subdomain of the population. If sampling units in a study can be easily ranked compared to quantification, McIntyre (1952) proposed to use the mean of n units based on a ranked set sample (RSS) to estimate the population mean, and observed that it provides an unbiased estimator with a smaller variance compared to SRS of the same size n.

In this paper we introduce the concept of stratified ranked set sample (SRSS) for estimating the population mean. SRSS combines the advantages of stratification and RSS to obtain an unbiased estimator for the population mean, with potentially significant gains in efficiency. 'The conclusion of this study is that by using SRSS the efficiency of the estimator relative to SSRS and SRS has strictly increased. Results from uniform distribution are given. Computer simulated results on other distributions are also given. An example using real data is presented to illustrate the computations.

Key Words

Simple random sample, stratified random sample, ranked set sample, stratified ranked set sample, order statistics.

1. Introduction

Ranked set sampling (RSS) was introduced by McIntyre (1952) to estimate the pasture yield. RSS procedure involves randomly drawing n sets of n units each from the population for which the mean is to be estimated. It is assumed that the units in each set can be ranked visually. From the first set of n units, the unit ranked lowest is measured. From the second set of n units, the unit ranked second lowest is measured. The process is continued until from the $n-th$ set of n units the $n-th$ ranked unit is measured. Talcahasi and Wakimoto (1968) warned that in practice the number of units which are easily ranked cannot be more than four.

A stratified simple random sample (SSRS), (for example see Hansen et al. 1953) is a sampling plan in which a population is divided into L mutually exclusive and exhaustive strata, and a simple random sample (SRS) of n_h elements is taken and quantified within each stratum h. The sampling is performed independently across the strata. In essence, we can think of a SSRS scheme as one consisting of L separate simple random samples.

A stratified ranked set sample (SRSS) is a sampling plan in which a population is divided into L mutually exclusive and exhaustive strata, and a ranked set sample (RSS) of n_h elements is quantified within each stratum, $h=1, 2, \ldots, L$. The sampling is performed independently across the strata. Therefore, we can think of a SRSS scheme as a collection of L separate ranked set samples.

In this paper, we introduce the concept of SRSS to estimate the population mean. This study showed that the estimator using SRSS is at least more efficient than the one using SSRS. In Section 2, we describe some sampling plans, discuss estimation of population mean using these plans, and give some useful definitions and general results and results for the uniform distribution. Simulation results from non-uniform distributions are given in Section 3. In Section 4, we illustrate the method using real data. The discussion is given in Section 5.

2. Samples and Estimation of Population Mean

Suppose that the population is divided into L mutually exclusive and exhaustive strata. Let:
$X^*_{h11}, X^*_{h12}, \ldots, X^*_{h1n_h}; X^*_{h21}, X^*_{h22}, \ldots, X^*_{h2n_h}; \ldots; X^*_{hn_h1}, X^*_{hn_h2}, \ldots, X^*_{hn_hn_h}$ be n_h independent random samples of size n_h each one is taken from each

stratum $(h=1,2,...,L)$. Assume that each element X_{hij}^{*} in the sample has the same distribution function $F_h(x)$ and density function $f_h(x)$ with mean μ_h and variance σ_h^2. For simplicity of notation, we will assume that X_{hij} denotes the quantitative measure of the unit X_{hij}^{*}. Then, according to our description $X_{h11}, X_{h21},..., X_{hn_h1}$ could be considered as the SRS from the $h-th$ stratum. Let $X_{hi(1)}^{*}, X_{hi(2)}^{*},..., X_{hi(n_h)}^{*}$ be the ordered statistics of the $i-th$ sample $X_{hi1}^{*}, X_{hi2}^{*},..., X_{hin_k}^{*}$ $(i=1,2,....n_k)$ taken from the $h-th$ stratum. Then, $X_{h1(1)}, X_{h1(2)},..., X_{hn_h(n_h)}$ denotes the RSS for the $h-th$ stratum. If $N_1, N_2,..., N_L$ represent the number of sampling units within respective strata, and $n_1, n_2,..., n_L$ represent the number of sampling units measured within each stratum, then $N=\sum_{h=1}^{L} N_h$ will be the total population size, and $n=\sum_{h=1}^{L} n_h$ will be the total sample size.

2.1 Definitions, notations and some useful results

The following notations and results will be used throughout this paper. For all $i, i=1,2,...,n_h$ and $h=1,2,...,L$, let $\mu_h = E(X_{hij}), \sigma_h^2 = Var(X_{hij})$, $\mu_{h(i)} = E(X_{hi(i)})$, $\sigma_{h(i)}^2 = Var(X_{hi(i)})$, for all $j=1,2,...,n_h$ and let $T_{h(i)} = \mu_{h(i)} - \mu_h$.

As in Dell and Clutter (1972), one can show easily that for a particular stratum $h,(1=1,2,...,L)$,

$$f_h(x) = \frac{1}{n_h} \sum_{i=1}^{n_h} f_{h(i)}(x),$$

and hence $\sum_{i=1}^{n_h} \mu_{h(i)} = n_h \mu_h$, $\sum_{i=1}^{n_h} T_{h(i)} = 0$ and $\sum_{i=1}^{n_h} \sigma_{h(i)}^2 = n_h \sigma_h^2 - \sum_{i=1}^{n_h} T_{h(i)}^2$

The mean μ of the variable X for the entire population is given by

$$\mu = \frac{1}{N} \sum_{h=1}^{L} N_h \mu_h = \sum_{h=1}^{L} W_h \mu_h \tag{2.1.1}$$

where $W_h = \dfrac{N_h}{N}$.

If within a particular stratum, h, we suppose to have selected SRS of n_h elements from N_h elements in the stratum and each sample element is measured with respect to some variable X, then the estimate of the mean ph using SRS of size n_h is given by

$$\bar{X}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{hi1}. \tag{2.1.2}$$

The mean and variance of $\bar{X}_h$ are known to be $E\left(\bar{X}_h\right) = \mu_h$ and $Var\left(\bar{X}_h\right) = \dfrac{\sigma_h^2}{n_h}$ respectively, assuming N_h's are large enough. The estimate of the population mean μ using SSRS of size n is defined by

$$\bar{X}_{SSRS} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{X}_h = \sum_{h=1}^{L} W_h \bar{X}_h \tag{2.1.3}$$

The mean and the variance of $\bar{X}_{SSRS}$ are known to be $E\left(\bar{X}_{SSRS}\right) = \mu$ and

$$Var\left(\bar{X}_{SSRS}\right) = \sum_{h=1}^{L} W_h^2 \left(\frac{\sigma_h^2}{n_h}\right) \tag{2.1.4}$$

respectively, assuming N_h's are large enough.

If within a particular stratum h, we suppose to have selected RSS of n_h elements from N_h elements in the stratum and each sample element is measured with respect to some variable X, then the estimate of the mean μ_h using RSS of size μ_h is given by

$$\bar{X}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} X_{hi(i)} \tag{2.1.5}$$

It can be shown that the mean and variance of $\bar{X}_{h(n_h)}$ are $E\left(\bar{X}_{h(n_h)}\right) = \mu_h$ and

$$Var\left(\bar{X}_{h(n_h)}\right) = \frac{\sigma_h^2}{n_h} - \frac{1}{n_h^2}\sum_{i=1}^{n_h} T_{h(i)}^2, \tag{2.1.6}$$

respectively, assuming N_h's are large enough. Therefore, the estimate of the population mean μ using SRSS of size n is defined by

$$\bar{X}_{SRSS} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{X}_{h(n_h)} = \sum_{h=1}^{L} W_h \bar{X}_{h(n_h)} . \qquad (2.1.7)$$

It can be shown straightforward algebra that the mean and the variance of $\bar{X}_{SRSS}$ are $E\left(\bar{X}_{SRSS}\right) = \mu$ (i.e., and unbiased estimator) and

$$Var\left(\bar{X}_{SRSS}\right) = \sum_{h=1}^{L} W_h^2 \left(\frac{\sigma_h^2}{n_h} - \frac{1}{n_h^2} \sum_{i=1}^{n_h} T_{h(i)}^2 \right), \qquad (2.1.8)$$

respectively, assuming N_h's are large enough.

Therefore, the relative efficiency of the estimator of the population mean μ using SRSS with respect to the one using SRSS can be defined by

$$RE = \frac{Var\left(\bar{X}_{SSRS}\right)}{Var\left(\bar{X}_{SRSS}\right)} = \frac{1}{\left\{ 1 - \frac{1}{Var\left(\bar{X}_{SSRS}\right)} \left(\sum_{h=1}^{L} \frac{W_h^2}{n_h^2} \sum_{i=1}^{n_h} T_{h(i)}^2 \right) \right\}} \qquad (2.1.9)$$

2.2 Results for the uniform distribution

Assume that a population of size N, with a variable X has a uniform distribution $U(U,\theta)$. Suppose we can divide this population into L strata with respect to some characteristics in the population. If we let $N_1, N_2, ..., N_L$ represent the number of sampling units within respective strata, and $n_1, n_2, ..., n_L$ represent the number of selected sampling units from respective strata, then $N = \sum_{h=1}^{L} N_h$ and $n = \sum_{h=1}^{L} n_h$. Assume that the random variable X_h has distribution $U(0,\theta_h)$. Thus, $\mu_h = E(X_h) = \frac{\theta_h}{2}$ and $\sigma_h^2 = Var(X_h) = \frac{\theta_h^2}{12}$. Also, $\theta = \frac{1}{N} \sum_{h=1}^{L} N_h \theta_h = \sum_{h=1}^{L} W_h \theta_h$.

The mean and variance of the estimate $\bar{X}_{SSRS}$ of the population mean μ using SSRS of size n ware $E\left(\bar{X}_{SSRS}\right) = \frac{\theta}{2} = \mu$ and

$$Var\left(\bar{X}_{SSRS}\right)=\sum_{h=1}^{L} W_h^2\left(\frac{\theta_h^2}{12n_h}\right) \tag{2.2.1}$$

respectively.

The mean and variance of the estimate $\bar{X}_{SRSS}$ of the population mean μ using SRSS of size n are $E\left(\bar{X}_{SRSS}\right)=\frac{\theta}{2}=\mu$ and

$$Var\left(\bar{X}_{SRSS}\right)=\sum_{h=1}^{L} W_h^2\left(\frac{\theta_h^2}{6n_h\left(n_h+1\right)}\right) \tag{2.2.2}$$

Also, if $n_h \geq 2$ then,

$$RE=\frac{Var\left(\bar{X}_{SSRS}\right)}{Var\left(\bar{X}_{SRSS}\right)}=\frac{\sum_{h=1}^{L} W_h^2\left(\frac{\theta_h^2}{n_h}\right)}{2\sum_{h=1}^{L} W_h^2\left(\frac{\theta_h^2}{n_h\left(n_h+1\right)}\right)}>1\,, \tag{2.2.3}$$

which implies that SRSS gives a more efficient unbiased estimator for the uniform population mean compared to SSRS.

3. Simulation Study

The normal and exponential distributions are used in the simulation. Sample sizes $N=10,20$ and 30 and number of strata $L=3$ are considered. For each of the possible combination of distribution, sample size and different choice of parameter, 2000 data sets were generated. The relative efficiencies of the estimate of the population mean using SRSS with respect to SSRS, SRS and RSS are obtained. All computer programs were written in Borland TURBO BASIC.

3.1 Result of the Simulation Study

The values obtained by simulation are given in Table 1. Our simulation indicates that estimating the population means using. SRSS is more efficient than using SSRS or SRS. In some cases, when the underlying distribution is normal with ($\mu_1=1.0,\ \mu_2=3.0,\ \mu_3=5.0$), the simulation indicates that estimating the population mean using SRSS is even more efficient than RSS.

Table 1: The relative efficiency of the simulation results

Distribution function	n	$RE\left(\bar{X}_{SRSS}, \bar{X}_{SSRS}\right)$	$RE\left(\bar{X}_{SRSS}, \bar{X}_{SRS}\right)$	$RE\left(\bar{X}_{SRSS}, \bar{X}_{RSS}\right)$
Normal	10	2.04	7.59	1.50
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4,$	20	3.19	11.63	1.29
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	30	4.45	16.30	1.25
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$				
Normal	10	2.08	3.48	0.72
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4,$	20	3.19	5.70	0.65
$\mu_1 = 1.0, \mu_2 = 2.0, \mu_3 = 3.0$	30	4.42	7.57	0.67
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$				
Normal	10	2.08	7.15	1.28
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4,$	20	3.38	10.50	1.19
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	30	4.32	13.94	1.10
$\sigma_1 = 1.0, \sigma_2 = 1.1, \sigma_3 = 1.2$				
Exponential	10	2.82	3.55	1.28
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4,$	20	3.04	3.78	0.96
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	30	3.50	4.17	0.86
Exponential	10	1.95	2.15	0.73
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4,$	20	2.85	3.25	0.71
$\mu_1 = 5.0, \mu_2 = 10.0, \mu_3 = 15.0$	30	3.53	4.15	0.74

4. Example: Body Mass Index Data

In Table 2 we present three sample of size 7 each, from baseline interview data for the Iowa 65+ Rural Health Study (RHS), which is a longitudinal cohort study of 3,673 individuals (1,420 men arid 2,253 women) ages 65 or older living in Washington and Iowa countries of the State of Iowa in 1982. This study is one of four supported by the National Institute on Aging and collectively referred to as EPESE, (Established Populations for Epidemiologic Studies of the Elderly), National Institute on Aging, 1986.

In the Iowa 65+ RHS there were 33 diabetic women aged 80 to 85, of whom 14 reported urinary incontinence. The question of interest is to estimate the mean body mass index (BMI) of diabetic women. The BMI is the ratio of the subject's weight (kilograms) divided by height (meters) squared. Note that, the BMI may be different for women with or without urinary incontinence. Thus, here is a situation where stratification might work well. The 33 women were divided into two strata, the first consists of those women without urinary incontinence and the second consists of those 14 women with urinary incontinence. Four samples of size $(n = 7)$ each were drawn from those women using SSRS, SRSS, RSS and SRS. Note that in case of SRSS and RSS the selecting samples are drawn with replacement. The calculated values of BMI are given in Table 2. These calculations indicate the same pattern of conclusions that were obtained earlier, and illustrate the method described in Section 2.

Table 2: Body Mass Index Samples of Diabetic Women Aged 80 to 85 Years with and without Urinary Incontinence

	SRS	**RSS**		**SSRS**	**SRSS**
	18.88	18.88	Stratum 1	23.45	23.45
	19.76	22.88		28.95	23.46
	20.57	23.45		30.17	30.10
	25.66	24.38	Stratum 2	19.61	19.61
	26.01	26.30		24.07	24.38
	28.95	27.31		27.49	31.31
	33.52	36.65		33.52	31.95
Estimated Mean	24.77	25.69		26.95	26.15
Standard Error	2.03	2.06		1.72	1.67

5. Discussion

The BMI data is a good example where we need stratification to find an unbiased estimator for the population mean of those diabetic women aged 80 to 85 years. Since the 33 women were divided into two strata, the first consists of those women without urinary incontinence and the second consists of those women with urinary incontinence. It is clear that the mean of the BMI in each stratum will be different. Also, women can be ranked visually according to their BMI. In this situation we recommend using SRSS to estimate the mean BMI of those women. SRSS will give an unbiased and more efficient estimate of the BMI mean. Moreover, SRSS can provide an unbiased and more efficient estimate for the mean of each stratum.

Remark: We could not find a closed form for optimal allocation of units and also for optimal allocation of resources for n_h using SRSS. However, the near optimal allocation can be obtained from the formulae obtained by using SSRS, for example see, Hansen et al. 1953.

References

Brock, D.B., Wineland T. Freeman, D.H., Lemke, J.H., Scherr, P.A. (1986). Demographic characteristics. In: Established Population for Epidemiologic Studies of the Elderly. Resource Data Book, Cornoni Huntley, J. Brock D.B., Ostfeld, A.M., Taylor, J.O. and Wallace, R.B. (eds). *National Institute on Aging*, NM Publication No. 86-2443. U.S. Government Printing Office, Washington, D.C.

Dell T.R. and Clutter J.L. (1972). Ranked Set Sampling Theory with Order Statistics Background. *Biometrics* 28, 545-555.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sampling Survey Methods and Theory*, Vols. 1 and 2, Wiley, New York.

McIntyre, G.A. (1952). A method of unbiased selective sampling, using ranked sets. *Australian J. Agri. Research* 8, 385-390.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the stratified sampling by means of ordering. *Ann. Inst. Statist.* Math., 20, 1-31.

CHAPTER TWO

USING RANKED SET SAMPLING FOR HYPOTHESIS TESTS ON THE SCALE PARAMETER OF THE EXPONENTIAL AND UNIFORM DISTRIBUTIONS

WALID ABU-DAYYEH,
UNIVERSITY OF UAE
AND HASSEN A. MUTTLAK,
DEAKIN UNIVERSITY

Abstract

The concept of ranked set sampling (RSS) was suggested by McIntyre (1952). Many authors including Takahasi and Wakimoto (1968). Stokes (1980) and Muttlak and McDonald (1990) have used RSS in estimation.

In this paper we will obtain the uniformly most powerful test (UMPT) and the likelihood ratio test (LRT) in case of exponential distribution and the UMPT in case of uniform distribution, using simple random sample (SRS) and then we will adapt the statistics of these tests to construct new tests using RSS. It turns out that the use of RSS gives much better results in terms of the power function compared to SRS.

Key Words

Ranked set sampling; simple random sample; power of the test, UMPT and LRT.

1. Introduction

In many applications it is very difficult or expensive to measure the sampling units, but the units can be ranked with out any cost. It turns out that in such cases the use of RSS gives better estimate of the population mean compared to the SRS. In agricultural and environmental studies, it is possible to rank the sampling units without actually measuring them. For some such applications see Cobby et al. (1985), Muttlak and McDonald (1992), Johnson et al. (1993) and Patil and Taillie (1993). For the sampling method of RSS see Stokes (1986).

Many other uses of RSS have been studied in the literature. Takahasi and Wakimoto (1968) independently suggested the same method that was considered by McIntyre (1952). They proved that the mean of RSS is an unbiased estimator of the population mean with smaller variance than the variance of the sample mean of a SRS with the same sample size. Stokes (1980) discussed the estimation of the variance based on RSS. Muttlak and McDonald (1990a, 1990b) developed RSS theory when the sampling units are selected with size based probability of selection.

The object of this paper is to obtain the UMPT for the one sided alternative and the LRT for the two sided alternative in case of the exponential distribution and the UMPT for the two sided alternative in case of uniform distribution using SRS and will adapt these tests to RSS data. It turns out that the tests based on RSS have higher power than the corresponding tests based on the SRS.

2. Exponential Distribution

Let $X_1, X_2, ..., X_n$ be a random sample from the exponential distribution with pdf

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{if } x \geq 0 \\ 0 & \text{Otherwise} \end{cases}$$

We are interested in testing the hypotheses

$$H_0 : \theta = \theta_0 \text{ vs. } H_\alpha : \theta > \theta_0 \quad (1)$$

It is well known that the UMPT of size α for testing (1) is given by

$$\phi UMPT = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i > C_\alpha \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Without lost of generality we may take $\theta_0 = 1$. Then $c_\alpha = \frac{\chi^2_{2n,1-\alpha}}{2}$, where χ^2_m is the chi-square distribution with m degrees of freedom. The power of the test (2) is given by

$$\beta_{\phi UMPT}(\theta) = P_\theta\left(\sum_{i=1}^{n} X_i > \frac{1}{2}\chi^2_{2n,1-\alpha}\right) = P_\theta\left(W > \frac{1}{\theta}\chi^2_{2n,1-\alpha}\right),$$

where W is distributed χ^2_{2n}

To obtain the test using RSS let $X_{11}, X_{12}, ..., X_{1n}$; $X_{21}, X_{22}, ..., X_{2n}$; ……; $X_{n1}, X_{n2}, ..., X_{nn}$ be the n groups of n independent random variables all with the same cumulative distribution function $F(x)$. Let $X_{i(1)}, X_{i(2)}, ..., X_{i(n)}$ be the order statistics of the variables $X_{i1}, X_{i2}, ..., X_{in}$ in the i-th group $(i = 1, 2, ..., n)$ Then $X_{1(1)}, X_{2(2)}, ..., X_{i(i)}, ..., X_{n(n)}$ denotes the ranked set sample, where $X_{i(i)}$ is the i-th order statistic in the i-th group. To simplify the notation, $X_{i(i)}$ will be denoted by Y_i through out this paper.

To test the same hypothesis (1) using the RSS we propose the following test

$$\phi_1 = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} Y_i > d \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where d is determined so that the test ϕ_1 has size α. To obtain the value of d, we need the distribution of $\sum_{i=1}^{n} Y_i$ under H_0. For this purpose we consider the following transformation:

$Z_1 = Y_1, Z_2 = Y_1 + Y_2, Z_3 = Y_1 + Y_2 + Y_3, ..., Z_n = \sum_{i=1}^{n} Y_i$. We know that $Y_1, Y_2, ..., Y_n$, are independent random variables with joint pdf:

$$g_\theta(y_1, y_2, ..., y_n) = \begin{cases} \left\{\prod_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!}\left[1 - e^{-y_i/\theta}\right]^{i-1}\right\}\frac{1}{\theta^n} e \\ \quad -\sum_{i=1}^{n} (n-i+1)\, y_i/\theta, y_i > 0, i = 1, ..., n \\ 0 \qquad \text{Otherwise} \end{cases} \quad (4)$$

Then the joint pdf of $Z_1, Z_2, ..., Z_n$ is given by

$$h_\theta(z_n, z_2, ..., z_n) = g_\theta(z_1, z_2 - z_1, z_3 - z_2, ..., z_n - z_{n-1}),$$

which implies that the pdf of Z_n is:

$$k_\theta(z_n) = \int_0^{z_n} \int_0^{z_{n-1}} ... \int_0^{z_n} g_\theta(z_1, z_2 - z_1, z_3 - z_2, ..., z_{n-1}) d_{z_1} d_{z_2 ... d_{z_{n-1}}} \tag{5}$$

Therefore the power function of the test (3) is given by

$$\beta_{\phi_1}(\theta) = P_\theta\left(\sum_{i=1}^{n} Y_i > d\right) \int_d^\infty k_\theta(z_n) d_{z_n},$$

To find d, we need to solve

$$\beta_{\phi_1}(1) = \alpha = \int_d^\infty k_{\theta=1}(z_n) d_{z_n} \tag{6}$$

It is not easy to find d for general n and α. Therefore we will find d for $n = 3, 4,$ and 5 and $\alpha = 0.05$. For $n = 3$, the pdf of z_3 is

$$k_\theta(z_3) = \frac{27}{2\theta} e^{-3z_3/\theta} \left\{ e^{2z_3/\theta} + 16e^{2z_3/\theta} - 4 \right\}$$
$$- \frac{27}{2\theta^3} e^{-3z_3/\theta} \left\{ 13\theta^2 + 10\theta e z_3 + 8\theta_{z_3} e^{z_3/\theta} + 2z_3^2 \right\} \tag{7}$$

For $\alpha = 0.05$ and $n = 3$ we found that $d = 5.532$, using a computer mathematical program. Therefore, the power of this test is given by

$$\beta_{\phi_1}(\theta) = P_\theta\left(\sum_{i=1}^{3} Y_i > d\right) = \int_{5.532}^\infty k_\theta(z_3) dz_3, \tag{8}$$

Similarly, the power function can be written for $n = 4$ and 5. Table (1) shows the results for $\alpha = 0.05$ and $n = 4$ and 5 and different values of θ.

It appears from Table (1) that the power of the tests $\phi UMPT$ and ϕ_1 increases as θ increases and also as n increases and that the power of ϕ_1 is larger than the power of $\phi UMPT$ i.e. using RSS gives higher power of the test compared to SRS.

Table (1): Values of $\beta_{UMPT}(\theta)$ and $\beta_{\phi_1}(\theta)$ for different values of θ and sample sizes $n = 3, 4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi UMPT}(\theta)$			$\beta_{\phi_1}(\theta)$		
	n = 3	$n = 4$	$n = 5$	$n = 3$	$n = 4$	$n = 5$
1.10	0.076	0.079	0.083	0.080	0.087	0.094
1.25	0.122	0.134	0.146	0.138	0.164	0.191
1.50	0.211	0.242	0.258	0.327	0.402	0.402
2.00	0.391	0.458	0.518	0.503	0.642	0.762
3.00	0.650	0.739	0.807	0.807	0.924	0.977
4.00	0.790	0.868	0.918	0.924	0.984	0.998
5.00	0.866	0.928	0.961	0.967	0.996	0.999
10.0	0.974	0.992	0.998	0.999	0.999	0.999

Next we will consider the LRT for testing the hypothesis

$$H_0 : \theta = 1 \text{ vs. } H_\alpha : \theta \neq 1 . \quad (9)$$

It is well known that the LRT of size a is given by

$$\phi_{LRT} = \begin{cases} 0 & \text{if } \frac{\chi^2_{2n,\alpha/2}}{2} < \sum_{i=1}^{n} X_i < \frac{\chi^2_{2n,1-\alpha/2}}{2} \\ 1 & \text{Otherwise} \end{cases}$$

which implies that its power function is given by

$$\beta_{\phi UMPT}(\theta) = \quad 1 - P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{2} < \sum_{i=1}^{n} X_i < \frac{\chi^2_{2n,1-\alpha/2}}{2}\right)$$

$$1 - P_\theta\left(\frac{\chi^2_{2n,\alpha/2}}{\theta} < W < \frac{\chi^2_{2n,1-\alpha/2}}{\theta}\right)$$

where $W = \sum_{i=1}^{n} X_i$ is distributed as χ^2_{2n}.

To test the same hypothesis using the RSS, the following test is proposed:

$$\phi_2 = \begin{cases} 0 & \text{if } k_1 < \sum_{i=1}^{n} Y_i < k_2 \\ 1 & \text{Otherwise} \end{cases}$$

The power function of the test ϕ_2 is

$$\beta_{\phi_2}(\theta) = 1 - P_\theta\left(k_1 < \sum_{i=1}^{n} Y_i < K_2\right) = 1 - \int_{K_1}^{k_2} k_\theta(z_n)\,dz_n ,$$

where $k_\theta(z_n)$ is defined in (5). To obtain the test of size α we need to find k_1 and k_2 to satisfy

$$\beta_{\phi_2}(1) = \alpha = 1 - \int_{k_1}^{k_2} k_{\theta=1}(z_n)\,d_{z_n} .$$

We will take $1-\int_0^{k_1} k_{\theta=1}(z_n)\,dz_n = \alpha/2$ and $1-\int_0^{k_2} k_{\theta=1}(z_n)\,dz_n = 1-\alpha/2$. To compare the two tests ϕ_{LRT} and ϕ_2, we take $\sigma = 0.05$ and $n = 3,4$ and 5. Table (2) shows the power for both tests for n = 3,4 and 5 and a = 0.05.

Considering Table (2) we conclude that the power of the tests ϕ_{LRT} and ϕ_2 increases as θ moves away from 1 in both directions and as n increases and the power of ϕ_2 is higher than ϕ_{LRT} i.e. using RSS will increase the power of the test. Also, we notice that ϕ_{LRT} appears to be unbiased test while ϕ_2 is an unbiased test.

Table (2): Values of $\beta_{LRT}(\theta)$ and $\beta_{\phi_2}(\theta)$ for different values of θ and sample sizes $n = 3,4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi LRT}(\theta)$			$\beta_{\phi_2}(\theta)$		
	$n=3$	$n=4$	$n=5$	$n=3$	$n=4$	$n=5$
0.05	0.999	0.999	0.999	0.999	0.999	0.999
0.10	0.946	0.995	0.999	0.999	0.999	0.999
0.25	0.450	0.633	0.776	0.822	0.969	0.996
0.50	0.129	0.1768	0.228	0.263	0.437	0.612
0.75	0.055	0.063	0.071	0.079	0.109	0.146
0.90	0.046	0.047	0.048	0.051	0.056	0.061
1.00	0.050	0.050	0.050	0.050	0.050	0.050
1.10	0.061	0.062	0.063	0.070	0.062	0.064
1.25	0.087	0.093	0.100	0.092	0.105	0.122
1.50	0.150	0.172	0.194	0.177	0.229	0.289
2.00	0.305	0.365	0.421	0.395	0.530	0.661
3.00	0.569	0.665	0.742	0.731	0.878	0.960
4.00	0.730	0.821	0.883	0.884	0.970	0.995
5.00	0.823	0.899	0.943	0.947	0.992	0.999
10.0	0.963	0.988	0.996	0.997	0.999	0.999

3. Uniform Distribution

Let $X_1, X_2, ..., X_n$ be a random sample from the uniform distribution **with** probability density function

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

We want to test

$$H_0 : \theta = \theta_0 \text{ vs. } H_\alpha : \theta \ne \theta_0 . \tag{11}$$

As was done in case of the exponential distribution we assume that $\theta_0 = 1$ w.l.o.g. Since the UMPT test for (11) exists there is no need to consider the LRT. The UMPT of side α is given by

$$\phi_\mu = \begin{cases} 1 & \text{if } X_{(n)} > 1 \text{ or } X_{(n)} \le n\sqrt{\alpha} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $X_{(n)}$ is the largest ordered statistic. Then the power function of this test is

$$\beta_{\phi_\mu}(\theta) = \begin{cases} 1 & \text{if } \theta \le n\sqrt{\alpha} \\ \frac{\alpha}{\theta_n} & \text{if } n\sqrt{\alpha} < \theta \le 1 \\ 1 + \frac{\alpha - 1}{\theta_n} & \text{if } \theta > 1 \end{cases}$$

To test the same hypothesis using the RSS we propose the following test

$$\phi_3 = \begin{cases} 1 & \text{if } \max\{Y_i\} < c \text{ or } \max\{Y_i\} > 1 \\ 0 & \text{otherwise} \end{cases}$$

To find the value of c we must solve the equation

$$a = P_{\theta=1}\left(\max\{Y_i\} < c\right) = \prod_{i=1}^{n} \left(P_{\theta=1}(Y_i) < c\right)$$

which can be written as

$$\alpha = \prod_{i=1}^{n} \left(\int_0^c \frac{n!}{(i-1)!(n-i)!} \left(\frac{y_i}{\theta}\right)^{i-1} \left(1-\frac{y_i}{\theta}\right)^{n-1} \frac{1}{\theta} dy_i \right)$$

Then the power of this test can be written as

$$\beta_{\phi_2}(\theta) = \begin{cases} 1 & \text{if } \theta \le c \\ \prod_{i=1}^{n} P_\theta(Y_i \le c) & \text{if } c < \theta \le 1 \\ \prod_{i=1}^{n} P_\theta(Y_i \le c) + 1 - \prod_{i=1}^{n} P_\theta(Y_i \le 1) & \text{if } \theta > 1 \end{cases}$$

To compare the two tests ϕ_μ and ϕ_3 we take $\alpha = 0.05$ and $n = 3, 4$ and 5. Table (3) shows the results for $n = 3, 4$ and 5 and $\alpha = 0.05$ with different values of θ.

Considering Table (3) we see that the power of the tests ϕ_μ and ϕ_3 increases as θ moves away from 1 in both directions and as n increases and the power of ϕ_3 is larger than ϕ_μ, i.e. using RSS will increase the power of the test.

Table (3): Values of $\beta_{\phi_\mu}(\theta)$ and $\beta_{\phi_3}(\theta)$ for different values of θ and sample sizes $n = 3, 4$ and 5 and $\alpha = 0.05$

θ	$\beta_{\phi\mu}(\theta)$			$\beta_{\phi_2}(\theta)$		
	$n = 3$	$n = 4$	$n = 5$	$n = 3$	$n = 4$	$n = 5$
0.25	0.999	0.999	0.999	0.999	0.999	0.999
0.50	0.400	0.800	0.999	0.946	0.999	0.999
0.60	0.232	0.386	0.643	0.497	0.999	0.999
0.75	0.119	0.158	0.211	0.194	0.330	0.528
0.90	0.069	0.076	0.085	0.084	0.104	0.129
1.00	0.050	0.050	0.053	0.050	0.053	0.050
1.10	0.286	0.351	0.410	0.298	0.374	0.445
1.25	0.514	0.611	0.689	0.561	0.683	0.779
1.50	0.719	0.812	0.875	0.795	0.899	0.955
2.00	0.881	0.941	0.970	0.947	0.988	0.998
3.00	0.965	0.988	0.996	0.993	0.999	0.999
4.00	0.985	0.997	0.999	0.999	0.999	0.999
5.00	0.992	0.999	0.999	0.999	0.999	0.999
10.0	0.999	0.999	0.999	0.999	0.999	0.999
10.0	0.963	0.988	0.996	0.997	0.999	0.999

References

Cobby, J.M., Ridout, M.S., Bassett, P.J. and Large, R.V. (1985). An investigation to the use of ranked set sampling on grass and grass-cover swards. *Grass and Forage Science*, 40, 257-263.

Johnson, G.D., Patil, G.P. and Sinha, A.K. (1993). Ranked set sampling for vegetation research. *Abstracta Botanica*, 17, 87-102.

McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Austr. J. Agri. Res.*, 3, 385-390.

Muttlak, H.A. and McDonald, L.L. (1990). Ranked set sampling with respect to concomitant variables and with size based probability of selection. *Commun. Statist. Theory Math.*, 19(1), 205-219.

Muttlak, H.A. and McDonald, L.L. (1990). Ranked set sampling with size based probability of selection. *Biometrics*, 46, 435-445.

Muttlak, H.A. and McDonald, L.L. (1992). Ranked set sampling and line intercept method: A more efficient procedure. *Biom. Journal* 46, 435-445.

Patil, G.P. and Taillie, C. (1993). Environmental sampling, observational economy and statistical inference with emphasis on ranked set sampling, encounter sampling and composite sampling. *Bulletin ISI Proceedings of 49 Session*, Firenze, 295-312.

Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics* 36, 35-42.

—. (1986). Ranked set sampling. *Encyclopedia of Statistical Sciences* 1986 ed.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. of Statist. Math.*, 20, 1-31.