APPLICATION OF RANKED SET SAMPLING IN ESTIMATION OF AVERAGE REID VAPOR PRESSURE OF GASOLINE

Kirtee K. Kamalja[1] and Rohan D. Koshti[2*]
Department of Statistics, School of Mathematical Sciences,
Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India.
*Corresponding author: e-mail: rohankoshti5@yahoo.co.in

Abstract

We aim to assemble the real-life examples and simulations studies spread over the literature where various ranked set sampling (RSS) methods are applied in bivariate setup. Further, we consider a bivariate Reid Vapor Pressure (RVP) data with the aim to estimate mean RVP of gasoline additive. We consider Morgenstern type bivariate logistic (MTBL) distribution to be a suitable fit to the bivariate data, review the RSS estimators under various RSS schemes, generate the samples under these schemes and estimate the mean RVP along with their variances. Our study confirms the theoretical results for MTBL distribution that extreme ranked set sampling (ERSS) scheme is the best to estimate both the mean and variance parameters.

Key words: Ranked set sampling, Concomitants of order statistics, Best linear unbiased estimator, Morgenstern type bivariate logistic distribution.

1. INTRODUCTION

There are number of practical situations where sampling unit is associated with several variables and the main variable of interest is difficult or expensive or time consuming to measure while some other correlated variable is easy or economically cheap to measure. For dealing with such situations, McIntyre (1952) introduced RSS scheme as a cost-effective alternative to simple random sampling (SRS) and applied to estimate the mean pasture yield. In this sampling scheme the sampling units are ranked by judgment method, without making the actual measurements for each of the selected n samples of n units. Now from each of the i^{th} ranked set, i^{th} unit is selected, and the measurement is made forthis i^{th} judgement ranked unit with respect to characteristic of interest, $i = 1,2, \dots, n$. This results in a ranked set sample of size n by McIntyre's method. Such a selected sample of n units is called one-cycle RSS of size n. If this complete procedure is repeated m times (m cycles), the RSS of size nm is generated and such a sample is called m-cycle RSS of size nm.

Judgement method of ranking is not suitable when there is ambiguity in discriminating the rank of one unit with another. Further the impact of ranking errors is also seen on the precision of estimator based on ranked set sample. The imperfect ranking of the units in RSS leads to larger mean square errors of RSS estimators. Stokes (1977) suggested using an auxiliary variable, correlated with the study variable, to rank the sampling units. Thus, Stokes considered a situation where the variable of interest, say Y, is difficult or expensive or time consuming to measure, but an auxiliary variable X correlated with Y is easily measurable and can be ordered exactly. Now each of the n sets of n samples are ranked with respect to the auxiliary variable and from each of the i^{th} ordered set, i^{th} ranked unit is selected for measurement with respect to the variable of interest Y for $i = 1,2, \dots, n$. This gives ranked set sample based on Stokes's method. In the subsequent period, variations in choosing ranked unit for Y-measurement led to different versions of RSS schemes such as Median RSS (MRSS), Lower and Upper RSS (LRSS and URSS), ERSS, Moving Extreme RSS (MERSS), maximum ranked set sampling with unequal sample sizes (MRSSU), etc. The aim of proposing the variety of RSS schemes was to obtain an improved estimator of parameter of interest than the usual SRS method keeping the constraint of cost of sampling in mind. For more detail see Samawi et al. (1996), Al-Omari and Bouza (2014), Biradar and Santosha (2014), Koshti (2021), Koshti and Kamalja (2017, 2021a, 2023), Kamalja and Koshti (2019) etc.

One can observe that McIntyre's RSS scheme requires univariate distributional setup for parameter estimation while due to role of auxiliary variable X correlated with variable of interest Y, Stokes's scheme requires bivariate distributional setup. Further in McIntyre's method, distribution of order statistics is required while in Stokes's method theory of concomitants of order statistics (COS) is needed. While implementing the RSS schemes to the given situation, the associated univariate or bivariate distribution, the family to which the distribution belongs, parameters of the distribution, order statistics and concomitants of order statistics etc. increase the scope of studies of RSS estimation. A vast literature is developed in connection with the univariate and bivariate setup giving more and more depth to this area.

This paper aims to overview a variety of real-life situations and datasets where RSS schemes are used for parameter estimation, specifically in bivariate setup. In general RSS applications in bivariate setup consider a situation where variable under study is difficult or costly or time consuming to measure but correlated with auxiliary variable. We summarize the variety of bivariate data sets dealt by the researchers along with the nature of auxiliary and study variable, the bivariate distribution which is assumed for it and RSS schemes used for estimation purpose. This brings the applications of RSS schemes in bivariate setup in one-sight summary form. Apart from this we also summarize about the simulation studies performed in the same lines.

Further we work on an interesting application of RSS in estimation for one real-life bivariate data. The main idea is to fit a suitable distribution to significantly correlated auxiliary and study variable marginally and then assume the suitable bivariate version of the marginals to (X, Y). Instead of just assuming the bivariate distribution to the data, this seems to be more appropriate to proceed with the application of RSS schemes for parameter estimation. The paper is organized as follows.

In section 2 we review the variety of the real-life bivariate data sets where RSS schemes are used for estimation and summarize in compact form. Section 3 describes the bivariate RVP data and searches for a suitable marginal distribution to each variate of the data. In section 4 we briefly review a literature on COS and estimators under RSS schemes for the bivariate distribution which fits to the RVP data. Finally, we present the numerical estimation of parameters under RSS schemes for RVP data in section 5. The concluding remarks are given in section 6.

2. APPLICATIONS OF RSS TO REAL-LIFE BIVARIATE DATA SETS

It is seen that among the variety of RSS schemes, a specific scheme performs better than the other ones for the given bivariate distribution. For different bivariate distributions, different RSS estimators are recommended by researchers. To achieve this, usually the researcher considers the efficiency of estimator under one RSS scheme over the other and proves that it crosses unit value unconditionally or under some condition. Numerical evaluations and trends in efficiencies across the parameter values and sample sizes are used when efficiency is complex function of parameters. As the simulated data best represents the parent distribution, simulation studies are also commonly used to establish the superiority of one scheme over the other. In this case the data is simulated from the bivariate distribution under consideration and parameters are estimated under the respective schemes along with their variances and efficiencies. Some recent statistical software facilitates the simulation for specific bivariate distributions. But simulation from bivariate distributions belonging to recently introduced families need special simulation algorithms. Table 1 summarizes some of the simulation studies done for recommending specific RSS scheme for different bivariate distributions like bivariate normal, bivariate Lomax, Morgenstern type bivariate Rayleigh (MTBR), Morgenstern/Cambanis type bivariate Uniform (MTBU/CTBU), Morgenstern/Cambanis type bivariate exponential (MTBE/CTBE) distributions.

Table 1 A summary of simulation studies for RSS estimation in bivariate setup

Reference	**Distribution under consideration**	**RSS schemes Used**
Philip et al. (2002)	Bivariate Normal distribution	RSS
Al-Saleh and Al-Ananbeh (2007)		MERSS
AlKadiri and Migdadi (2019)	MTBU, MTBE	Bivariate RSS
Chacko (2017)	MTBE	RSS, URSS
Koshti and Kamalja (2021 a)	CTBU	RSS, ERSS, LRSS, URSS
Koshti and Kamalja (2021 b)	Bivariate Lomax distribution	RSS, LRSS, MRSS
Basikhasteh et al. (2021)	MTB Rayleigh distribution	RSS, URSS, MRSSU
Kamalja and Koshti (2022)	CTBE	RSS, LRSS, ERSS

The proposed RSS scheme can be justified as the best when variance of estimator based on this scheme is least among all other RSS schemes. It also becomes necessary for the researcher to present at least one real-life situation where the developed results are useful. Hence usually the research studies present the applicability of developed results through real-life examples. The variety of RSS schemes have been used in many real-life situations by researchers. Here, we briefly describe about some of the real situations and bivariate data sets found in literature for estimation under RSS schemes.

Philip et al. (2002) estimated mean RVP of the new reformulated gasoline in United States (US) based on RSS scheme with set size three and number of cycles five using two variables as: field measurement (X) and lab measurement (Y) of RVP of the gasoline. Before this, Nussbaum and Sinha (1997) also used RSS scheme for estimation of mean RVP of the gasoline for the same situation without using any specific bivariate distribution. They considered a ranked set sample of only 12 gasoline samples to send for full laboratory RVP measurement (with set size three and four cycles). That one sample from each of the three sets of the three field measurements is sent for laboratory measurement and this is repeated for four times.

Al-Saleh and Al-Hadrami (2003) and Al-Saleh and Al-Ananbeh (2007) used the data of heights and diameters of 1103 trees (obtained from Prodan (1968)) and remove smallest 20 observations to achieve normality. They proved that MERSS is more efficient than usual SRS in estimating the population mean, and the MERSS sample carries more information about the parameter than a SRS of equivalent size.

Chacko and Thomas (2007) estimated the actual average parental income of students from various departments of University of Kerala based on RSS and LRSS scheme under the assumption that (X, Y) follows bivariate Pareto distribution. Here X is parental annual income reported on admission record whereas Y is actual parental annual income which is obtained by asking questions to students on the various components of parental income. This situation is also studied by Thomas and Philip (2018). Koshti and Kamalja (2021b) estimated average annual parental income assuming (X, Y) follow bivariate Lomax distribution.

Chacko and Thomas (2009) used the bivariate data relating to Confir (*Pinus Palustrine*) trees where X is diameter (in cm) of the Confir tree at breast height and Y is height (in ft) of the tree. They assumed that (X, Y) follows Morgenstern type bivariate Logistic (MTBL) distribution to estimate the parameters associated with Y based on RSS scheme.

Tahmasebi and Jafari (2012) considered a bivariate data on 256 purslane plants (portulaca oleracea). The mean shoot diameter (Y) is estimated using RSS, ERSS and MERSS schemes using shoot height as auxiliary variable (X) and assuming (X, Y) to have MTBU distribution. Koshti and Kamalja (2021 a) assumed CTBU distribution to the same bivariate data and estimated mean shoot diameter using RSS, ERSS, LRSS and URSS schemes.

A bivariate data from marine biological research in the Persian Gulf relating to hawksbill turtle (eretmochelys imbricata) eggs is used by Tahmasebi and Jafari (2015). In this application X is weight (in gm) of eggs and Y is diameter (in mm) of the hawksbill turtle eggs. They assumed that (X, Y) follows MTB Gamma distribution to estimate parameter associated with Y using RSS, ERSS and MERSS schemes.

The study in reappraisal of caloric requirements in healthy women by Owen et al. (1986) shown that the body weight of women (X) was highly related to the Resting Metabolic Rate (RMR) (Y) of the women. Here X can be measured very easily whereas Y was difficult to measure. This situation is considered by Tahmasebi et al. (2017) to estimate parameter associated with Y-variate using RSS schemes under the assumption that (X, Y) follows MTB Rayleigh distribution. For this data Basikhasteh et al. (2021) obtained Bayes estimators based on bivariate RSS, ERSS and maximum ranked set sampling with unequal sample size.

Lange et al. (1993) studied the influence of water chemistry on mercury concentration in largemouth bass from 53 different Florida lakes. The data consist of amount of alkalinity (mg/l), calcium (mg/l), chlorophyll (mg/l) etc. in each of the water samples. The sample of fishes was taken from each lake to measure the minimum mercury concentration $(\mu g/g)$. Lange et al. (1993) observed that the bioaccumulation of mercury in the largemouth bass was strongly influenced by the chemical characteristics of the lakes. Accordingly, the amount of alkalinity in water sample (X) and the minimum mercury concentration in the sampled fish (Y) forms the bivariate data. This data is also used by Mohsin et al. (2014) and Chacko (2017). Mohsin et al. (2014) fitted a bivariate exponential distribution to the data while Chacko (2017) assumed (X, Y) to have MTBE distribution and used RSS and ERSS schemes for estimation of minimum mercury concentration. Kamalja and Koshti (2022) estimated the mean minimum mercury concentration in largemouth bass from different Florida lakes using RSS, LRSS and ERSS schemes assuming $(X, Y) \sim CTBE$ distribution.

Koshti and Kamalja (2021 b) considered the gross income (in US$) in 2005 (X) and 2010 (Y) of 77 Professors, Associate Professors and Assistant Professors who worked at University of California (UC) for more than 5 years in the Department of Statistics. Shih et al. (2019) fitted the bivariate Pareto distribution to this data using the Frank copula and Sankaran and Nair bivariate Pareto (SNBP) distributions by using maximum likelihood estimation. Koshti and Kamalja (2021 b) estimated mean gross income for the professors at UC system under RSS, LRSS and MRSS schemes assuming bivariate Lomax distribution to (X, Y).

The above review shows that the researchers considered both parametric and nonparametric methods for the estimation under RSS schemes for bivariate data. The parameter estimation associated with bivariate distributions under various RSS schemes has further been used to estimate other population quantities like mean, variance etc. In the following Table 2 we present a summary of bivariate real-life data sets discussed above along with the assumed bivariate distribution and RSS schemes used for parameter estimation.

Table 2 A summary of real-life bivariate data sets where RSS estimation is used

Reference	**Details of data set**	**Details about of variables**		**Distribution under consideration**	**Variant of RSS schemes used**
		Study variable	**Auxiliary variable**		
Philip et al. (2002)	Environmental Protection Agency (EPA) data (Size:15)	RVP measurement at laboratory	RVP measurement at field	Bivariate Normal distribution	RSS
Chen et al. (2003)	RVP data (Size: 90)			-	RSS, URSS
Al-Saleh and Al-Hadrami (2003)	Tree data by Prodan (1968) (Size: 1083)	Diameter of the tree	Height of tree	Bivariate Normal distribution	MLE based on MERSS
Al-Saleh and Al-Ananbeh(2007)					
Chacko and Thomas (2007), Thomas and Philip (2018)	Parental income of PG students (Size: 64)	Actual parental annual income	Reported parental annual income in records	Bivariate Pareto distribution	RSS, LRSS
Koshti and Kamalja (2021 b)				Bivariate Lomax distribution	

Reference	Details of data set	Details about of variables		Distribution under consideration	Variant of RSS schemes used
		Study variable	**Auxiliary variable**		
Chacko and Thomas (2009), Singh and Mehta (2013)	Confir (*Pinus Palustrine*) tree (Size: 396)	Height of the tree	Diameter of the tree at breast height	MTB Logistic	RSS
Tahmasebi and Jafari (2012)	Biological study on purslane plants (Size: 256)	Shoot diameter	Shoot height	MTB Uniform	RSS, ERSS, MERSS
Koshti and Kamalja (2021 a)				CTB Uniform	RSS, ERSS, LRSS, URSS
Tahmasebi and Jafari (2015)	Marine biological science in Persian Gulf (Size: 300)	Diameter of the hawksbill turtle eggs	Weight of the eggs	MTB Gamma	RSS, ERSS, MERSS
Tahmasebi et al. (2017)	Health data on Women (Size: 44)	Resting Metabolic Rate of women	Body weight of women	MTB Rayleigh	RSS, ERSS, MERSS
Basikhasteh et al. (2021)					RSS, URSS, MRSSU (Bayes estimators)
Chacko (2017)	Mercury concentration in largemouth bass from Florida (Size: 52)	Minimum mercury concentration in the sampled fish	Amount of alkalinity in water sample	MTB Exponential	RSS, ERSS, (Bayes estimators)
Kamalja and Koshti (2022)				CTB Exponential	RSS, LRSS, ERSS
Koshti and Kamalja (2021 b)	Gross income data at UC (Size: 77)	Gross income in 2010	Gross income in 2005	Bivariate Lomax distribution	RSS, LRSS, MRSS

In most of the above situations the researchers assumed a suitable bivariate distribution to the real-life bivariate data and estimated the parameter of interest using suitable RSS scheme. These studies are more focused on estimation under RSS schemes. In a very few cases statistical goodness of fit tests are used to validate the hypothetical bivariate distribution. The reason for this is, checking the validity of specific distributional assumption for univariate data is quite easy with the help of existing statistical software but to validate the fit of bivariate distribution, limited statistical software with only a few multivariate distributions are available.

To deal with bivariate data, it seems more realistic to guess a suitable marginal distribution to each variable and validate it statistically. Once the confirmation of validity of specified distribution for each of the univariate data is done, the significance of correlation between the two variables will set up the assumption of the corresponding bivariate distribution. We motivated to follow this procedure to deal with a real-life bivariate data for RSS estimation instead of mere assumption of the bivariate distribution.

We consider a real-life situation described by Philip et al. (2002) where RSS scheme is suitable to apply in bivariate set up. We fit some possible suitable distributions to both the study and auxiliary variables marginally and choose the one with the best goodness of fit measure. The correlation between the two variates builds the assumption about the corresponding bivariate distribution to the data and the appropriate RSS scheme for parameter estimation can then be used.

3. DEALING WITH RVP DATA

Philip et al. (2002) considered one motivational application of RSS from the Environmental Protection Agency (EPA) of the US to evaluate gasoline quality which is measured by the Reid Vapor Pressure. Unburned hydrocarbons emitted from automobile tailpipes produce ground level ozone and smog. But the recent advancements in automobile technology, many of the hydrocarbons evaporate off the manifold. One way to reduce this evaporated emission is to control the use of gasoline and vehicle. This can be achieved by reducing volatility of the gasoline which is measured by the RVP value. In US it is recommended to use reformulated gas which has certain limiting RVP value.

An EPA inspector occasionally visits gas pump in a city, takes sample of gasoline and measures RVP at the field which produces cheap and quick measurement. Once in a while, the inspector after measuring RVP at the field will ship a gasoline sample to the laboratory for a measurement likely with higher precision at a higher cost. Thus, the pair of measurement is collected at field and laboratory. Here laboratory measurements (Y) of RVP are much more expensive than measurement at field (X) because of special packing to be used to ship a gasoline sample from a field to a laboratory. Cheap field measurements of RVP are correlated with laboratory measurements. A cost-effective sampling method is extremely desirable in this situation as discussed by Nussbaum and Sinha (1997). Philip et al. (2002) efficiently estimated the mean RVP in gasoline consumed by the public under the assumption that (X, Y) follow bivariate normal distribution. For some other references related with this study see, Chen et al. (2003) and Wolfe (2012).

We consider a data used by Chen et al. (2003) for which 90 field (X) and laboratory (Y) measurements are given. Chen et al. (2003) generated a sample of size 15 using both SRS and RSS method for estimation of mean laboratory measurement Y without any assumption about the distribution of (X, Y). Chen et al. (2003) used the 90 X-values from Table 3 of Nussbaum and Sinha (1997) and generated 90 corresponding Y-values using fitted simple linear regression model. This data and its descriptive statistics are shown in Table 3 and Table 4 respectively.

Table 3 The $(Field, lab)$ measurements values

No.	X	Y	No.	X	Y	No.	X	Y
1	7.27	7.42043	31	7.47	7.40194	61	7.35	7.30603
2	7.57	7.55349	32	7.54	7.47252	62	7.41	7.54441
3	7.47	7.49534	33	7.54	7.49941	63	7.28	7.17364
4	7.27	7.15787	34	7.45	7.43206	64	7.41	7.37344
5	7.51	7.70336	35	6.42	6.48164	65	7.37	7.37414
6	8.03	7.97076	36	8.21	8.09954	66	7.63	7.54699
7	7.37	7.40452	37	8.69	8.80488	67	7.37	7.40092
8	7.16	7.13687	38	8.64	8.61522	68	7.45	7.29943
9	8.32	8.26775	39	7.86	7.95413	69	7.47	7.53020
10	8.30	8.30437	40	8.22	8.20800	70	7.37	7.43612
11	7.51	7.43280	41	7.35	7.21393	71	7.32	7.33868
12	7.01	6.96980	42	7.37	7.14588	72	7.30	7.31769
13	7.52	7.57230	43	7.41	7.36116	73	7.22	7.10908
14	6.53	6.43238	44	7.45	7.49847	74	7.47	7.41043
15	7.01	6.92487	45	7.44	7.46515	75	7.54	7.58248
16	7.54	7.45800	46	8.34	8.29940	76	7.31	7.36234
17	7.31	7.32792	47	8.56	8.64405	77	7.25	7.36149
18	7.59	7.55373	48	7.32	7.22311	78	7.37	7.26742
19	7.37	7.23511	49	7.35	7.45366	79	7.32	7.23074
20	7.47	7.49863	50	7.50	7.35767	80	7.28	7.33903
21	7.56	7.66931	51	7.47	7.49101	81	7.38	7.49330

No.	X	Y	No.	X	Y	No.	X	Y
22	7.34	7.456940	52	7.37	7.35279	82	7.22	7.11519
23	7.56	7.51281	53	7.43	7.30706	83	7.76	7.69649
24	7.45	7.58576	54	7.41	7.41159	84	7.45	7.42787
25	7.60	7.49504	55	7.37	7.2815	85	7.51	7.56681
26	7.63	7.53542	56	7.31	7.28819	86	7.47	7.54080
27	7.16	7.23144	57	7.59	7.50309	87	7.38	7.33081
28	7.54	7.56595	58	7.47	7.52883	88	7.79	7.77385
29	7.51	7.49295	59	7.43	7.42279	89	7.38	7.46263
30	7.52	7.62099	60	7.40	7.56357	90	7.14	7.03363

~ Source: Chen et al. (2003)

Table 4 Descriptive statistics of RVP of gasoline data

Variable	Minimum	Q_1	Median	Mean	Q_3	Maximum	SD	$\hat{\rho}_{XY}$
X	6.420	7.35	7.445	7.493	7.540	8.690	0.3594	0.9760
Y	6.432	7.32	7.445	7.484	7.552	8.805	0.3736	

We fit lognormal, gamma, normal, Weibull, exponential and logistic distribution to both the variables X and Y. The goodness of fit statistics for respective fits which include, Kolmogorov-Smirnov (K-S) statistic, Cramer-von Mises (C-M) statistic, Anderson-Darling (A-D) statistic along with the analytical measures Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) are reported in Table 5. The lower values of all these statistics and analytical measures are desirable for confirming one distribution among all the candidate distributions.

Table 5 Goodness of fit statistics and analytical measures of the competing distributions for the RVP data

Variable	Distribution	Goodness of fit with			Goodness of fit criterion	
		K-S statistic	C-M statistic	A-D statistic	AIC	BIC
X	Logistic	0.1449	0.5721	4.1478	52.0516	57.0512
	Log Normal	0.2160	1.1734	6.3291	70.3452	75.3448
	Gamma	0.2195	1.2059	6.4883	71.5140	76.5136
	Normal	0.2265	1.2757	6.8288	74.2233	79.2229
	Weibull	0.2758	2.0393	10.3618	111.9971	116.9967
	Exponential	0.5854	8.3533	38.191	544.5091	547.0089
Y	Logistic	0.1460	0.3716	2.7188	62.1091	67.1087
	Log Normal	0.2141	0.8254	4.5066	77.3904	82.3900
	Gamma	0.2177	0.8544	4.6512	78.5229	83.5225
	Normal	0.2249	0.9174	4.9658	81.1799	86.1795
	Weibull	0.2710	1.6851	8.7192	118.193	123.193
	Exponential	0.5814	8.2976	37.9537	544.286	546.786

From Table 5, Logistic distribution seems to be the best one among all the other distributions. The MLEs of parameters of Logistic distribution are reported in Table 6 along with their standard errors.

Table 6 Logistic distribution fit summary for X and Y

Variable	$\hat{\mu}$	$\hat{\sigma}$	$SE(\hat{\mu})$	$SE(\hat{\sigma})$
X	7.4518	0.1619	0.0279	0.0154
Y	7.4485	0.1742	0.0303	0.0163

For further confirmation, we plot the empirical CDF of Logistic distribution using 'fitdist' function in 'fitdistrplus' (Delignette-Muller and Dutang (2014)) package in R for both variables X and Y. These are presented in Figure 1.

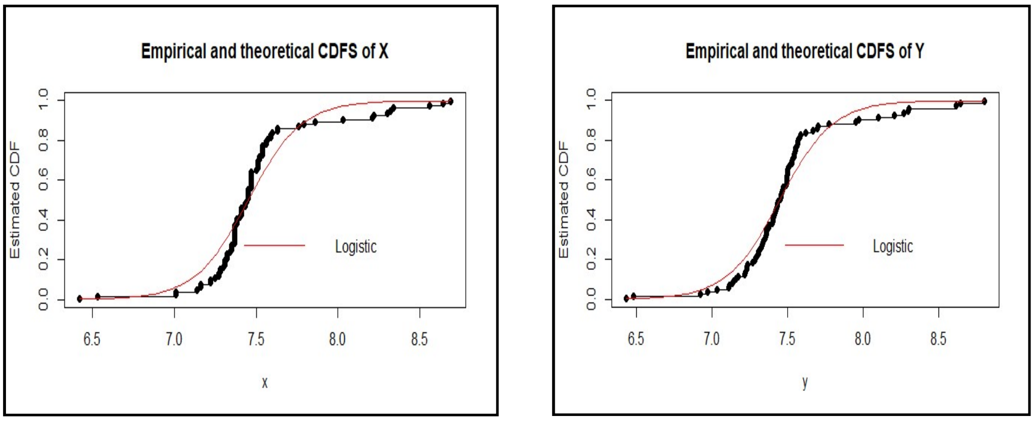

Figure 1: Empirical CDF of Logistic distribution for X and Y

All the above results support the Logistic fit to be the most suitable among all other considered competitor distributions for both X and Y. Thus $X \sim Logistic(7.4518, 0.1619)$ and $Y \sim Logistic(7.4485, 0.1742)$. The correlated random variables motivate us to assume MTBL distribution to RVP data. We briefly review the estimators of parameters for MTBL distribution under RSS schemes proposed by Chacko and Thomas (2009) and Lesitha et al. (2010) in the next section.

4. REVIEW OF RSS ESTIMATION FOR MTBL DISTRIBUTION

We briefly review results on COS for MTBL distribution. Let (X, Y) have MTBL distribution with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha)$, denoted as $MTBL(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha)$. The pdf of $MTBL(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha)$ distribution given by Kotz et al. (2000) is,

$$f(x,y) = \frac{\exp\left(-\frac{x-\mu_1}{\sigma_1}\right)}{\sigma_1\left(1+\exp\left(-\frac{x-\mu_1}{\sigma_1}\right)\right)^2} \frac{\exp\left(-\frac{y-\mu_2}{\sigma_2}\right)}{\sigma_2\left(1+\exp\left(-\frac{y-\mu_2}{\sigma_2}\right)\right)^2} \left(1 + \alpha\left(\frac{1-\exp\left(-\frac{x-\mu_1}{\sigma_1}\right)}{1+e\ \left(-\frac{x-\mu_1}{\sigma_1}\right)}\right)\left(\frac{1-e\ \left(-\frac{y-\mu_2}{\sigma_2}\right)}{1+\exp\left(-\frac{y-\mu_2}{\sigma_2}\right)}\right)\right);$$

$$x, y, \mu_1, \mu_2 \in \mathbb{R};\ \sigma_1, \sigma_2 > 0; \alpha \in [-1,1].$$

The marginal distribution of Y is Logistic with parameters μ_2, σ_2 with pdf

$$f_2(y) = \frac{exp\left(-\frac{y-\mu_2}{\sigma_2}\right)}{\sigma_2\left(1+exp\left(-\frac{y-\mu_2}{\sigma_2}\right)\right)^2}; \qquad y, \mu_2 \in \mathbb{R}; \sigma_2 > 0.$$

where μ_2 is location parameter and σ_2 is scale parameter.

Chacko and Thomas (2009) and Lesitha et al. (2010) studied the COS for $MTBL$ distribution and proposed RSS estimators of μ_2 and σ_2 using the results of COS. Let $Y_{[r]r}$ be the concomitants of r^{th} order statistic corresponding to r^{th} order statistic $X_{(r)r}$ in the r^{th} ordered sample $X_{(1)r}, X_{(2)r}, \dots, X_{(n)r}, r = 1,2,..n$. The mean and variance of $Y_{[r]r}$ are as follows.

$$E\left(Y_{[r]r}\right) = \mu_2 + \xi_r\sigma_2, Var\left(Y_{[r]r}\right) = \delta_r\sigma_2^2$$

where $\xi_r = -\frac{\alpha(n-2r+1)}{(n+1)}$, $\delta_r = \frac{\pi^2}{3} - \alpha^2\left(\frac{n-2r+1}{n+1}\right)^2$.

- **Estimation under usual RSS scheme**

Let $Y_{[1]1}, Y_{[2]2}, \dots, Y_{[n]n}$ be RSS sample of size n from MTBL distribution. The RSS estimators of μ_2 and σ_2 and their variances based on the RSS sample are under the assumption that the association parameter α from $MTBL(\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha)$ distribution is known. To review the estimators given by Chacko and Thomas (2009) consider the following notations.

$\boldsymbol{Y_{[n]}} = [Y_{[1]1} \quad Y_{[2]2} \quad \cdots \quad Y_{[n]n}]'$ is a column vector of ranked set sample,

$\boldsymbol{\xi} = (\xi_1 \quad \xi_2 \quad \cdots \quad \xi_n)'$,

$\mathbf{1} = (1, 1, \dots, 1)'$ is a column vector of size n with all entries 1,

$\boldsymbol{G} = diag(\delta_1, \delta_2, \dots, \delta_n)$ is a $n \times n$ diagonal matrix,

$\Delta = \left(\boldsymbol{\xi}'\boldsymbol{G}^{-1}\boldsymbol{\xi}\right)\left(\mathbf{1}'\boldsymbol{G}^{-1}\mathbf{1}\right) - \left(\boldsymbol{\xi}'\boldsymbol{G}^{-1}\mathbf{1}\right)^2$,

$[.]$ is the greatest integer function and

$T_r = \left(Y_{[r]r} - Y_{[n-r+1]\overline{n-r+1}}\right), r = 1,2, \dots, n.$

The unbiased estimators $\hat{\mu}_{2,RSS}$, $\hat{\sigma}_{2,RSS}$ and BLUEs $\hat{\mu}_{2,BLUE}$ and $\hat{\sigma}_{2,BLUE}$ of μ_2 and σ_2 based on RSS sample given by Chacko and Thomas (2009) are summarized in the following.

Scheme	Estimator	Var (Estimator)
RSS	$\hat{\mu}_{2,RSS} = \frac{1}{n}\sum_{r=1}^{n} Y_{[r]r}$	$Var(\hat{\mu}_{2,RSS}) = \frac{\sigma_2^2}{n}\left(\frac{\pi^2}{3} - \frac{\alpha^2}{n}\sum_{r=1}^{n}\left(\frac{n-2r+1}{n+1}\right)^2\right)$
	$\hat{\sigma}_{2,RSS} = \frac{1}{\sum_{r=1}^{\left[\frac{n}{2}\right]}\xi_r}\sum_{r=1}^{\left[\frac{n}{2}\right]} T_r$	$Var(\hat{\sigma}_{2,RSS}) = \frac{\sigma_2^2}{2\left(\sum_{r=1}^{\left[\frac{n}{2}\right]}\xi_r\right)^2}\sum_{r=1}^{\left[\frac{n}{2}\right]}\left(\frac{\pi^2}{3} - \alpha^2\left(\frac{n-2r+1}{n+1}\right)^2\right)$
RSS BLUE	$\hat{\mu}_{2,BLUE} = \Delta^{-1}\left(\boldsymbol{\xi}'\boldsymbol{G}^{-1}(\boldsymbol{\xi}\mathbf{1}' - \mathbf{1}\boldsymbol{\xi}')G^{-1}\right)\boldsymbol{Y}_{[\boldsymbol{n}]}$	$Var(\hat{\mu}_{2,BLUE}) = \frac{\sigma_2^2(\boldsymbol{\xi}'\boldsymbol{G}^{-1}\boldsymbol{\xi})}{\Delta}$
	$\hat{\sigma}_{2,BLUE} = \Delta^{-1}\left(\mathbf{1}'\boldsymbol{G}^{-1}(\mathbf{1}\boldsymbol{\xi}' - \boldsymbol{\xi}\mathbf{1}')\boldsymbol{G}^{-1}\right)\boldsymbol{Y}_{[\boldsymbol{n}]}$	$Var(\hat{\sigma}_{2,BLUE}) = \frac{\sigma_2^2(\mathbf{1}'\boldsymbol{G}^{-1}\mathbf{1})}{\Delta}$

Chacko and Thomas (2009) obtained moment estimator of α based on sample correlation coefficient $\hat{\rho}$ as follows.

$$\hat{\alpha} = \begin{cases} -1 & if\,\hat{\rho} \leq -\frac{3}{\pi^2} \\ \frac{\hat{\rho}\pi^2}{3} & if -\frac{3}{\pi^2} < \hat{\rho} < \frac{3}{\pi^2} \\ 1 & if\,\hat{\rho} \geq \frac{3}{\pi^2} \end{cases}$$

- **Estimation under ERSS scheme**

Lesitha et al. (2010) reported that the maximum information about the parameters μ_2 and σ_2 is available in concomitants of extreme order statistics for MTBL distribution. This information is utilized to obtain estimator of μ_2 and σ_2 under ERSS scheme. In this scheme Lesitha et al. (2010) considered $\left[\frac{n}{2}\right]$ independent sets of samples each with n units and measure Y-variable associated with the lower and upper extremes with respect to X in each sample. They use such a generated sample to obtain estimators of μ_2 and σ_2.

For the sake of application of ERSS scheme, we consider the usual ERSS scheme proposed by Samawi et al. (1996) to estimate parameters. It consists of generating n random samples each of n units from the population. The even-size sample under ERSS scheme is $\left(Y_{[1]1}, Y_{[n]2}, \ldots, Y_{[1]n-1}, Y_{[n]n}\right)$ and is denoted by ERSS_1. While an odd-size ERSS sample is $\left(Y_{[1]1}, Y_{[n]2}, Y_{[1]3} \dots, Y_{[n]n-1}, Y_{\left[\frac{n+1}{2}\right]n}\right)$ and is denoted by ERSS_2. The BLUEs of μ_2 and σ_2 their variances using ERSS_1 and ERSS_2 samples are as follows.

$$\hat{\mu}_{2,ERSS} = \Delta^{-1}\left(\boldsymbol{\xi}'\boldsymbol{G}^{-1}\left(\boldsymbol{\xi}\mathbf{1}' - \mathbf{1}\boldsymbol{\xi}'\right)G^{-1}\right)\boldsymbol{Y}_{[\boldsymbol{n}]}, \qquad Var\left(\hat{\mu}_{2,ERSS}\right) = \frac{\sigma_2^2(\boldsymbol{\xi}'\boldsymbol{G}^{-1}\boldsymbol{\xi})}{\Delta},$$

$$\hat{\sigma}_{2.ERSS} = \Delta^{-1}\left(\mathbf{1}'\boldsymbol{G}^{-1}\left(\mathbf{1}\boldsymbol{\xi}' - \boldsymbol{\xi}\mathbf{1}'\right)\boldsymbol{G}^{-1}\right)\boldsymbol{Y}_{[\boldsymbol{n}]}, \qquad Var(\hat{\sigma}_{2.ERSS}) = \frac{\sigma_2^2(\mathbf{1}'\boldsymbol{G}^{-1}\mathbf{1})}{\Delta}.$$

Here respective $\boldsymbol{Y}_{[\boldsymbol{n}]}$, $\boldsymbol{\xi}$ and $\boldsymbol{G}$ are to be used as specified below for even and odd sample sizes, that is, for ERSS_1 and ERSS_2 samples.

$$\boldsymbol{Y}_{[\boldsymbol{n}]} = \begin{cases} \left[Y_{[1]1}, Y_{[n]2}, \dots, Y_{[1]n-1}, Y_{[n]n}\right]' & if\ n\ even \\ \left[Y_{[1]1} \quad Y_{[n]2} \quad \cdots \quad Y_{\left[\frac{n+1}{2}\right]n}\right]' & if\ n\ odd \end{cases}$$

$$\boldsymbol{\xi} = \begin{cases} (\xi_1 \quad \xi_n \quad \cdots \quad \xi_n)' & if\ n\ even \\ \left(\xi_1 \quad \xi_n \quad \cdots \quad \xi_{\frac{n+1}{2}}\right)' & if\ n\ odd \end{cases},$$

$$\boldsymbol{G} = \begin{cases} diag(\delta_1, \delta_n, \dots, \delta_n) & if\ n\ even \\ diag\left(\delta_1, \delta_n, \dots, \delta_{\frac{n+1}{2}}\right) & if\ n\ odd \end{cases}.$$

5. ESTIMATION OF PARAMETERS FOR RVP DATA UNDER RSS SCHEMES

In this section we estimate RVP measurement of the gasoline at laboratory using RSS and ERSS schemes. From the bivariate RVP data $(X_i, Y_i), i = 1, 2, \dots, 90$, RSS sample of size 8 is generated using 'RSSampling' (Sevinc et al. (2019)) package in R. To choose ERSS sample we use the same 8×8 array of 8 random samples each of size 8 as given by the package for RSS sample selection. The RSS and ERSS samples selected are presented in Table 7.

Table 7 The RSS and ERSS samples for laboratory measurement of RVP

Scheme	Sample values for Y- variable							
RSS	7.0336	7.1091	7.4537	7.4228	7.6210	7.6693	7.9708	7.6965
$ERSS_1$	7.0336	7.5825	7.2314	8.2678	7.3279	8.2678	6.4324	7.6965

As the estimation under RSS and ERSS schemes are under the assumption that the association parameter α is known, we need to estimate α. We estimate α using correlation between the bivariate data as given by Chen et al. (2003). The $\hat{\rho}_{XY} = 0.9760$ leads to an estimate of α as $\hat{\alpha} = 1$ as given by Chacko and Thomas (2009). We use $\alpha = 1$ in RSS and ERSS estimators and their variances as reported in Section 4. Table 8 shows the estimates of μ_2 and σ_2 under RSS and $ERSS_1$ schemes along with their variances assuming $(X, Y) \sim MTBL(\mu_1, \mu_2, \sigma_1, \sigma_2, 1)$ distribution.

Table 8 Estimates of μ_2 and σ_2 under RSS and ERSS schemes

Scheme	Estimator of μ_2	Estimate of μ_2	$\frac{V(\hat{\mu}_{2,Scheme})}{\sigma_2^2}$	Estimator of σ_2	Estimate of σ_2	$\frac{V(\hat{\sigma}_{2,Scheme})}{\sigma_2^2}$
RSS	$\hat{\mu}_{2,RSS}$	7.4971	0.3788	$\hat{\sigma}_{2,RSS}$	0.5452	1.9178
	$\hat{\mu}_{2,BLUE}$	7.4918	0.3766	$\hat{\sigma}_{2,BLUE}$	0.5195	1.3603
$ERSS_1$	$\hat{\mu}_{2,ERSS}$	7.4800	0.3356	$\hat{\sigma}_{2,ERSS}$	0.6090	0.5548

The results reported in Table 8 conclude that as expected, an estimator of μ_2 and σ_2 based on ERSS scheme has least variance as compared to RSS scheme. It is observed that estimator of population mean under ERSS scheme is close to the true value.

6. CONCLUSIONS

In this paper we summarize various real-life bivariate data sets and simulation studies where the estimation of parameters is done under different RSS schemes. We consider the situation where RVP measurement at laboratory is much more expensive than measurement at field but highly correlated with laboratory measurement. This situation encourages to use RSS scheme for sample selection as it gives more efficient estimator than the usual SRS scheme. We consider the sample observations on the two variables as given by Chen et al. (2003) and fit various distributions to each variable marginally and arrive at conclusion that logistic distribution is the most suitable among all other competitor distributions. Logistic marginals with correlation support to assume MTBL distribution to the bivariate data under study. Further we briefly review the estimation of location and scale parameter of MTBL distribution under RSS and ERSS schemes. Finally, we establish that, among the usual RSS and ERSS scheme, estimate of RVP of gasoline under ERSS scheme is better than RSS assuming the MTBL distribution to the data.

REFERENCES

[1] Al Kadiri, M. and Migdadi, M. (2019). Estimating parameters of Morgenstern type bivariate distribution using bivariate ranked set sampling. *Electronic Journal of Applied Statistical Analysis*, 12, 190-208.

[2] Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: modifications and applications. *Investigación Operacional*, 35, 215-235.

[3] Al-Saleh, M.F. and Al-Ananbeh, A.M. (2007). Estimation of the means of the bivariate normal using moving extreme ranked set sampling with concomitant variable. *Statistical papers*, 48, 179-195.

[4] Al-Saleh, M.F. and Al-Hadrami, S.A. (2003). Parametric estimation for the location parameter for symmetric distributions using moving extremes ranked set sampling with application to trees data. *Environmetrics*, 14, 651–664.

[5] Basikhasteh, M., Lak, F. and Tahmasebi, S. (2021). Bayesian Estimation of Morgenstern Type Bivariate Rayleigh Distribution Using Some Types of Ranked Set Sampling. *Revista Colombiana de Estadística*, 44, 279-296.

[6] Biradar, B.S. and Santosha, C.D. (2014). Estimation of the mean of the exponential distribution using maximum ranked set sampling with unequal samples. *Open Journal of Statistics*, 4, 641-649.

[7] Chacko, M. (2017). Bayesian estimation based on ranked set sample from Morgenstern type bivariate exponential distribution when ranking is imperfect. *Metrika*, 80, 333-349.

[8] Chacko, M. and Thomas, P.Y. (2007). Estimation of a parameter of bivariate Pareto distribution by ranked set sampling. *Journal of Applied Statistics*, 34, 703-714.

[9] Chacko, M. and Thomas, P. Y. (2009). Estimation of parameters of Morgenstern type bivariate logistic distribution by ranked set sampling. *Journal of the Indian Society of Agricultural Statistics*, 63, 77-83.

[10] Chen, Z, Bai, Z. and Sinha, B. (2003). Ranked set sampling: theory and applications. Springer Science & Business Media.

[11] Delignette-Muller, M. L., and Dutang, C. (2014). fitdistrplus: an R package for distribution fitting methods.

[12] Kamalja, K.K. and Koshti, R. D. (2019). Estimation of scale parameter of Morgenstern type bivariate generalized uniform distribution by ranked set sampling. *Journal of Data Science*, 17, 513-534.

[13] Kamalja, K. K. and Koshti, R. D. (2022). Application of Ranked Set Sampling in Parameter Estimation of Cambanis-Type Bivariate Exponential Distribution. *Statistica*, 82, 145-175.

[14] Koshti, R. D. (2021). A Study on Concomitants of Order Statistics and its applications in Ranked Set Sampling, Ph.D. Thesis, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. (http://hdl.handle.net/10603/396780)

[15] Koshti, R. D. and Kamalja, K. K. (2017). Estimation of scale parameter of a bivariate Lomax distribution by ranked set sampling. *Model Assisted Statistics and Applications*, 12, 107-113.

[16] Koshti, R. D. and Kamalja, K. K. (2021a). Parameter estimation of Cambanis-type bivariate uniform distribution with ranked set sampling. *Journal of Applied Statistics*, 48, 61-83.

[17] Koshti, R. D. and Kamalja, K. K. (2021b). Efficient estimation of a scale parameter of bivariate Lomax distribution by ranked set sampling. *Calcutta Statistical Association Bulletin*, 73, 24-44.

[18] Koshti, R. D. and Kamalja, K. K. (2023). A Review on Concomitants of Order Statistics and its Application in Parameter Estimation under Ranked Set Sampling. *Journal of the Korean Statistical Society* (Accepted for publication)

[19] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). Continuous Multivariate Distributions. New York; Wiley.

[20] Lange, T. R., Royals, H. E. and Connor, L. L. (1993). Influence of water chemistry on mercury concentration in largemouth bass from Florida lakes. *Transactions of the American Fisheries Society*, 122, 74-84.

[21] Lesitha, G., Thomas, P. Y. and Chacko, M. (2010). Applications of ranked set sampling in estimating parameters of Morgenstern type bivariate logistic distribution. *Calcutta Statistical Association Bulletin*, 62, 71-90.

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

[23] Mohsin, M., Kazianka, H., Pilz, J. and Gebhardt, A. (2014). A new bivariate exponential distribution for modeling moderately negative dependence. *Statistical Methods & Applications*, 23, 123-148.

[24] Nussbaum, B. D. and Sinha, B. K. (1997). Cost effective gasoline sampling using ranked set sampling. In Proceedings of the Section on Statistics and the Environment, *American Statistical Association*, 83-87.

[25] Owen, O. E., Kavle, E., Owen, R. S., Polansky, M., Caprio, S., Mozzoli, M. A., Kendrick, Z. V., Bushman, M., and Boden, G. (1986). A reappraisal of caloric requirements in healthy women. *The American Journal of Clinical Nutrition*, 44, 1-19

[26] Philip L.H., Sun Y., Sinha B.K. (2002) Estimation of the common mean of a bivariate normal population. *Annals of the Institute of Statistical Mathematics,* 54, 861-878.

[27] Prodan M. (1968). Forest Biometrics–Pergamon Press. London, U.K., 447.

[28] Samawi, H. M., Ahmed, M. S. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, 38, 577-586.

[29] Sevinc, B., Cetintav, B., Esemen, M. and Gürler, S. (2019). RSSampling: A pioneering package for ranked set sampling. *R Journal*, 11.

[30] Shih, J. H., Lee, W., Sun, L. H. and Emura, T. (2019). Fitting competing risks data to bivariate Pareto models. *Communications in Statistics-Theory and Methods*, 48, 1193-1220.

[31] Singh, H. P. and Mehta, V. (2013). An improved estimation of parameters of Morgenstern type bivariate logistic distribution using ranked set sampling. *Statistica*, 73, 437-461.

[32] Stokes, S.L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics-Theory and Methods*, 6, 1207-1211.

[33] Tahmasebi, S. and Jafari, A. A. (2012). Estimation of a scale parameter of Morgenstern type bivariate uniform distribution by ranked set sampling. *Journal of Data Science*, 10, 129-141.

[34] Tahmasebi, S. and Jafari, A. A. (2015). A review on unbiased estimators of a parameter from morgenstern type bivariate gamma distribution using ranked set sampling. *Azerbaijan Journal of Mathematics*, 5, 3-12.

[35] Tahmasebi, S., Eskandarzadeh, M. and Almaspoor, Z. (2017). Inferences on a Scale Parameter of Bivariate Rayleigh Distribution by Ranked Set Sampling. *Pakistan Journal of Statistics and Operation Research*, 13, 1-16.

[36] Thomas, P.Y. and Philip, A. (2018) Induced Ranked Set Sampling when Units are Inducted from Several Populations. *Statistica*, 78, 57-79.

[37] Wolfe, D. A. (2012). Ranked set sampling: its relevance and impact on statistical inference. ISRN Probability and Statistics.