IAPQR Transactions
Journal of the Indian Association for Productivity, Quality and Reliability

CONTENTS

Volume 47, Numbers 1 & 2, 2022-2023

PRINTS PUBLICATIONS PVT LTD
Viraj Tower-2, 4259/3, Ansari Road, Darya Ganj,
New Delhi–110002, INDIA
Tel : +91-11-45355555 | Fax : +91-11-23275542
contact@printspublications.com
www.printspublications.com

IAPQR Transactions VOLUME 47 Numbers 1 & 2 2022-2023

ISSN 0970-0102

IAPQR Transactions

Journal of The Indian Association for Productivity, Quality and Reliability

VOLUME 47 Numbers 1 & 2 2022-2023

Prints Publications Pvt Ltd
New Delhi
www.printspublications.com

Editorial Office: Indian Association for Productivity, Quality & Reliability
E-mail : iapqrtransactions@gmail.com
Website : www.iapqr.org

Annual Subscription Price

Rs. 1000 (National)	U.S. $ 100.00 (International)
(Online version free with print subscription)	

All remittance must be paid in favour of **Prints Publications Pvt Ltd**, *payable at* **New Delhi.**

Periodicity : This Journal is published Bi-annually.

RULES

(1) Two issues of the Journal will be published each year.

(2) Articles, notes and communications meant for publication in this Journal should have a bearing on the theory and/or application of productivity enhancing and regulatory techniques.

(3) Contributions to IAPQR Transactions are expected to contain non-trivial applications of quantitative methods which enhance the substantive content of the contribution.

(4) A soft copy of the complete manuscript should be submitted to the Editors by e-mail at iapqrtransactions@gmail.com, with a copy to kuntalbakuli17@gmail.com, composed using Microsoft Word (equation version 3.0).

Articles published in this journal are abstracted/reviewed in renowned scientific periodicals like Zentralblatt fur Mathematik, Mathematical Reviews, Statistical Theory and Methods Abstracts, Quality Control and Applied Statistics and some other such publications.

We are pleased to inform you that Prints Publications Pvt Ltd has become member of the DOI Crossref, USA. The DOI prefix allotted for Prints Publications is 10.32381. Also DOI allotted for this Journal is:

https://doi.org/10.32381/IAPQRT.

CONTENTS

OBITUARY

Prof. Hemshankar Ray

A Polymath

(19th June, 1940 - 28th April, 2023)

A great material scientist, teacher, a polymath - above all a great mind, Hemshankar Ray breathed his last in the morning hours on 28th April, 2023 after a brief illness at Kolkata. As vice-president of IAPQR, he enthusiastically initiated several activities and would monitor even the smallest of details to ensure flawless execution. It is unbelievable that such a lively person can leave us all of a sudden.

Born on June 19, 1940 in Jabalpur, India. Hemshankar spent his childhood, schooling and initial college at his ancestral place Bhagalpur, Bihar. He ranked second in the ISC examination of Bihar Board examination.

He received his Bachelor of Technology with honours degree in metallurgy from the Indian Institute Technology, Kharagpur, West Bengal, India in 1962, then a degree of Master of Applied Science, University of Toronto, 1963 and Doctor of Philosophy, from the same university in 1966. He joined the Indian Institute of Technology, Kanpur as a lecturer in 1967 and continued as an assistant professor till 1980. In between, he spent his time at Pilkington Brothers Ltd., Lathom, England, (1977-1979) initially as a senior technologist and then as a manager, R&D. He was extremely happy when he got the opportunity of teaching at his alma mater IIT, Kharagpur as a professor (1980-1990) where he also carried out the responsibility of Dean, Student Affairs (1988-90). To all the students, he was manna from heaven because of devising several plans for their welfare.

'A new horizon opened in my life'- that is what Ray said when he took over the charge of a laboratory of the Council of Scientific & Industrial Research (CSIR), Regional Research Laboratory [now Institute of Mineral & Materials

Technology (IMMT)], Bhubaneswar, Orissa as its Director departing the academic arena of work so far in 1990 and joining the technology of application world. To him it was a ‘great opportunity to interact with industries and face their real life problems.’ He steered the R&D activities for ten years. During his tenure as a Director of RRL he served as a consultant to Research and Development, Steel Authority of India Ltd, Ranchi, (1985-1989), and Metallurgical and Engineering Consultants (India) Ranchi, (1993-1994). The staff of IMMT will remember him for his leading from the front in rescue operation in the worst disaster with storm/rain in 1999, October at Bhubaneswar. He has been listed as a noteworthy laboratory administrator; researcher by Marquis Who's Who.

Finally, he anchored himself at Central Glass & Ceramic Research Institute (CGCRI) at Kolkata when he joined as an Emeritus Scientist of CSIR in 2000 and continued for five years and thereafter three more years as Emeritus Scientist of AICTE. After these assignments, he taught at Indian Institute of Engineering Science and Technology, Sibpur for four years as an adjunct professor.

Among the awards, he received National Metallurgist Day award from Govt. of India, 1984, Kamani Gold Medal from Indian Institute of Metals, 1979, G.D Birla award, Indian Institute of Metals in 1991, Silver Jubilee Research award from IIT, Kharagpur and Distinguished Alumnus Award IIT, Kharagpur, 2014 are a few noteworthy ones.

A prolific reader and writer, he was proficient in articles of different genres, be they related to science or otherwise. Writing 500 technical papers and 24 books including four on management science and six on other topics goes to his credit and the last book published a couple of months before titled; ‘Mind Your Mind’. His tender attributes made him so humane. A story writer, singer, esraj, tabla player, painter and an amateur sorcerer coupled with his research work he deserved to be called a true polymath. He knew the magical trick –how to endear people with ease whoever came in touch with him. By his passing away IAPQR is deprived of his guidance and becomes much weaker in vigour: but the losses to the society are manifold, as it loses a human being who has been a beacon of hope and inspiration to many. Society loses a Man, a human being.

Subrata Ghosh

IAPQR Transactions
Vol. 47, No. 1-2, 2022–2023

WHITHER STATISTICS?

S.P. MUKHERJEE*
Mentor, Indian Association for Productivity, Quality and Reliability

***Abstract*:** The present article reflects on a few issues regarding the on-going trend in theory and applications of Statistics. The aspect of applications and the problems with sample surveys in several domains have been emphasized.

***Keywords and Phrases*:** Probabilistic Induction, Akaike Information Criterion, Human Development Index, Human Poverty Index, Disease Adjusted Life Years, Fault Tree Analysis, Labour Force.

1. INTRODUCTION

Recent times have witnessed the emergence of several distinct trends in the pursuit of Statistics and its applications that cannot be readily reconciled one with the other. Some look upon Statistics as a highly sophisticated scientific-method that alone can take care of uncertainty – of course to some extent only-prevailing in every field of human enquiry. Going beyond 'deduction', as providing the building block of Pure Mathematics, Statistics makes use of 'abduction' and 'induction' based on data that are regarded as realizations of random variables. Statistics remains a somewhat 'abstract' branch of knowledge. Contributions by this 'exclusive' band of scientists often involve quaint symbols and/or high-flown mathematical logic. A second group focuses on data and data analysis, supported by computerized methods and tools primarily to throw some light on an observed phenomenon affected by uncertainty. A continuing middle path works on probability models to take care of uncertainty in observed data and model-based inferences, sometimes

*Invited Article; Communicating email: prof.mukherjee@gmail.com
DOI-https://doi.org/10.32381/IAPQRT.2023.47.01-02.1

venturing into model-free inductions also. Alongside we have a wide spectrum covering naïve applications of descriptive statistics (including relevant parts of categorical data analysis) to applications of sophisticated methods and techniques of analysis applied to data which have been collected by adopting some valid designs through applications marked by routine use of softwares for methods and techniques that suit needs of the investigation. We also have to reckon with the rapidly increasing use of statistics to provide inputs for planning, implementing, monitoring and evaluating policies in both public and private sectors. And in this context also, one comes across a host of developments marked by use of context-specific non-traditional methods for data collection and data analysis.

From being simple 'political arithmetic' or being branded as 'statecraft' to being accepted and appreciated as a scientific method cutting across disciplines and applied by a host of investigators in a wide array of social, economic and natural phenomena to enrich the substantive content of their findings must be acknowledged as a big transition. Such a transition could be brought about by the visionary and diligent work of many thinkers who would not forget the roots and would not remain satisfied with only the stem growing. They would always look beyond. and sometimes at situations and solutions far removed from the present scenario.

And those who contributed the most towards the big transition were not motivated just by the intellectual urge for more sophistication but were guided more often by the urge to satisfy a felt need to improve the existing state. Of course, the felt need could pose a challenge and that is what has been taken up by the big contributors.

The present article considers only a few issues to reflect on the on-going trend in theory and applications of Statistics. These concern probability models to represent different phenomena, to extract information and draw conclusions from observed data, and use of statistical methods in diverse investigations to enhance the substantive content and hence the knowledge derived from such investigations on many emerging aspects of the Economy, the Society, and the Environment. The use of sample surveys and the role of summary measures like indices have been discussed briefly.

2. LOOKING BEYOND AND AHEAD

Statistics with its emphasis on data-based inductive inferencing supported and assessed in terms of probability considerations, has often been used and will continue to be used in peeping into the future. Prediction (with similar tasks) has been a big commitment for Statistics. Of interest to us in this context is the emphasis by Hume on the assumption of similarity in behaviour of the underlying entity in future as has been observed so far to justify probability-based induction. At the same time, we all appreciate that the only 'constant' in the universe is 'change'. Thus, continuity of the same behaviour is at stake. Bayesians have attempted to provide some answer to this vexing problem.

Trying to look ahead based on what has happened in the past and on some assumption regarding changes likely to take place in future, we come across exercises branded as predictions, projections and forecasts, leaving aside estimates usually relating to the near future. Some statisticians tend to differentiate among these on different grounds like the data needed, the methods used, the assumptions made, and the time horizon covered, while others take such differentiation as nothing significant.

A major concern of Statistics is to make inductive inferences based on empirical observations (as an important element of the 'premises') based on probabilistic logic. Such inductions proceed from the observed 'part' to the 'whole' with a large 'unobserved' remainder. Thus, we may like to infer about the future, based on our past and current observations on a dynamic phenomenon. The other aspect of inductive inferencing to look beyond the data in the premises has been being widely used to arrive at conclusions meant to validly apply to the entire domain of interest from which the observed data constitute a sample. Sometimes, the sample is selected or designed to facilitate the process of inferencing, while in many situations the observed data are taken to represent a (random) sample. Going beyond the part to comprehend the whole is again an exercise in uncertainty, caused by likely differences between the part and the remainder of the whole.

Hume (1748) questioned the propriety of the assumption that the unobserved future will behave in the same way as the observed present. Thus a model developed from a part (training) of the observed present (including past) and tested to yield a good fit (based on the validation part) may not be unique and may necessarily hold good for the unobserved future. To resolve Hume's problem, Bayesian methods have been adduced, while in some other instances the problem has been just ignored.

The problem merits special attention in situations where we like to project or predict some aspect of a rare phenomenon (with few past observations) but having a huge impact. The problem of estimating a very small probability has been sometimes attempted using Fault Tree Analysis. This approach, however, implies the availability of credible estimates of probabilities for the base (bottom)–level events, using sufficient data. Problems of this nature lie in the 'fourth quadrant' and point out to limitations of Statistics as mentioned by Taleb (2008). Uncertainty about the future can at best be reduced, but never completely removed.

3. Use of Probability Models

To make some model-based inference using an observed data set, we require to choose some probability model that would fit the observed data 'satisfactorily' To validate the choice, we compute the value of a statistic which is primarily based on the divergence between the observed frequencies or cumulative proportions or some other sequence of measures and their counterparts expected from the model (using the model estimated from the sample observations) and use an asymptotic distribution for this statistic.

The last few decades have witnessed the emergence of a virtual deluge of probability models criss-crossing one another, obtained through several methods like exponentiation, transmutation, compounding, use of quantile function (of a reference distribution), incorporation of some additional parameter(s) and the like. Few new ideas are involved, several known methods and results in mathematics have been used, and the resulting models have been shown to be more flexible to accommodate some typical behaviour of real-life data sets. In fact, just in the context of reliability and survival analysis, one can comfortably cite several hundred articles which deal with these newly developed models and their properties. Some of the models involve as many as six parameters. Incidentally, Johnson had remarked that not more than four parameters are needed to meaningfully represent features of interest in various data sets. It is, of course, true that access to powerful computing facilities can downplay the problem of estimating more than a handful of parameters, at least in terms of providing reasonable approximations. And the contributors point out a better fit to some typical data set by a newly developed, though somewhat cumbersome, probability model than by the existing classical and simpler models. And in some cases, these 'typical' data sets could be 'pathological'.

Just to illustrate the above point, we may consider the problem of modelling life-time or time-to-failure data in reliability and survival analysis. The most widely used one-parameter exponential model with density and reliability (survival) function given by

$$f(x) = \lambda \exp(-\lambda x) \text{ and } S(x) = \exp(-\lambda x)$$

and a constant failure rate λ fails to represent non-constant (age-dependent) failure rate situations. Thus, a two-parameter generalization viz. the Weibull model with a monotonic (increasing or decreasing) failure rate was proposed with the density function,

$$f(x) = \alpha \lambda x^{\alpha - 1} \exp(-\lambda x^{\alpha})$$

where the shape parameter α determines the failure rate pattern with a value more than unity yielding an increasing failure rate, a value less than unity corresponding to a decreasing failure rate and a unity value reducing the distribution to the exponential. It was subsequently pointed out that some failure mechanisms could justify a unimodal or inverted bath-tub shaped failure rate behaviour and hence the inverse Weibull distribution came to be studied. Several mechanisms generating failure times in real life situations were identified to result in data that could be appropriately fitted by the inverse Weibull distribution with the survival function given by

$$S(x) = \exp(-\alpha x^{-\lambda}), \ \alpha, \lambda > 0$$

In fact, the inverse Weibull distribution was found to provide a good fit to data relating to degradation phenomena in mechanical components like crankshafts and pistons in diesel engines. And the added advantage was not at the cost of any additional parameter. With a proportional reversed hazard behaviour, this model is quite appealing and simple. The inverse Rayleigh distribution which was developed independently by Voda (1972) is just a special case corresponding to $\alpha = 2$. A reflected Weibull distribution obtained by rotating the Weibull distribution about the equiangular line was proposed by Cohen (1973) to get an alternative to Pearsonian type VI distribution. However, this has not found much use in reliability analysis. To accommodate both unimodality and an initially decreasing pattern, Gusmao et al. (2013) proposed a three-parameter generalized (really an exponentiated) inverse Weibull model. Oluyede and Yang (2014) provided a detailed review of generalizations of the inverse Weibull distribution and corresponding applications.

To develop a wider class, a Beta distribution was taken as the generator with the generalized inverse Weibull as the parent distribution. Later came the Reflected Beta generated Generalized Inverse Weibull distribution. No doubt, this class is pretty general that includes as particular cases many existing simpler models involving fewer parameters. Failure characteristics of this model are quite complex, estimation of parameters even through numerical methods is not convenient and one may not be able to associate any worthwhile failure mechanism with this model.

All this relates only to one direction in which the simple and useful Weibull model has been extended or modified. Extensions and generalizations on the ground (pretext?) of covering wider patterns of variation in observed data sets are taking place rather rapidly on a continuing basis with no end in sight. One can conveniently cite more than a few dozens of published research work just concerning the Weibull distributions. No wonder we have a handbook by Rinne (2008) devoted entirely to the Weibull distribution.

After all, if a simpler model can provide a reasonably good fit to the data on hand, the necessity or desirability of studying a more general distribution in practical applications is questionable. Of course, there is pleasure in deriving such models and working their properties in terms of a mathematical exercise. One wonders if this motivated quite a few scholars working on extensions, modifications, and generalizations of such models.

The claim for a better fit to an observed data set by a more general and usually a more sophisticated model with more parameters is mostly pointed out in terms of penalized maximized likelihood or the model-reproduced covariance matrix computed at the sample values. We must remember that these measures of goodness of fit are only approximately obtained by us and small differences in the measure should be accepted with a pinch of salt. Moreover, the data sets considered in some of the contributions may be branded as pathological' and much reproducible. All these considerations may well pose the question "how significant are these developments?"

Among the more commonly used distance (divergence)–based tests of goodness-of-fit, the Chi-square and the Kolmogorov-Smirnov tests are worth mentioning. In the first case, one problem that is likely to surface is the possibility of test results leading to different conclusions or providing different degrees of support to the model choice as may result from different groupings of the observed data. Adoption of two different numbers of classes and/ or two different choices of class-boundaries will generally result in two different values of the statistic.

The second problem that is common to any distance-based procedure is the fact that the model-expected frequencies or cumulative proportions or other measures depend on the estimates of the model parameters as derived from the observed data. And there do exist in most cases more than one method for such estimation. There cannot be any guarantee that the statistics summarizing the divergences between observed and expected quantities will remain the same, irrespective of the method of estimation used. Moreover, even when we use a particular well-accepted method of estimating the parameters in the initially chosen model, we should probe whether the associated standard errors are relatively small or not. In fact, a robustness investigation may be taken up to find out if the same conclusion about fit of the model to the observed data set holds if we compute expected results based on the upper and lower confidence limits for the estimate of each of the model parameters estimated.

In case we use any criterion like Akaike Information or Bayes Information, problems as above cannot be avoided. completely. When we compare two models for their ability to fit the given data set, we require to compute the likelihood of the observed data assuming each model and this depends on the estimated parameter values using the method of maximum likelihood. Though the method of parameter estimation remains the same, likelihood values computed at the upper and lower confidence limits of the estimated parameters may not yield the same choice between the models in a pair and eventually may not result in the choice of the same model in preference to all others.

One more point to be noted is the fact that with the current availability of a whole host of models, including quite a few more or less similar to one another in several respects, it is possible that instead of just fixing the choice of a particular model and validating the same by way of a goodness-of-fit test, it may be worthwhile to consider a few more models for fitting the observed data and making a final choice based on other considerations like simplicity and the like.

4. Quantification and Summarization

Quantification of all the concepts and constructs involved in a study is not a mandate for fruitful applications of statistical methods and techniques to enrich the substantive content of the study. Since the emergence and development of comprehensive methods for dealing with categorical data, one can avoid quantification in situations where some amount of subjectivity is inherent in any attempt to quantify a qualitative feature. Of course, there exist cases where we need some sort of a scoring or scaling to evaluate such a feature before putting an individual value in some category. For example, we tend to note an individual's intelligence by applying a mental test, scoring the responses and placing the final score within some pre-assigned range. In any exercise to quantify a concept—primary or derived—a certain extent of subjectivity or selectivity can hardly be avoided. Notwithstanding this limitation, more and more qualitative features are being quantified—not necessarily uniquely, nor always properly.

When we make use of statistics relating to a phenomenon like participation of people belonging to the eligible group in the 'labour force' or 'incidence of unemployment' in the 'labour force; for the purpose of temporal or spatial comparison or even for the purpose of identifying the factors that explain the current level, we should attach the highest priority to developing and accepting a standard definition of the underlying variables and to adoption during data collection. There are, and there can exist different operational definitions of a term like 'labour force' or of 'an employed person', however, comparisons are justified if and only if the same definitions are used. Thus, one can define the labour force as the totality of all persons in an age-group like 15-59 or15 + or 10 + or18 + (depending the ground reality prevailing in a given context) who are currently available for some productive work (including creation and delivery of service) and are actively engaged in some economic activity or are searching for some such activity. We have to specify whether productive work or economic activity should include work done without pay or profit and / or done wholly for own consumption or should include only work done for others or for the market (at least partly) and against pay or profit. To come out with a national picture about employment-unemployment, more important than developing and using a sound sampling design with few assumptions is the need to develop and practice a standard definition of various terms and phrases used in data collection. No amount of sophistication in sampling design, in taking acre of non-response and in preparation of tables can compensate for deficiencies in data resulting from non-uniform operational definitions adopted during data collection.

Apparently comparable figures compiled through two different surveys bearing on the same phenomenon may not really throw up figures which are strictly comparable. Thus, the number of 'employed' persons found in a household-based survey seeking information about usual principal status of household members should not be compared with the total employment in an establishment-based survey where information is sought from establishments about persons engaged in those establishments. Based on usual principal status, there are household members who are not working in any establishments, not even in household establishments. And the possibility of counting a self-employed person more than once cannot be ruled out in an establishment-based survey also. Thus, what is badly needed is to understand the operational definitions used to collect information in any survey before we jump on comparing or combining figures from related surveys.

5. PITFALLS IN SUMMARISING

Aggregation and averaging are quite often used to conveniently comprehend the connotation of a relatively large set of data or, simply put, to summarise the data or to replace the data set by a single figure. In certain situations, we find ratios between two aggregates or take just the proportion of units or individuals exhibiting some specified feature (s) within a given aggregate of such units or individuals. In the process, we run the risk of missing a lot of information contained in the data set. An important illustration of this fact is provided by the use of income per capita to assess economic development of a country or a region. Gross Domestic Product (GDP) is widely accepted as a measure of national Income (though its computation may involve the product approach, the expenditure approach and the income approach to deal with different sectors of the National Economy) and this total is divided by the total population. The denominator includes the non-productive segment of the population or, say, the segment outside the labour force currently. Thus a comparison of the per capita income across countries or even over time periods may not always valid if features of the denominator vary. The more important problem concerns the fact that the single measure viz. GDP per capita provides no idea about inequality in incomes. We all appreciate the fact that economic growth and economic development are not identical and that two important constraints to be satisfied by a situation of economic growth continuing for several years in a row to be recognized as one of economic development are that inequality in incomes must not increase and that the number below an absolute poverty line shall not increase. Evidences in many countries indicate that these constraints are not satisfied even when per capita income has been growing, clearing pointing to accumulation of income in the hands of a few at the cost of increasing impoverishment of large segments of the population.

The deficiency of GDP to take cognizance of the depletion or degradation of natural as well as man-made resources during the production process renders it inappropriate in the context of sustainable development. This limitation of GDP can be partly removed by considering the Net Domestic Product (NDP) which is not generally computed, posing as it does quite a few problems in accounting for resource depletion or degradation during current production. The fact that workers from some relatively poor countries participate in the production processes in some country and add to the latter's GDP without being a part of its population will lead to an exaggerated per capita GDP for the country. With 'migrant' workers not considered as part of the country's population.

Several such summary measures are often combined into some single indices which are used to compare or rank countries in terms of their performance on economic or social or political fronts. The most notable example is the Human Development Index (HDI) computed by the United Nations Development Programme (UNDP) along with the Human Poverty Index and several development-related Indices. As is well-known, the Human Development Index is the unweighted average of three component indices related to Health, Education and Income. All the four parameters involved are averages or ratios of aggregates. Let us consider the parameter "combined gross enrolment ratio" and the way it is computed in various countries from the relevant administrative records generated and used by them. Problems arise from definitions of 'tertiary' level of education to limiting ages for enrolment, recognition of open education courses as also non-formal courses offered off-line or on-line, collating data from privately managed institutions, and so on. Similar problems are faced in respect of other parameters as well. The very concept of an international dollar or purchasing power parity dollar invites several problems in its quantification. And then comes the question of adjusting the per capita GDP to account for the fact that beyond a certain threshold, a rise in per capita income does not lead to a proportionate increase in quality of life.

To remove some of the drawbacks in HDI as an indicator of human development, UNDP has put forward another index viz. Human Poverty Index (HPI) which again is based on averages and suffers from limitations associated with the use of averages.

Even working out a framework defined in terms of aggregates and averages, we need to bring in methodological improvements to focus on details and to extract more information about the underlying phenomenon. Let us take the case of health care. One major aspect of health is revealed by mortality rates which are usually computed separately for the two sexes and for different ages/ age-groups. In fact, these rates are essential inputs for constructing life tables and deriving the life expectance or complete expectation of life at age 0 (a parameter involved in HDI). However, to reveal the impacts of different diseases or disease groups, we may calculate the aggregate years of life lost due to some diseases. Going beyond to take into account the morbidity aspect and its impact, we can even go for a more sophisticated figure viz. disease adjusted years of life (DALY) and the Disease-adjusted life expectancy. The impact of morbidity on productivity and hence on national income-somewhat indirectly though is not difficult to appreciate, though this has not been quantitatively analysed.

Going beyond the quite defensible price indices required in various contexts and the oft-used measures of poverty including the multi-dimensional poverty index used in some countries and the widely accepted HDI, we come across a virtual deluge of indices which suffer from indirect, inadequate and sometimes even incorrect quantification of the underlying variables, inadequate sample sizes and doubtful data. A few to mention are Global Hunger Index, Cato Institute of Human Freedom Index, World Press Freedom Index, Democracy Index, Happy Planet Index and the Human Happiness Index. Some of these indices make use of administrative records which usually suffer from lack of credibility and incomplete coverage besides a considerable time lag. Others make use of results thrown up by opinion polls carried out on selective samples of respondents and some others make use of both. All these summary measures hide more than what they reveal about the situation that really prevails.

6. The Problem with Sample Surveys

To meet requirements of research investigations as well as of policy planning, monitoring and evaluation, we need estimates of parameters characterizing populations or domains of interest which are usually quite large. And limitations on resources coupled with the need for data within stipulated time, we take recourse to sample surveys. It can be easily understood that the suitability of the estimates derived from the unit-level responses for any of the above two purposes depends on the survey design including the sample size and the approach adopted to elicit and record unit level responses. Many sample surveys carried out by government agencies or private bodies do not take due care of these two aspects. Of course, there exist genuine problems and to tide over those we need to make reasonable assumptions, not necessarily warranted by available evidences. Thus, for example, the sample size in many sample surveys is determined by stipulating a fixed relative standard error of the estimate to be thrown up for the parameter of interest, like the population total or the mean of a study variable. We should first note that the relative standard error (which is the same for both the estimated total or the estimated mean) is essentially the coefficient of variation of the estimate and hence is unknown. It can only be estimated in terms of the sample data once these are available. However, for the purpose of determining the sample size, we can only 'guess' its value by assuming some likely value for the population standard deviation. And, that way a 'guessed' value for the r.s.e. can be equated to a pre-specified value like 10% or 5% to find the sample size. To add to our worry, most surveys are used to estimate more than one parameter sometimes related to more than one study variables. The associated problem of sample size determination remains to bother us, particularly in surveys taken up for the first time.

Findings from surveys with a small sample size for a large country should be accepted only with a big pinch of salt. And, strangely, some such surveys with grossly inadequate sample sizes are sometimes carried out by some 'international 'agencies to compare countries in respect of parameters like charity-giving index appearing on the web. Even when the sample is reasonably large, quality of data as are elicited in terms of responses from informants in households or establishments suffers badly on several counts including inadequate training of investigators, lack of relevant information with the informants, denial or non-cooperation from the informants, etc.

Sample surveys meant to generate useful information on a situation of interest like the conditions of living and work by 'domestic workers 'adopt some definitions of terms and phrases which are accepted more or less internationally to enable cross-country comparisons. In this case, 'domestic workers 'include persons like private tutors engaged by households to coach some members of the household in terms of some agreement between the household and the tutor Similarly, physicians or healthcare professionals engaged by some households to check health status and / or suggest appropriate treatment of members of the household on a regular basis should also be recognized as domestic workers. However, unless investigators specifically ask probing questions about the engagement of such domestic workers, responding households are quite likely to miss these types of domestic workers. May be a few households or a few investigators realize this point and we get some responses. Once processed, the summary may present a misleading picture. Arguing that data relating to such professional domestic workers are really not of interest to us can only justify narrowing down the scope of 'domestic workers' by excluding such professionals.

A point to remember in connection with sample surveys or, for that matter, even for censuses or complete enumeration studies is that not all information that we need in respect of the units or individuals of interest can be collected in terms of responses given out by the informants for a host of reasons. In fact, a question like how much calorie intake can be attributed to a member of a household cannot possibly be worked out simply by noting the quantity of each food item consumed by the household during a certain reference – carefully chosen to be representative, avoiding recall lapse, and making use of some table to find out the calorie content per unit of each food items. One obvious difficulty arises from the fact that some members may consume some food items outside the household and the informant on behalf of the household may have no idea about the items taken and their quantities. Incidentally, such items are often quite calorie rich. What is important to make sense out of survey data in such situations is to carry out small-scale type studies by specially trained investigators using the participatory approach, if possible, to collect detailed data and thus to derive some idea about the credibility of usual survey data.

7. An Issue for the National Statistical System

In India as also in quite a few other countries there exist National Statistical Systems to collect, edit, and analyse data bearing on a whole range of issues, including data needed as inputs in policy formulations. In some situations, such data to be collected once for all to examine the possible impact of feasible policy alternative and policy interventions have to be partly collected from existing administrative records and partly from some specially designed sample survey. Let us consider one such situation that has not been taken into account anywhere so far and may not be that important at the national level in any country. However, this illustration is just indicative of some points that should be kept in mind by a National Statistical System. The national government of a country would like to reduce the existing extent of income inequality within its population and has identified the following two possible routes:

(1) increase rates of corporate tax to raise more funds to allow for increase in allowances and subsidies to the disadvantaged sections of its people and (2) retain corporate tax as it is and insist on higher wages and salaries to employees in the corporate sectors. The first is likely to dampen investment in production within the country and the second is likely to reduce operating surplus, again leading to reduced investment.

Any decision by way of a choice between the two would require some relevant figures. How accurate do we need these figures to be? And how do we go about getting these figures? Looking at administrative records relating to direct taxes paid, we require to check the quality of data, carry out imputations for missing entries or outliers by adopting appropriate methods, and development a meta-data base by integrating such records with results of sample surveys. Data Integration should be an important exercise, that can even throw up—as it has already done—interesting theoretical issues in Statistics and Computer Science.

Coming to specially designed sample surveys to generate data of the type mentioned, the problem of sampling design becomes a challenging task. Any such survey involves several estimation variables, requires estimation of more than one parameters, cannot derive much strength from designs and findings of previous surveys and has to be guided by considerations of time and cost primarily. The use of relative standard error and of design effect as also schemes for stratification as well as for selection of ultimate stage units has to be made with due caution and with recognition of the possible impact on data quality. Of course, control of non-sampling errors continues to remain a vexing problem.

8. Concluding Remarks

In the ultimate analysis, Statistics—covering both statistical methods and statistical data—should serve the purpose of National Development, providing inputs and instruments for decisions and actions by all concerned. Thus, we need to convey messages through findings of relevant investigations to the entire community of decision-makers at all levels in a language they understand and not in a manner that is not intelligible to a large section of such people. Simplicity does not necessarily imply a compromise with rigour.

Developments in Statistical theory and Methods are also expected to describe, analyse, predict, and even optimize aspects and features of interest in almost all branches of human enquiry. And such aspects and features are affected by chance ascribable to a host of factors including complexity and resulting partial ignorance as well as those associated with attempts to reveal truths behind the underlying phenomena.

In both the areas, the end or the objective should justify the means reflected in the level of sophistication of the models, methods and techniques to be developed and/or adopted. In fact, this level should differ from the case of an exploratory research to a confirmatory one, from a research in the perceptual world to one in the conceptual world, from a one-shot investigation into an incident to a longitudinal study of an evolving phenomenon, from a study where more than adequate data are conveniently accessible to one in which data are sparse and from a situation where uncertainty can be ascribed to partial ignorance and consequent lack of control to one in which fuzziness is an inherent feature of available data.

Curious readers get confused in the deluge of models, methods and techniques along with softwares proposed to study the same phenomena and sometimes get lost in unnecessary sophistications and complexities.

One wonders if the trend of developments indicate a configuration resembling that of a pyramid, with the base representing the highest extent of sophistication in dealing with data (in a generic sense) that justify few of the assumptions involved in using the data which conform substantially to the requirements of a traditional or classical method of analysis. The fundamental conflict between sophistication in analysis and uncertainty about the underlying phenomena poses almost a baffling problem.

References

1. Cohen, A.C. (1973): The reflected Weibull distribution, *Technometrics,* 15(4), 867-873.
2. Gusmao, F.R.S. (2011): The generalized inverse Weibull distribution, *Statistical Papers*, 52, 591-619.
3. Hume, D. (1748): *An enquiry concerning human understanding.* Oxford University Press.
4. Oluyede, B.O. and Yang, T. (2014): Generalizations of the inverse Weibull distribution with applications, *Electronic Jour. App. Statist. Analysis,* 7 (1), 94-116.
5. Rinne, H. (2008): *The Weibull Distribution: A Handbook.* Chapman and Hall, USA.
6. Taleb, N.N. (2008): *The Fourth Quadrant:* Map of the Limits of Statistics. Edge.org.
7. Voda, V.G. (1972): On the inverse Rayleigh distributed random variable, *Rep. Statist. App. Res. JUSE,* 19, 13-21.

IAPQR Transactions
Vol. 47, No. 1-2, 2022–2023

USE OF RANKED SET SAMPLING IN THE ESTIMATION OF PARAMETERS μ_2 AND σ_2 OF MORGENSTERN TYPE BIVARIATE EXPONENTIAL DISTRIBUTION

N. K. SAJEEV KUMAR[1*] AND SUMI. A. R.[2]
[1,2]*Department of Statistics, University College, Trivandrum, Kerala, India.*

Abstract: In this article we have derived the best linear unbiased estimator (BLUE) of the parameters μ_2 and σ_2 involved in the Morgenstern type bivariate exponential distribution by ranked set sampling (RSS). Also we have evaluated the BLUE's of μ_2 and σ_2 involved in the Morgenstern type bivariate exponential distribution by concomitants of ranked set sampling for some specific values of the sample size n. The efficiency comparison of the proposed estimators with that of the estimators using concomitants of order statistics and modified ranked set sampling is also made.

Keywords and Phrases: Morgenstern Family of Distribution, Morgenstern Type Bivariate Exponential Distribution, Best Linear Unbiased Estimation, Concomitants of Order Statistics, Concomitants of Ranked Set Sampling, Modified Ranked Set Sampling.

1. INTRODUCTION

The concept of ranked set sampling was first introduced by McIntyre ([8], 1952) as a process of improving the precision of the sample mean as an estimator of the population mean. Ranked set sampling as described in McIntyre ([8], 1952) is applicable whenever ranking of a set of sampling units

*Corresponding author email: sajeevkumarnk@gmail.com
DOI-https://doi.org/10.32381, IAPQRT.2023.47.01-02.2

can be done easily by a judgment method, for more details, see Chen et al., ([4], 2004). The process of ranked set sampling involves selecting n sets of units, each of size n, and ordering the units of each of the set by judgment method or by using values obtained on the units by a relatively inexpensive method, without making actual measurement on the units. Then the unit ranked as one from the first set is actually measured, the unit ranked as two from the second set is measured. The process continues in this way until the unit ranked as n from the n^{th} set is measured. Then the observations resulting out of actual measurements made on the units chosen as described in the above process is known as ranked set sample.

Ranking by a judgment method is not recommendable in case when there is ambiguity in discriminating the rank of one unit with another (for example, ranking based on intensity of yellowish color of newborn babies, as mentioned in Chen et al., ([4], 2004, p. 7), become difficult, as we known that probability of ties in this situation is not zero as the characteristic under consideration is not continuous). In certain situation, one may prefer exact measurement of some easily measurable variable associated with the study variable rather than ranking the units by a crude judgment method. Suppose the variable of interest say Y, is difficult or much expensive to measure, but an auxiliary variable X correlated with Y is readily measurable and can be ordered exactly. In this case as an alternative to McIntyre's ([8], 1952) method of ranked set sampling, Stokes ([13], 1977) used an auxiliary variable for the ranking of the sampling units. Stokes ([13], 1977) using auxiliary variate is as follows: Choose n^2 independent units, arrange them randomly into n sets each with n units and observe the value of the auxiliary variable X on each of these units. In the first set, that units for which the measurement on the auxiliary variable is the smallest is chosen. In the second set, that unit for which the measurement on the auxiliary variable is the second smallest is chosen. The procedure is repeated until in the last set, that unit for which the measurement on the auxiliary variable is the largest is chosen. The measurements on the Y variate of the resulting new set of n units chosen by the above method gives a ranked set sample as suggested by Stokes ([13], 1977). If $X_{(r)r}$ is the observation measured on the auxiliary variable X from the unit chosen from the r^{th} set then we write $Y_{[r]r}$ to denote the corresponding measurement made on the study variable Y on this unit so that $Y_{[r]r}$, r =1, 2, …, n from the ranked set sample.

Clearly, $Y_{[r]r}$ is the concomitants of r^{th} order statistic arising from the r^{th} sample. By using the expression giving by Yang ([16], 1977) for the mean and variance of concomitants of order statistics, we have the following. For $1 \le r \le n$,

$$E[Y_{[r]r}] = E[E(Y \mid X_{(r)r})] \tag{1}$$

and

$$Var[Y_{[r]r}] = Var[E(Y \mid X_{(r)r})] + E[Var(Y \mid X_{(r)r})] \tag{2}$$

Since $Y_{[r]r}$ and $Y_{[s]s}$ for $r \neq s$ are drawn from two independent samples we have,

$$Cov[Y_{[r]r,[s]s}] = 0, \; r \neq s. \tag{3}$$

Stokes ([14], 1995) has considered the estimation of parameters of location-scale family of distributions using RSS. Lam et al., ([6], 1994, [7], 1995) have obtained the BLUEs of location and scale parameters of exponential and logistic distributions using ranked set sample. Applications of ranked set sampling on exponential distributions are available in Abu-Dayyeh and Huttlak ([1], 1996), Barnett and Moore ([2],1997), Muttlak and Abu-Dayyesh ([9], 2010), Samuh and Qtait ([10],2015), Sengupta and Mukhuti ([12], 2008). Stokes ([13], 1977) has suggested the ranked set sample mean as an estimator for the mean of the study variable Y, when an auxiliary variable X is used for ranking the sample units, under the assumption that (X, Y) follows a bivariate normal distribution. Barnett and Moore ([2], 1997) have improved the estimator of Stokes ([13], 1977) by deriving the Best linear unbiased estimator(BLUE) of the mean of the study variate Y, based on ranked set sample obtained on the study variate Y.

1.1. Morgenstern family of distributions

The most important aspect to the considered in a modeling problem is regarding the flexibility of the chosen family, in the sense that it should contain a wide variety of models capable of representing any data situation. Another consideration is the sort of prior information available in the choice of the model. In problems involving several random variables, the analyst may make reasonable assumption about the marginal distributions. Then the question is to construct a joint distribution function with a set of given marginal. The Morgenstern family of distributions assumes importance in such contexts as a highly flexible system.

Let X and Y be two random variables with cdfs given by $F_X(x)$ and $F_Y(y)$ respectively with pdfs $f_X(x)$ and $f_Y(y)$ and jointly distributed with cdf $F(x, y)$ given by (see, Kotz et al., ([15], 2000), p. 52).

$$F(x,y) = F_X(x)F_Y(y)\{1+\alpha(1-F_X(x))(1-F_Y(y))\},\ -1 \le \alpha \le 1. \quad (4)$$

Then the family of distributions having the above form of cdf is called the Morgenstern Family of Distributions (MFD). The system provides a very general expression of a bivariate distribution from which members can derived by substituting expression of any desired set of marginal distributions. The joint pdf corresponding to the cdf defined in (4) is given by,

$$f(x,y) = f_x(x)f_Y(y)\{1+\alpha(1-2F_X(x))(1-2F_Y(y))\},\ -1 \le \alpha \le 1 \quad (5)$$

The parameter α is considered as the association parameter, the two random variables are independent when the value of α is zero and the maximum correlation attained between X and Y is $\frac{\alpha}{3}$. For suitable choices of $F_X(x)$ and $F_Y(y)$, the MFD is primarily useful as an alternative to the standard bivariate normal distribution (see, Conway, ([5],1983), p. 28-31).

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be a sequence of independent observations drawn from a distribution with pdf given by (5), where $F_X(x)$ and $F_Y(y)$ are the cdfs and $f_X(x)$ and $f_Y(y)$ are the pdfs of X and Y respectively. Scaria and Nair ([11], 1999) have given convenient forms for the pdf $f_{[r:n]}(y)$ of $Y_{[r:n]}$, the concomitants of r^{th} order statistics and joint pdf $f_{[r,s:n]}(y_1, y_2)$ of $Y_{[r:n]}$ and $Y_{[s:n]}$ (concomitants of r^{th} and s^{th} order statistics) of a random sample of size n arising from a distribution belonging to MFD with cdf (4) are given below:

For $1 \le r \le n$;

$$f_{[r:n]}(y) = f_Y(y)\{1+\alpha\left(\frac{n-2r+1}{(n+1)}\right)[1-2F_Y(y)]\} \quad (6)$$

And for $1 \le r < s \le n;\quad y_1, y_2 \in R$

$$f_{[r,s:n]}(y_1,y_2) = f_Y(y_1)f_Y(y_2)\{1+\alpha\frac{(n-2r+1)}{(n+1)}[1-2F_Y(y_1)]+\alpha\frac{(n-2s+1)}{(n+1)}$$

$$[1-2F_r(y_2)]+\alpha^2[\frac{n-2s+1}{n+1}-\frac{2r(n-2s)}{(n+1)(n+2)}][1-2F_Y(y_1)][1-2F_Y(y_2)]\} \quad (7)$$

It is clear to note that we can represent $2F_Y(y)f_Y(y)$ as the pdf of the largest order statistic $Y_{2:2}$ of a random sample of size two drawn from the marginal distribution of random variable Y. Then we can write the pdf defined in (6) and the joint pdf defined in (7) as,

For $1 \le r \le n$,

$$f_{[r:n]}(y) = f_Y(y) + \alpha \frac{(n-2r+1)}{(n+1)}[f_Y(y) - f_{2:2}(y)] \quad (8)$$

And for $1 \le r < s \le n$,

$$f_{[r,s:n]}(y_1, y_2) = f_Y(y_1)f_Y(y_2) + \alpha \frac{(n-2r+1)}{(n+1)} f_Y(y_2)[f_Y(y_1) - f_{2:2}(y_1)]$$

$$+ \alpha \frac{(n-2s+1)}{(n+1)} f_Y(y_1)[f_Y(y_2) - f_{2:2}(y_2)]$$

$$+ \alpha^2 [\frac{n-2s+1}{n+1} - \frac{2r(n-2s)}{(n+1)(n+2)}][f_Y(y_1) - f_{2:2}(y_1)][f_Y(y_2) - f_{2:2}(y_2)],$$

$$y_1, y_2 \in R. \quad (9)$$

From (8) we get the expression for the k^{th} moment of the concomitant $Y_{[r:n]}$ is given by

$$E(Y_{[r:n]}^k) = \mu^{(k)} + \alpha \frac{(n-2r+1)}{(n+1)}(\mu^{(k)} - \mu_{2:2}^{(k)}),\ 1 \le r \le n, \quad (10)$$

where $\mu^{(k)} = E[Y^k]$ and $\mu_{2:2}^{(k)} = E[Y_{2:2}^k]$, $k = 1,2,..$ and we write μ to denote $\mu^{(1)}$.

Now the product moment of r^{th} and s^{th} concomitants is given by

$$E[Y_{[r:n]}, Y_{[s:n]}] = \mu^2 + 2\alpha \frac{[n-(r+s)+1]}{(n+1)} \mu(\mu - \mu_{2:2})$$

$$+ \ \alpha^2 \{\frac{n-2s+1}{n+1} - \frac{2r(n-2s)}{(n+1)(n+1)}\}(\mu - \mu_{2:2})^2, 1 \le r < s \le n. \quad (11)$$

In this work, we have derived the BLUE of the parameters μ_2 and σ_2 of Morgenstern type bivariate exponential distribution (MTBED) using concomitants of ranked set sampling. This work is an extension of the work done by Chako and Thomas ([3], 2008).

2. ESTIMATION OF THE PARAMETERS μ_2 AND σ_2 INVOLVED IN MTBED USING RSS

An important member of Morgenstern family of distribution is (MFD) is Morgenstern type Bivariate exponential distribution (MTBED). The bivariate random variable (X, Y) is said to have a Morgenstern type bivariate exponential distribution (MTBED) if its pdf is given by

$$f(x,y) = \frac{1}{\sigma_1\sigma_2}\exp\{(-\frac{x-\mu_1}{\sigma_1}) + (-\frac{y-\mu_2}{\sigma_2})\}[1+\alpha\,(2\exp\{-\frac{x-\mu_1}{\sigma_1}\}-1)\,(2\exp\{-\frac{y-\mu_2}{\sigma_2}\}-1)]$$

$$x > \mu_1\ ,\ y > \mu_2\ ;\ -1 \le \alpha \le 1\ ;\ \mu_1 > 0\ ,\ \mu_2 > 0\ ,\ \sigma_1 > 0\ ,\ \sigma_2 > 0$$

$= 0$ *otherwise.* (12)

The pdf defined in (12) is an extension of the pdf of MTBED given in Chacko and Thomas ([3], 2008). Here, we have introduced new location parameters μ_1 and μ_2. Clearly the marginal distribution of X and Y variables are univariate exponential distributions. The joint pdf of standard MTBED is obtained by making the substitutions $U = \frac{X-\mu_1}{\sigma_1}$ and $V = \frac{Y-\mu_2}{\sigma_2}$ in (12) and is given by,

$$f_{U,V}(u,v) = \exp\{-u-v\}[1+\alpha(2\exp\{-u\}-1)(2\exp\{-v\}-1)],\ u > 0,\ v > 0 \quad (13)$$

Clearly the marginal distribution of each U and V of (U,V) with pdf defined by (13) are standard exponential distributions.

Suppose X and Y are the two characteristics of interest on the units of a population and (X, Y) has a distribution with pdf (12). Suppose there are n independent random samples each of size n drawn from (12) and in each sample, units are ranked according to the values of the observations measured on the easily measurable variable X on the units. Though the variable Y is of primary interest, we assume that it is difficult to measure. Then to deal with the inference problem on the distribution of the variable Y we obtain a ranked set sample by measuring the value of Y on the r^{th} ordered unit of the r^{th} sample, r =1,2,…, n. We write $Y_{[r]r}$ to denote this observation measured on Y of the r^{th} ordered unit of the r^{th} set, r = 1,2,…, n. Then clearly $Y_{[r]r}$ is distributed as the concomitants of r^{th} order statistic r = 1,2,…, n. Suppose there are n independent random samples of size n drawn from (13) and in each sample, units are ranked according to the value of the observations measured on the easily measurable variable U on the unit. We obtain a ranked set sample by measuring the value of V on the r^{th} ordered unit of the r^{th} sample,

$r = 1,2,\ldots, n$. We write $V_{[r]r}$ to denote this observation measured on V of the r^{th} ordered unit of the r^{th} set, $r = 1, 2,.., n$. Then clearly $V_{[r]r}$ is distributed as the concomitants of r^{th} order statistic of a random sample of size n arising from (13). Then the pdf $f^*_{[r]r}(v)$ of $V_{[r]r}$, $r = 1, 2,\ldots, n$ is given by

$$f^*_{[r]r}(v) = \exp\{-u\}\{1+\alpha(\frac{n-2r+1}{n+1})2\exp\{-v\}-1\}\text{ , v > 0} \tag{14}$$

Since the marginal distribution of V of (U, V) is standard exponential distributions and,

Using spacings of exponential order statistics, we have the following.

$$\mu_{2:2} = E[V_{2:2}] = \frac{3}{2} \tag{15}$$

and $$\mu^{(2)}_{2:2} = E[V^2_{2:2}] = \frac{7}{2}, \tag{16}$$

where $V_{2:2}$ is the largest order statistics of a random sample of size 2 drawn from the distribution of marginal random variable X. On substituting (15) to (16) in (10) and (11) we get, for $n \geq 1$,

$$E[V_{[r]r}] = 1 - \frac{\alpha}{2}\left(\frac{n-2r+1}{n+1}\right) = \xi_r \text{ , } for \ 1 \leq r \leq n \tag{17}$$

and $$E[(V_{[r]r})^2] = 2 - \frac{3}{2}\left(\frac{n-2r+1}{n+1}\right) \tag{18}$$

Thus the variance and covariance of concomitants of ranked set sample of size n arising from (13) are given by for $1 \leq r \leq n$,

$$Var[V_{[r]r}] = 1 - \frac{\alpha}{2}\left(\frac{n-2r+1}{n+1}\right) - \frac{\alpha^2}{4}\left(\frac{n-2r+1}{n+1}\right)^2 = \beta_{r,r},\ 1 \leq r \leq n \tag{19}$$

and for $1 \leq r < s \leq n$

$$Cov[V_{[r]r}, V_{[s]s}] = 0 = \beta_{r,s},\ r \neq s,\ r, s = 1,2,\ldots, n\ . \tag{20}$$

Since $Y_{[r]r}$ and $Y_{[s]s}$ are drawn for $r \neq s$ are drawn from two independent samples.

Let $(X_i, X_i), i = 1,2,...$ be a ranked set sample of size n drawn from a population with pdf defined by (12). Clearly we have $X_i = \mu_1 + \sigma_1 U_i$ and

$Y_i = \mu_2 + \sigma_2 V_i$ for $i = 1,2,...,n$. Then by using (18), (19), and (20) we have for $1 \le r \le n$

$$E(Y_{[r]r}) = \mu_2 + \sigma_2 \xi_r \tag{21}$$

$$Var\,[Y_{[r]r}] = \sigma_2^2\, \beta_{r,\,r} \tag{22}$$

and for $1 \le r < s \le n$,

$$Cov\,[Y_{[r]r}, Y_{[s]s}] = 0,\;\; r \ne s \tag{23}$$

where ξ_r, $\beta_{r,r}$ and $\beta_{r,s}$ are defined by (17), (18), and (20) respectively. Clearly from (17), (19), (20), it follows that ξ_r, $\beta_{r,r}$ and $\beta_{r,s}$ are known constants provided α is known. If we write

$\underline{Y}_{[n]} = (Y_{[1]1}, Y_{[2]2},..., Y_{[n]n})'$, $\underline{\xi} = (\xi_1, \xi_2,.., \xi_n)'$ and $G = ((\beta_{i,j}))$ then we can write

$$E[\underline{Y}_{[n]}] = \underline{1}\;\; \mu_2 + \underline{\xi}\;\; \sigma_2, \tag{24}$$

where $\underline{1}$ is a column vector of n ones and $\underline{\xi} = \left(\xi_1, \xi_2,..., \xi_n\right)'$.

and $$Var[\underline{Y}_{[n]}] = G\sigma_2^2, \tag{25}$$

Now we derive the BLUE of μ_2 and σ_2 involved in (12) and is given in the following theorem.

Theorem 2.1: Let (X_i, Y_i), $i = 1,2,...,n$ be a random sample of size n drawn from a population with pdf defined in (12). Let $\underline{Y}_{[n]} = (Y_{[1]1}, Y_{[2]2},..., Y_{[n]n})'$ be the vector of random samples arising from (12). Let $\underline{\xi} = (\xi_1, \xi_2,..., \xi_n)'$ and $G = ((\beta_{i,j}))$ respectively be the mean vector and dispersion matrix of concomitants of ranked set sample arising from (13). Then BLUEs of the parameters μ_2^* of μ_2 and σ_2^* of σ_2 is given by

$$\mu_2^* = -\; \frac{\underline{\xi}' G^{-1}(\underline{1}\underline{\xi}' - \underline{\xi}\underline{1}')G^{-1}}{(\underline{\xi}' G^{-1}\underline{\xi})(\underline{1}' G^{-1}\underline{1}) - (\underline{\xi}' G^{-1}\underline{1})^2} \underline{Y}_{[n]} \tag{26}$$

$$\sigma_2^* = \frac{\underline{1}' G^{-1}(\underline{1}\underline{\xi}' - \underline{\xi}\underline{1}')G^{-1}}{(\underline{\xi}' G^{-1}\underline{\xi})(\underline{1}' G^{-1}\underline{1}) - (\underline{\xi}' G^{-1}\underline{1})^2} \underline{Y}_{[n]}, \tag{27}$$

and corresponding variances are given by

$$Var(\mu_2^*) = \frac{(\underset{\sim}{\xi}' G^{-1} \underset{\sim}{\xi}) \ \sigma_2^2}{(\underset{\sim}{\xi}' G^{-1} \underset{\sim}{\xi})(\underset{\sim}{1}' G^{-1} \underset{\sim}{1}) - (\underset{\sim}{\xi}' G^{-1} \underset{\sim}{1})^2} \tag{28}$$

And $$Var(\sigma_2^*) = \frac{(\underset{\sim}{1}' G^{-1} \underset{\sim}{1}) \ \sigma_2^2}{(\underset{\sim}{\xi}' G^{-1} \underset{\sim}{\xi})(\underset{\sim}{1}' G^{-1} \underset{\sim}{1}) - (\underset{\sim}{\xi}' G^{-1} \underset{\sim}{1})^2}. \tag{29}$$

Proof: Let (X_i, Y_i), i =1, 2,.., n be a random sample of size n drawn from a population with pdf defined by (12). If $\underset{\sim}{Y}_{[n]} = (Y_{[1]1}, Y_{[2]2}, ..., Y_{[n]n})'$ be a vector ranked set sample with concomitant variables of size n arising from (12). Let $\underset{\sim}{\xi} = (\xi_1, \xi_2, .., \xi_n)'$ and $G = ((\beta_{i,j}))$ respectively be the mean vector and dispersion matrix of vector of concomitants of ranked set sample arising from (13), then by using (24) and (25) together defines a generalized Gauss-Markov set up. Then by Generalized Gauss-Markov theorem we get the BLUE's of μ_2 and σ_2 and their corresponding variances.

3. Estimation of the Parameters μ_2 and σ_2 Involved in MTBED Using Concomitants of Order Statistics and Using Modified Ranked Set Sampling

To compare the efficiency of our estimator proposed in section 2, we take two other estimators, one using concomitants of order statistics and other using modified ranked set sampling and are explained in this section.

3.1 Estimation of the parameters μ_2 and σ_2 involved in using concomitants of order statistics

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ be the observations of random sample of size n drawn from (12). The Y component associated with r^{th} order statistic $X_{r:n}$ is defined as the concomitants of r^{th} order statistic and it is denoted by $Y_{[r:n]}, r = 1, 2, ..., n.$ Also let $(U_i, V_i), i = 1, 2, ..., n$ be the sequence of independent observations drawn from (13). Let $V_{[r:n]}$ be the concomitants of r^{th} order statistic $U_{r:n}$ using (13), then the pdf of $V_{[r:n]}$ and joint pdf of $V_{[r:n]}$ and $V_{[s:n]}$ for $1 \le r < s \le n$ are respectively given in (30) and (31).

$$f^*_{[r:n]}(v) = \exp\{-u\}\{1+\alpha(\frac{n-2r+1}{n+1})2\exp\{-v\}-1\}\text{ , v>0} \tag{30}$$

And for $1 \le r \le s \le n$

$$f^*_{[r,s:n]}(v_1,v_2) = \exp\{-v_1-v_2\}[1+\alpha\,(\frac{n-2r+1}{n+1})\,(2\exp\{-v_1\}-1)+\alpha\left(\frac{n-2s+1}{n+1}\right)(2\exp\{-v_2\}-1)$$

$$+\alpha^2\left(\frac{n-2s+1}{n+1}-\frac{2r(n-2s)}{(n+1)(n+2)}\right)(2\exp\{-v_1\}-1)\,(2\exp\{-v_2\}-1)],$$

$$v_1, v_2 > 0. \tag{31}$$

Now $E[V_{[r:n]}] = \xi_r$ as defined in (17) and $Var[V_{[r:n]}] = \beta_{r,r}$ as defined in (19). Also for $1 \le r < s \le n$,

$$Cov[V_{[r:n]}, V_{[s:n]}] = \beta_{rs} = \frac{\alpha^2}{4}\left[\frac{n-2s+1}{n+1}-\frac{2r(n-2s)}{(n+1)(n+2)}-\frac{(n+2r+1)(n-2s+1)}{(n+1)^2}\right] \tag{32}$$

Let (X_i, Y_i) , $i = 1,2,..$ be sample of size n drawn from a population with pdf defined in (12). Clearly $X_i = \mu_1 + \sigma_1 U_i$ and $Y_i = \mu_2 + \sigma_2 V_i$ for i =1, 2,…n. Then, for $1 \le r \le n$,

$$E[Y_{[r:n]}] = \mu_2 + \sigma_2 \xi_r \tag{33}$$

$$V[Y_{[r:n]}] = \sigma_2^2 \beta_{rs} \tag{34}$$

and for $1 \le r < s \le n$

$$Cov[Y_{[r:n]}, Y_{[s:n]}] = \sigma_2^2\,\beta_{r,s},\ r \ne s \tag{35}$$

where $\xi_r, \beta_{r,r}$ and $\beta_{r,s}$ are respectively defined in (17), (19), and (32).

Let $Y_{[n]} = (Y_{[1:n]}, Y_{[2:n]}, ..., Y_{[n:n]})'$ be the vector of concomitants of order statistics arising from (12). Let $\underset{\sim}{\xi} = (\xi_1, \xi_2, ..., \xi_n)'$ and $B = ((\beta_{ij}))$ respectively the mean vector and dispersion matrix of $Y_{[n]}$. Then we can write $E[Y_{[n]}] = \underset{\sim}{1}\mu_2 + \underset{\sim}{\xi}\sigma_2$, where $\underset{\sim}{1}$ is a column vector of n ones and $\underset{\sim}{\xi} = (\xi_1, \xi_2, ..., \xi_n)'$ and $V[Y_{[n]}] = B\sigma_2^2$.

Now using generalized Gauss markov theorem the $BLUE's$ of μ_2 and σ_2 involved in (12) is given by

$$\tilde{\mu}_2 = -\frac{\underline{\xi}' B^{-1}(\underline{1}\underline{\xi}' - \underline{\xi}\underline{1}')B^{-1}}{(\underline{\xi}' B^{-1}\underline{\xi})(\underline{1}' B^{-1}\underline{1}) - (\underline{\xi}' B^{-1}\underline{1})^2} Y_{[n]} \tag{36}$$

$$\tilde{\sigma}_2 = \frac{\underline{1}' B^{-1}(\underline{1}\underline{\xi}' - \underline{\xi}\underline{1}')B^{-1}}{(\underline{\xi}' B^{-1}\underline{\xi})(\underline{1}' B^{-1}\underline{1}) - (\underline{\xi}' B^{-1}\underline{1})^2} Y_{[n]} \tag{37}$$

$$Var(\tilde{\mu}_2) = \frac{(\underline{\xi}' B^{-1}\underline{\xi})\sigma_2^2}{(\underline{\xi}' G^{-1}\underline{\xi})(\underline{1}' G^{-1}\underline{1}) - (\underline{\xi}' G^{-1}\underline{1})^2} \tag{38}$$

$$Var(\tilde{\sigma}_2) = \frac{(\underline{1}' B^{-1}\underline{1})\sigma_2^2}{(\underline{\xi}' G^{-1}\underline{\xi})(\underline{1}' G^{-1}\underline{1}) - (\underline{\xi}' G^{-1}\underline{1})^2}. \tag{39}$$

3.2. Estimating the parameters μ_2 and σ_2 using modified ranked set sampling

In this sub section, we derived the BLUE of μ_2 and σ_2 involved in MTBED using modified ranked set sampling similar to that of given in Sengupta and Mukhuti ([12], 2008), p. 911.

Let $X_{(i)r}$ denote the i^{th} order statistic of X observations in the r^{th} sample of size n taken from (12) and let $Y_{[i]r}$ denote the corresponding concomitants of $X_{(i)r}$, the i^{th} order statistic of X observations in the r^{th} sample, I =1, n and r = 1,2,…,n. Hence, $Y_{[1]1}, Y_{[n]2}, ..., Y_{[n]n}$ (That is in the first set take concomitants of first order statistic and remaining sets of sample take each of concomitants of n^{th} order statistic) constitute a modified ranked set sampling.

Let (U_i, V_i), I =1, 2,…,n be a sequence of independent observations drawn from (13). Let $U_{(i)r}$ denote the i^{th} order statistic of U observations in the r^{th} sample of size n taken from (13) and let $V_{[i]r}$ denote the corresponding concomitants of $U_{(i)r}$, i=1, n and r =1, 2,…,n.

Now for i = 1, n and for $1 \le r \le n$

$$E\left[V_{[i]r}\right] = 1 - \frac{\alpha}{n}\left(\frac{n-2r+1}{n+1}\right) = \psi_r \tag{40}$$

$$V\left[V_{[i]r}\right] = 1 - \frac{\alpha}{2}\left(\frac{n-2r+1}{n+1}\right) - \frac{\alpha^2}{n}\left(\frac{n-2r+1}{n+1}\right)^2 = b_{r,r} \tag{41}$$

And for i =1, n and $1 \le r < s \le n$

$$Cov\left(V_{[i]r}, V_{[i]s}\right) = b_{r,s} = 0 \tag{42}$$

Now clearly for i =1, n and r =1, 2,…,n

$$E[Y_{[i]r}] = \mu_2 + \sigma_2 \psi_r \tag{43}$$

$$V[Y_{[i]r}] = \sigma_2^2 b_{r,r} \tag{44}$$

$$Cov[Y_{[i]r}, Y_{[i]s}] = 0 = b_{r,s}, \quad r \neq s \tag{45}$$

Let $\underline{Y}_{[n]} = (Y_{[1]1}, Y_{[n]2}, ..., Y_{[n]n})'$, $\underline{\psi} = (\psi_1, \psi_2, ..., \psi_n)'$, $\Omega = \left(\left(b_{r,s}\right)\right)$. Then we can write

$$E(\underline{Y}_{[n]}) = \underline{1}\mu_2 + \psi\sigma_2, \tag{46}$$

where $\underline{1}$ is a column vector of n ones and

$$V[Y_{[n]}] = \Omega\sigma_2^2 \tag{47}$$

Now using (46) and (47) and by generalized Gauss markov theorem, the BLUE of μ_2 and σ_2 and their variances are given by

$$\tilde{\tilde{\mu}}_2 = -\frac{\underline{\psi}'\Omega^{-1}(\underline{1}\underline{\psi}' - \underline{\psi}\underline{1}')\Omega^{-1}}{(\underline{\psi}'\Omega^{-1}\underline{\psi})(\underline{1}'\Omega^{-1}\underline{1}) - (\underline{\psi}'\Omega^{-1}\underline{1})^2}\underline{Y}_{[n]} \tag{48}$$

$$\tilde{\tilde{\sigma}}_2 = \frac{\underline{1}'\Omega^{-1}(\underline{1}\underline{\psi}' - \underline{\psi}\underline{1}')\Omega^{-1}}{(\underline{\psi}'\Omega^{-1}\underline{\psi})(\underline{1}'\Omega^{-1}\underline{1}) - (\underline{\psi}'\Omega^{-1}\underline{1})^2}\underline{Y}_{[n]} \tag{49}$$

$$V(\tilde{\tilde{\mu}}_2) = \frac{(\underline{\psi}'\Omega^{-1}\underline{\psi})\sigma_2^2}{(\underline{\psi}'\Omega^{-1}\underline{\psi})(\underline{1}'\Omega^{-1}\underline{1}) - (\underline{\psi}'\Omega^{-1}\underline{1})^2} \tag{50}$$

and

$$V(\tilde{\tilde{\sigma}}_2) = \frac{(\underline{1}'\Omega^{-1}\underline{1})\sigma_2^2}{(\underline{\psi}'\Omega^{-1}\underline{\psi})(\underline{1}'\Omega^{-1}\underline{1}) - (\underline{\psi}'\Omega^{-1}\underline{1})^2}. \tag{51}$$

4. COMPUTATION SECTION

We have evaluated the coefficients of the BLUE μ_2^* (given in (26)) of μ_2 and evaluated the coefficients of the BLUE σ_2^* (given in (27)) of σ_2 using ranked set sampling for different values of α and sample size n, are given in Table 1. Also we have evaluated the variance of the estimators μ_2^* (given in (28)), $\tilde{\mu}_2$ (given in (38)) and $\tilde{\tilde{\mu}}_2$ (given in (50)) and evaluated the relative efficiency of μ_2^* relative to $\tilde{\mu}_2$ (RE_1), relative efficiency of μ_2^* relative to $\tilde{\tilde{\mu}}_2$ (RE_2) for different values of α and n are presented in Table 2.

We have also evaluated variance of the estimators μ_2^* (given in (29)), $\tilde{\sigma}_2$ (given in (39)) and $\tilde{\tilde{\sigma}}_2$ (given in (51)) and evaluated the relative efficiency of σ_2^* relative to $\tilde{\sigma}_2$ (RE_3), relative efficiency of σ_2^* relative to $\tilde{\tilde{\sigma}}_2$ (RE_4) for different values of α and n are presented in Table 3.

Also to find the asymptotic variances of the MLE's of the parameters μ_2 and σ_2, we have derived the following.

$$\frac{d\log f(x,y)}{d\mu_2} = \frac{1}{\sigma_2}\left[1+\frac{2\alpha\exp\{-\frac{y-\mu_2}{\sigma_2}\}(2\exp\{\frac{x-\mu_1}{\sigma_1}\}-1)}{1+\alpha(2\exp\{-\frac{x-\mu_1}{\sigma_1}\}-1)(2\exp\{-\frac{y-\mu_2}{\sigma_2}\}-1)}\right], \quad (52)$$

where $f(x, y)$ is defined in (12). Now the fisher information measure about the parameter μ_2 involved in (12) based on a sample of size n arising from (12) is given by

$$I_{\mu_2}(\alpha) = nE\left[\frac{d\log f}{d\mu_2}\right]^2.$$

On simplification we get

$$I_{\mu_2}(\alpha) = \frac{n}{\sigma_2^2}\left[1+4\alpha^2\int_0^\infty\int_0^\infty \frac{e^{-3v}(2e^{-u}-1)^2 e^{-u}}{1+\alpha(2e^{-u}-1)(2e^{-v}-1)}dudv\right]. \quad (53)$$

Again we have

$$\frac{d\log f(x,y)}{d\sigma_2} = \frac{1}{\sigma_2}\left[-1+(\frac{y-\mu_2}{\sigma_2})+\frac{2\alpha(\frac{y-\mu_2}{\sigma_2})\exp\{-\frac{y-\mu_2}{\sigma_2}\}(2\exp\{-\frac{x-\mu_1}{\sigma_1}\}-1)}{1+\alpha(2\exp\{-\frac{x-\mu_1}{\sigma_1}\}-1)(2\exp\{-\frac{y-\mu_2}{\sigma_2}\}-1)}\right] \quad (54)$$

Table 1: Coefficients of the BLUE's μ_2^* (given in (26)) and σ_2^* (given in (27)) using RSS

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	-0.75	2	-3.5	4.5								
		3	-2.161	-0.012	3.173							
		4	-1.553	-0.576	0.644	2.485						
		5	-1.209	-0.652	-0.011	0.816	2.057					
		6	-0.989	-0.629	-0.233	0.238	0.853	1.761				
		7	-0.837	-0.584	-0.315	-0.009	0.360	0.842	1.543			
		8	-0.724	-0.538	-0.342	-0.126	0.121	0.415	0.816	1.377		
		9	-0.639	-0.495	-0.346	-0.186	-0.0079	0.199	0.453	0.781	1.241	
		10	-0.571	-0.457	-0.34	-0.216	-0.081	0.071	0.249	0.466	0.746	1.132

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	-0.75	2	4	-4								
		3	2.498	0.337	-2.835							
		4	1.809	0.821	-0.401	-2.229						
		5	1.415	0.850	0.205	-0.621	-1.849					
		6	1.162	0.795	0.395	-0.077	-0.689	-1.587				
		7	0.985	0.728	0.454	0.147	-0.222	-0.701	-1.392			
		8	0.855	0.664	0.465	0.248	-0.0008	-0.296	-0.691	-1.244		
		9	0.756	0.608	0.456	0.294	0.115	-0.093	-0.344	-0.669	-1.123	
		10	0.677	0.559	0.440	0.314	0.177	0.025	-0.153	-0.369	-0.645	-1.024

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	-0.5	2	-5.5	6.5								
		3	-3.497	-0.005	4.503							
		4	-2.555	-0.920	1.006	3.469						
		5	-2.011	-1.066	-0.004	1.251	2.831					
		6	-1.657	-1.042	-0.371	0.389	1.285	2.395				
		7	-1.408	-0.976	-0.514	-0.004	0.574	1.251	2.077			
		8	-1.225	-0.905	-0.566	-0.200	0.204	0.661	1.194	1.835		
		9	-1.083	-0.837	-0.578	-0.302	-0.003	0.327	0.700	1.132	1.644	
		10	-0.971	-0.775	-0.571	-0.356	-0.125	0.125	0.401	0.713	1.070	1.490

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	-0.5	2	6	-6								
		3	3.832	0.335	-4.169							
		4	2.808	1.168	-0.759	-3.217						
		5	2.213	1.265	0.202	-1.052	-2.628					
		6	1.826	1.208	0.535	-0.225	-1.120	-2.225				
		7	1.554	1.119	0.655	0.145	-0.433	-1.109	-1.931			
		8	1.352	1.030	0.690	0.323	-0.081	-0.538	-1.069	-1.707		
		9	1.197	0.948	0.688	0.412	0.113	-0.218	-0.590	-1.020	-1.531	
		10	1.073	0.876	0.671	0.455	0.224	-0.027	-0.303	-0.613	-0.969	-1.387

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	-0.01	2	-298.90	299.90								
		3	-199.5	0	200.5							
		4	-149.5	-49.949	50.049	150.45						
		5	-119.57	-60.067	0.0001	60.268	120.38					
		6	-99.763	-59.887	-19.897	19.925	60.144	100.48				
		7	-85.393	-57	-28.535	0	28.607	57.285	86.036			
		8	-74.624	-53.389	-32.107	-10.778	10.791	32.215	53.687	75.206		
		9	-66.401	-49.85	-33.266	-16.65	0	16.683	33.399	50.15	66.934	
		10	-59.822	-46.553	-33.261	-19.946	-6.606	6.612	20	33.413	46.85	-60.313

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	-0.01	2	299.40	-299.40								
		3	199.83	0.333	-200.16							
		4	149.80	50.199	-49.799	-150.20						
		5	119.77	60.267	0.200	-60.068	-120.18					
		6	99.93	60.053	20.064	-19.759	-59.977	-100.31				
		7	85.536	57.143	28.678	0.143	-28.464	-57.143	-85.893			
		8	74.749	53.514	32.232	10.903	-10.666	-32.09	-53.562	-75.081		
		9	66.512	49.961	33.377	16.761	0.111	-16.572	-33.288	-50.039	-66.823	
		10	59.922	46.653	33.361	20.046	6.706	-6.512	-19.900	-33.313	-46.75	-60.213

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	0.01	2	299.90	-298.90								
		3	200.5	0	-199.5							
		4	150.45	50.049	-49.949	-149.55						
		5	120.37	60.268	0.0001	-60.067	-119.57					
		6	100.47	60.144	19.925	-19.897	-59.887	-99.763				
		7	86.036	57.285	28.607	0	-28.535	-57	-85.383			
		8	75.206	53.686	32.214	10.79	-10.778	-32.107	-53.388	-74.624		
		9	66.934	50.15	33.399	16.683	0	-16.65	-33.266	-49.85	-66.401	
		10	60.313	46.85	33.413	20	6.612	-6.606	-19.946	-33.261	-46.583	-59.822

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	0.01	2	-299.40	299.40								
		3	-200.16	0.333	199.83							
		4	-150.20	-49.799	50.199	149.80						
		5	-120.17	-60.068	0.200	60.267	119.97					
		6	-100.31	-59.977	-19.759	20.064	60.053	99.93				
		7	-85.893	-57.143	-28.464	0.143	28.678	57.143	85.536			
		8	-75.081	-53.561	-32.089	-10.665	10.903	32.232	53.513	74.749		
		9	-66.823	-50.039	-33.288	-16.572	0.111	16.761	33.377	49.961	66.512	
		10	-60.261	-46.75	-33.313	-19.900	-6.512	6.623	20.046	33.361	46.653	59.922

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	0.5	2	6.5	-5.5								
		3	4.503	-0.005	-3.497							
		4	3.469	1.006	-0.920	-2.555						
		5	2.831	1.251	-0.0048	-1.066	-2.011					
		6	2.395	1.285	0.389	-0.371	-1.042	-1.657				
		7	2.077	1.251	0.574	-0.004	-0.514	-0.976	-1.408			
		8	1.835	1.194	0.661	0.204	-0.200	-0.566	-0.905	-1.225		
		9	1.644	1.132	0.700	0.327	-0.0034	-0.302	-0.578	-0.837	-1.083	
		10	1.406	1.048	0.732	0.444	0.176	-0.078	-0.323	-0.563	-0.801	-1.042

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	0.5	2	-6	6								
		3	-4.168	0.335	3.832							
		4	-3.217	-0.759	1.168	2.808						
		5	-2.628	-1.052	0.202	1.265	2.213					
		6	-2.225	-1.112	-0.225	0.535	1.208	1.826				
		7	-1.931	-1.109	-0.433	0.145	0.655	1.119	1.554			
		8	-1.707	-1.069	-0.538	-0.081	0.323	0.690	1.030	1.352		
		9	-1.530	-1.02	-0.59	-0.218	0.11	0.412	0.688	0.948	1.197	
		10	-1.312	-0.950	-0.630	-0.341	-0.072	0.182	0.425	0.663	0.899	1.1336

Estimator	α	n	Coefficients									
			a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
μ^*	0.75	2	4.5	-3.5								
		3	3.173	-0.012	-2.161							
		4	2.485	0.644	-0.576	-1.553						
		5	2.057	0.816	-0.011	-0.652	-1.209					
		6	1.761	0.853	0.238	-0.233	-0.629	-0.989				
		7	1.543	0.842	0.36	-0.0093	-0.315	-0.584	-0.837			
		8	1.375	0.815	0.422	0.120	-0.127	-0.342	-0.538	-0.724		
		9	1.241	0.781	0.453	0.199	-0.0079	-0.186	-0.346	-0.495	-0.639	
		10	1.132	0.746	0.466	0.249	0.071	-0.081	-0.216	-0.340	-0.457	-0.571

Estimator	α	n	Coefficients									
			b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
σ^*	0.75	2	-4	4								
		3	-2.835	0.337	2.498							
		4	-2.229	-0.401	0.821	1.809						
		5	-1.849	-0.621	0.205	0.850	1.415					
		6	-1.587	-0.689	-0.077	0.395	0.795	1.162				
		7	-1.392	-0.701	-0.222	0.147	0.454	0.728	0.985			
		8	-1.242	-0.690	-0.300	-0.0002	0.248	0.465	0.664	0.855		
		9	-1.123	-0.669	-0.344	-0.093	0.115	0.294	0.456	0.608	0.756	
		10	-1.024	-0.645	-0.369	-0.153	0.025	0.177	0.314	0.440	0.559	0.677

Table 2: $V_1 = Var(\mu_2^*)$, $V_2 = Var(\tilde{\mu}_2)$, $V_3 = Var(\tilde{\tilde{\mu}}_2)$ **and** $RE_1 = (V2/V1)$, $RE_2 = (V3/V1)$ **for different values of** α **and n.**

$\alpha = -0.75$						$\alpha = -0.5$					
n	V_1	V_2	V_3	RE_1	RE_2	*n*	V_1	V_2	V_3	RE_1	RE_2
2	30.992	28.531	30.992	0.921	1	2	71.002	68.519	71.002	0.965	1
3	13.204	12.001	9.307	0.909	0.705	3	30.992	29.855	22.291	0.963	0.719
4	8.041	7.465	5.271	0.928	0.706	4	19.168	18.665	13.105	0.974	0.685
5	5.689	5.501	3.746	0.967	0.658	5	13.706	13.580	9.585	0.991	0.699
6	4.369	4.449	2.974	1.018	0.680	6	10.609	10.734	7.786	1.012	0.734
7	3.533	3.809	2.516	1.078	0.712	7	8.630	8.934	6.709	1.035	0.777
8	2.962	3.401	2.216	1.148	0.748	8	7.262	7.699	5.998	1.060	0.826
9	2.542	3.092	2.005	1.216	0.789	9	6.262	6.804	5.495	1.086	0.878
10	2.226	2.875	1.849	1.292	0.831	10	5.501	6.126	5.122	1.114	0.931
$\alpha = -0.01$						$\alpha = 0.01$					
n	V_1	V_2	V_3	RE_1	RE_2	*n*	V_1	V_2	V_3	RE_1	RE_2
2	179281	179278	179281	0.999	1	2	179281	179278	179281	0.999	0.751
3	79998.7	79997.59	59949.03	0.999	0.749	3	79998.7	79997.998	60040.02	0.999	0.751
4	49999.139	49998.59	36980.78	0.999	0.739	4	49999.14	49998.59	37091.88	0.999	0.742
5	36028.102	36027.865	28124.43	0.999	0.781	5	36028.102	36027.89	28237.04	0.999	0.784
6	28044.301	28044.257	23482.27	0.999	0.837	6	28044.301	28044.25	23594.31	0.999	0.841
7	22856.681	22856.77	20684.69	1.0001	0.905	7	22856.68	22830.557	20795.79	0.999	0.909
8	19257.78	19257.905	18825.615	1.0001	0.978	8	19257.72	19257.91	18935.92	1.0001	0.983
9	16666.24	16666.502	17522.80	1.0001	1.0514	9	16666.24	16666.49	17632.18	1.0001	1.058
10	14685.85	14686.16	16550.52	1.0001	1.127	10	14685.85	14694.342	16659.18	1.0001	1.134

$\alpha = 0.5$						$\alpha = 0.75$					
n	V_1	V_2	V_3	RE_1	RE_2	n	V_1	V_2	V_3	RE_1	RE_2
2	71.002	68.519	71.002	0.965	1	2	30.992	28.531	30.992	0.921	1
3	30.992	29.875	24.197	0.964	0.781	3	13.204	12.072	10.499	0.914	0.795
4	19.168	18.523	15.177	0.966	0.792	4	8.041	7.344	6.529	0.913	0.812
5	13.706	13.303	11.647	0.971	0.849	5	5.689	6.203	4.965	1.090	0.873
6	10.609	10.354	9.811	0.976	0.925	6	4.369	4.021	4.146	0.920	0.948
7	8.630	8.474	8.696	0.982	1.008	7	3.533	3.272	3.646	0.926	1.032
8	7.262	7.178	7.952	0.988	1.095	8	2.959	2.798	3.309	0.946	1.118
9	6.262	6.232	7.420	0.995	1.185	9	2.542	2.389	2.697	0.939	1.060
10	5.501	5.513	7.021	1.0022	1.345	10	2.226	2.092	2.889	0.939	1.298

Table 3: $V_4 = Var(\sigma_2^*)$, $V_5 = Var(\tilde{\sigma}_2)$, $V_6 = Var(\tilde{\tilde{\sigma}}_2)$ **and** $RE_3 = (V_5/V_4)$, $RE_4 = (V_6/V_4)$ **for different values of** α **and** n

α = — 0.75						α = — 0.5					
n	V_4	V_5	V_6	RE_3	RE_4	n	V_4	V_5	V_6	RE_3	RE_4
2	31.000	28.5	31.00	0.919	1	2	71.006	68.506	71.005	0.965	0.999
3	13.216	11.839	9.940	0.896	0.751	3	30.998	29.755	23.266	0.959	0.750
4	8.053	7.229	5.955	0.897	0.739	4	19.173	18.518	14.173	0.966	0.739
5	5.699	5.216	4.422	0.915	0.776	5	13.711	13.404	10.656	0.978	0.777
6	4.379	4.129	3.635	0.943	0.830	6	10.614	10.537	8.844	0.993	0.839
7	3.542	3.464	3.163	0.978	0.893	7	8.634	8.722	7.752	1.010	0.897
8	2.969	3.036	2.849	1.025	0.959	8	7.266	7.477	7.027	1.029	0.967
9	2.550	2.709	2.628	1.062	1.031	9	6.265	6.572	6.512	1.049	1.039
10	2.234	2.478	2.463	1.059	1.103	10	5.504	5.887	6.129	1.069	1.114

α = — 0.01						α = 0.01					
n	V_4	V_5	V_6	RE_3	RE_4	n	V_4	V_5	V_6	RE_3	RE_4
2	179281	179278	179281	0.999	1	2	179281	179278	179281	0.999	1
3	79998.7	79997.59	59999.02	0.999	0.749	3	79998.7	79998.00	59999.02	0.999	0.749
4	49999.14	49998.59	37036.33	0.999	0.741	4	49999.14	49998.597	37036.33	0.999	0.740
5	36028.102	36027.86	28180.74	0.999	0.782	5	36028.102	36027.898	28180.739	0.999	0.782
6	28044.301	28044.253	23538.29	0.999	0.839	6	28044.301	28044.25	23538.29	0.999	0.939
7	22856.681	22856.77	20740.24	1.0001	0.907	7	22856.68	22830.59	20740.24	1.130	0.974
8	19257.776	19257.901	18801.699	1.0001	0.976	8	19257.72	19257.91	18880.83	1.0001	0.980
9	16666.24	16666.49	17577.493	1.0001	1.055	9	16666.24	16666.49	17577.49	1.0001	1.055
10	14685.848	14686.16	16604.85	1.0001	1.131	10	14685.85	14694.342	16604.85	1.0001	1.131
$\alpha = 0.5$						$\alpha = 0.75$					

n	V_4	V_5	V_6	RE_3	RE_4	n	V_4	V_5	V_6	RE_3	RE_4
2	71.006	68.506	71.006	0.965	1	2	31.00	28.5	31.00	0.919	1
3	30.998	29.939	23.235	0.966	0.749	3	13.216	12.15	9.893	0.9193	0.749
4	19.173	18.624	14.123	0.971	0.706	4	8.053	7.477	5.880	0.928	0.730
5	13.711	13.426	10.593	0.979	0.773	5	5.699	6.483	4.328	1.137	0.759
6	10.614	10.490	8.772	0.989	0.826	6	4.379	4.204	3.528	0.9600	0.806
7	8.634	8.621	7.674	0.998	0.889	7	3.542	3.469	3.045	0.979	0.859
8	7.266	7.332	6.944	1.0091	0.956	8	2.968	3.008	2.725	1.013	0.814
9	6.265	6.39	6.425	1.019	1.026	9	2.550	2.605	2.078	1.022	0.815
10	5.504	5.677	6.038	1.0314	1.097	10	2.234	2.313	2.327	1.035	1.046

Now fisher information measure about the parameter σ_2 involved in (12) based on a sample of size n arising from (12) is sample is given by

$$I_{\sigma_2}(\alpha) = nE\left[\frac{d}{d\sigma_2}\log f(x,y)\right]^2 \tag{55}$$

On simplification we get

$$I_{\sigma_2}(\alpha) = \frac{n}{\sigma_2^2}\left\{1 + 4\alpha^2\int_0^\infty\int_0^\infty \frac{v^2 e^{-3v}(2e^{-u}-1)e^{-u}}{\{1+\alpha\,(2e^{-u}-1)\,(2e^{-v}-1)\}}\,dudv\right\}. \tag{56}$$

Now the asymptotic variance of the MLE $\hat{\mu}_2$ of μ_2 involved in MTBED based on a sample of size n is given by $Var(\hat{\mu}_2) = \dfrac{1}{I_{\mu_2}(\alpha)}$ and the asymptotic variance of the MLE $\hat{\sigma}_2$ of σ_2 involved in MTBED based on a sample of size n is given by $Var(\hat{\sigma}_2) = \dfrac{1}{I_{\sigma_2}(\alpha)}$. We have evaluated the asymptotic variances of the MLE's of μ_2 and σ_2 for different values of α and sample size *n* and are presented respectively in Tables 4 and 5. Also we have drawn the graphs of asymptotic variance of the MLE of μ_2 and σ_2 involved in (12) for different values of α against different sample sizes and are presented in Figures 1 and 2.

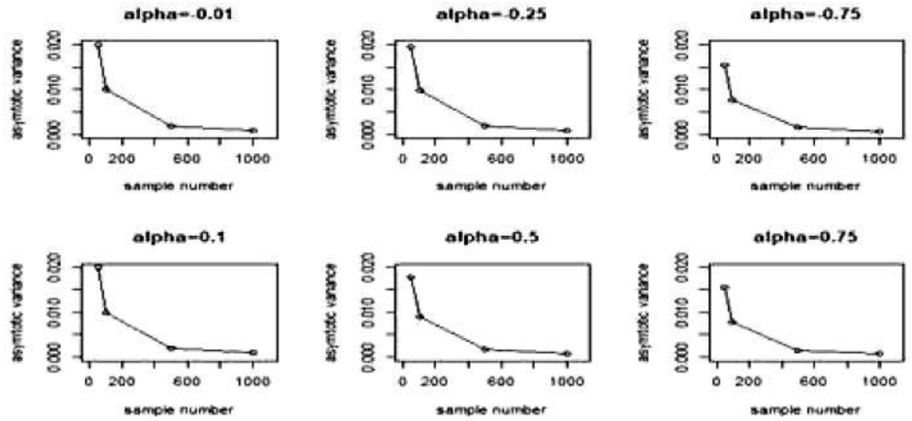

Figure 1: Graph of asymptotic variance of MLE of μ_2 for different values of α and sample size *n*

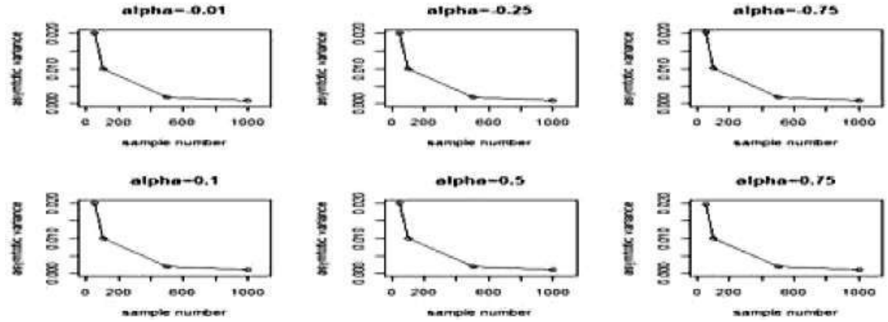

Figure 2: Graph of asymptotic variance of MLE of σ_2 for different values of α and sample size *n*.

5. CONCLUSION

From Tables 2 and 3, it may be noted that for sample sizes (n), greater than or equal to seven most of the cases the performances of the estimators μ_2^* and σ_2^* using ranked set sampling is much better than the estimators using concomitants of order statistics and using modified ranked set sampling for different values of α. But when the sample sizes is less than seven the estimators using modified ranked set sampling is much better than that of the estimators using ranked set sampling for different values of α.

Table 4: Asymptotic variance of MLE of μ_2 for different values of α and n.

	$\sigma_2^{-2} Var(\hat{\mu}_2) = \dfrac{\sigma_2^{-2}}{I_{\mu_2}(\alpha)}$			
α	**n = 50**	**n = 100**	**n = 500**	**n = 1000**
-0.75	0.01542	0.00771	0.00154	0.00077
-0.5	0.01788	0.00894	0.00179	0.00089
-0.25	0.01945	0.00973	0.00195	0.00097
-0.1	0.01991	0.00996	0.00199	0.00100
-0.01	0.02000	0.01000	0.00200	0.00100
0.01	0.02000	0.01000	0.00200	0.00100
0.1	0.01991	0.00996	0.00199	0.00100
0.25	0.01945	0.00973	0.00195	0.00097
0.5	0.01788	0.00894	0.00179	0.00089
0.75	0.01542	0.00771	0.00154	0.00077

Table 5: Asymptotic variance of MLE of σ_2 for different values of α and n.

	$\sigma_2^{-2}Var(\hat{\sigma}_2)=\dfrac{\sigma_2^{-2}}{I_{\mu 2}(\alpha)}$			
α	**n = 50**	**n = 100**	**n = 500**	**n = 1000**
-0.75	0.02015	0.01008	0.00202	0.00101
-0.5	0.02004	0.01002	0.00200	0.00100
-0.25	0.02000	0.01000	0.00200	0.00100
-0.1	0.02000	0.01000	0.00200	0.00100
-0.01	0.02000	0.01000	0.00200	0.00100
0.01	0.02000	0.01000	0.00200	0.00100
0.1	0.02000	0.01000	0.00200	0.00100
0.25	0.02000	0.01000	0.00200	0.00100
0.5	0.01996	0.00998	0.00200	0.00100
0.75	0.01985	0.00992	0.00198	0.00099

6. Acknowledgements

The authors wish to thank the referee for giving valuable comments.

References

1. Abu-Dayyeh, H., and Huttlak, H.A (1996): Using ranked set sampling for hypothesis tests on the scale parameter of the exponential and uniform distributions, *Pakistan J. Stat*., 12, 131-138.
2. Barnett, V. and Moore, K. (1997): Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering, *J. Appl. Stat.*, 24(6), 697-710.
3. Chacko, M. and Thomas, P.Y. (2008): Estimation of parameters of a parameter of Morgenstern type bivariate exponential distribution by ranked set sampling, *Ann Inst Stat Math*., 60(2), 301-318.
4. Chen, Z., Bai, Z. and Sinha, B.K. (2004): Lecture notes in Statistics (Vol. 176), Springer-Verlag, New York.
5. Conway, D.A. (1983): Ferlie-Gumbel-Morgenstern distributions. In: Kotz S, Johnson N.L. Encyclopaedia of Statistical Sciences Vol. 3, Wiley, New York.

6. Lam, K., Sinha, B.K., and Wu, Z. (1994): Estimation of parameters in a two parameter exponential distribution using ranked set sample, *Ann Inst Stat Math*., 46(4), 723-736.
7. Lam, K., Sinha, B.K., and Wu, Z. (1995): Estimation of location and scale parameters of a logistic distribution using a ranked set sample, *Statistical Theory and Applications* Papers in Honor of Herbert A. David. (Eds. H.N. Nagaraja, P.K. Sen and D.F. Morrison), Springer, New York.
8. MacIntyre, G.A. (1952): A method for unbiased selective sampling using ranked sets, *Aust. J. Agric. Res.*, 3(4), 385-390.
9. Muttlak, H.A, Abu-Dayyesh, W.A, Saleh, M.F and Al-Sawi, E. (2010): Estimating P(Y<X) using ranked set sampling in case of the exponential distribution, *Commun. Stat. Theo. Meth*., 39(10), 1855-1868.
10. Samuh, M.H and Qtait, A. (2015): Estimation for the parameters of the exponentiated exponential distribution using a median ranked set sampling, *J. Appl. Statist. Meth*., 14(1), 215-237.
11. Scaria, J. and Nair, N.U. (1999): On concomitants of order statistics from Morgenstern family, *Biometrical Journal*, 41(4), 483-489.
12. Sengupta, S. and Mukhuti, S. (2008): Unbiased estimation of $P(X > Y)$ for exponential populations using order statistics with application in ranked set sampling, *Commun. Stat. Theo. Meth*., 37 (6), 898-916.
13. Stokes, S.L. (1977): Ranked set sampling with concomitant variables, *Commun. Stat. Theo. Meth*., 6, 1207-1211.
14. Stokes, S.L. (1995): Parametric ranked set sampling, *Ann Inst Stat Math*., 47(3), 465-482.
15. Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000): *Continuous Multivariate Distributions*. Vol.1 (2nd ed.), Wiley, New York.
16. Yang, S.S. (1977): General distribution theory of the concomitants of order statistics, *Ann. Stat.*, 5, 996-1002.

IAPQR Transactions
Vol. 47, No. 1-2, 2022–2023

WEIGHTED ALTERNATED CHARTING STATISTIC CONTROL CHART TO MONITOR THE MEAN VECTOR

M.P. GADRE[1]
Shriram Institute of Information Technology, Solapur, India

***Abstract*:** In this article, modification of the 'Alternated Charting Statistic' (*ACS*) scheme with the traditional Hotelling χ^2 chart to control the mean vector of a bivariate normal process is proposed. With the bivariate *ACS* scheme in use, the two quality characteristics (*X*, *Y*) are controlled in an alternating fashion. The proposed control chart is called as 'Weighed Alternated Charting Statistic' (*WACS*) control chart to monitor the mean vector by using variable inspection. In *WACS* chart, which one of the two quality characteristic is to be inspected, is explained in detail. While defining the rule of the status of the process, the weights of each of the two quality characteristics are considered. It is numerically illustrated that the *WACS* chart performs better as compared to the χ^2chart and *ACS* chart.

***Keywords and Phrases*:** Average Run Length, Average Time to Signal, Alternated Chart Statistic.

1. Introduction

Control charts are monitoring tools specially designed to detect assignable causes. $\overline{X}$ and T^2 charts are the standard tools to control the mean and the mean vector. Wu and Spedding (2000) proposed Synthetic chart to detect shift in the process mean. For multivariate processes, Hotelling (1947) introduced χ^2 and T^2 charts to monitor the mean vector. Ghute and Shirke (2008) proposed the 'Synthetic' control chart to monitor a Mean vector' (*Syn-M*) chart. For both the above mentioned charts, 'Average Run Length' (*ARL*) criterion is used.

---

1. Corresponding author email: drmukund.gadre@gmail.com
DOI-https://doi.org/10.32381/IAPQRT.2023.47.01-02.3

For univariate case, Wu et al. (2001) introduced the term ‘Average Time to Signal’ (*ATS*) and used *ATS* criterion in place of *ARL* criterion. Gadre and Kakade (2016, 2019) developed some run length based control charts to monitor mean vector of a multivariate process by using *ATS* criterion. Woodall and Ncube (1985) proposed ‘Multivariate Cumulative SUM’ (*MCUSUM*) chart and compared with T^2 chart. Hamed et al., (2016) discussed a procedure of *MCUSUM* chart to monitor mean vector. They also discussed the related applications. Xia et al. (2018) developed an effective chart to monitor the mean vector. This proposed *D-MCUSUM* control chart performs better as compared with *MCUSUM* control chart. Reynold and Chou (2011) proposed multivariate control chart to monitor mean vector and covariance matrix with variable sampling interval. Kang et al. (2007) introduced a Shewhart type control chart for coefficient of variation (SH CV chart). This control chart is widely applicable in various fields like clinical chemistry. In such situations, neither the process mean the process variability is constant, but the coefficient of variation is constant.

In multivariate applications, more expensive and time-consuming measurements are observed. To reduce the time consumption and expenses for measurement, Leoni and Costa (2017) introduced the ‘Alternated Charting Statistic’ (*ACS*) control chart to monitor the mean vector of bivariate and trivariate processes and verified that the chart is easier to operate and faster than the classical HotellingT^2 chart in signaling changes in the mean vector.

In a multi-attribute situation, Melo et al. (2017) developed an attribute control chart to monitor a mean vector of a bivariate process. The chart is named as Max *D* chart. This chart is a competitive alternative to the Hotelling T^2 chart.

Note: It is to be noted that, though some of the authors [like Ghute and Shirke (2008), Gadre and Kakade (2016), Leoni and Costa (2017) and Gadre and Nisha (2021)] mentioned that the proposed charts perform better as compared to the Hotelling T^2 chart, the comparison was carried out between the proposed charts and Hotelling χ^2 chart, but not with Hotelling T^2 chart (Refer Montgomery D. C. (2008)).

For bivariate processes, while defining the rule of the status of the process, the weights of each of the two quality characteristics are considered. The proposed control chart is called as ‘Weighted Alternated Charting Statistic’ (*WACS*) control chart to monitor the mean vector by using variable inspection.

Remainder of the paper is organized as follows. In Section 2, we cover a brief review of the multivariate control charts like Hotelling χ^2 and *ACS* charts. Section 3 covers some basic notations, terms and operation of the proposed chart. Derivation of the *ATS* expression of the *WACS* chart is given in the next section. In Section-5, for some numerical illustrations, *ATS* performances of the proposed chart with the Hotelling χ^2 and *ACS* charts are studied. A real life example is included and the *ATS* performances of the proposed chart with the related control charts are studied in the same section. Concluding remarks are given in Section 6.

2. A BRIEF REVIEW OF THE MULTIVARIATE CONTROL CHARTS

In the following, we brief some related multivariate control charts to monitor mean vector.

2.1 Hotelling χ^2 chart (Refer Montgomery D.C. (2008))

For bivariate processes, suppose a sample of size n is taken from $N_2\,(\underline{\mu}, \Sigma)$ distribution. In bivariate processes in order to monitor process mean vector, Hotelling χ^2 chart can be used. The T^2statistic is given by,

$$\chi^2 = n\left(\underline{\overline{X}} - \underline{\mu}_0\right)^{/}\Sigma_0^{-1}\left(\underline{\overline{X}} - \underline{\mu}_0\right) \tag{1}$$

where, $\underline{\overline{X}}$ and $\underline{\mu}_0 = (\mu_{0x}, \mu_{0y})^{/}$ are the sample mean vector and in-control mean vector. Also Σ_0 is the in control covariance matrix of a bivariate process respectively. Let (X, Y) be the two quality characteristics. Thus, for a bivariate process, the χ^2statistic is given by,

$$\chi^2 = n\begin{bmatrix}\overline{X} - \mu_{0x}\\ \overline{Y} - \mu_{0y}\end{bmatrix}^{/}\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}\overline{X} - \mu_{0x}\\ \overline{Y} - \mu_{0y}\end{bmatrix} \tag{2}$$

Here ρ is the correlation coefficient between the quality characteristic X and Y. When the process is in control, the above χ^2 statistic is distributed as a chi-square random variable with two degrees of freedom (χ_2^2).

The quantity 'k' is the only design parameter of the χ^2chart. 'Average Run Length' (*ARL*) is the average number of samples inspected by the time the process has gone out of control. The *ARL* criterion is

$$\left.\begin{array}{ll}\text{Minimize} & ARL_1\\ \text{subject to} & ARL_0 \geq \tau^{/}\end{array}\right\} \tag{3}$$

where, $\tau^{/}$ is the minimum required value of ARL_0. The *UCL* of the chart is adjusted to obtain the desired in control *ARL* (ARL_0). i.e. $UCL=\chi^2_{2,\alpha}$, where α is a reciprocal of ARL_0.

Once an assignable cause occurs, the mean vector changes from $\underline{\mu}_0$ *to* $\underline{\mu}_1$ = $(\mu_{0x}+\delta_{1x}\sigma_X,\ \mu_{0y}+\delta_{1y}\sigma_Y)$, where $\underline{\boldsymbol{\delta}}_1 = (\delta_{1x},\ \delta_{1y})' = \left(\frac{\mu_{1x} - \mu_{0x}}{\sigma_{\bar{x}}},\ \frac{\mu_{1y} - \mu_{0y}}{\sigma_{\bar{y}}}\right)'$ as the shift in the standardized mean vector due to the process going out of control. δ_{1x} and δ_{1y} are the input parameters, which define what is the acceptable tolerance on the process mean vector, variation. σ_X, σ_Y are the standard deviations of the sample means of *X* and *Y* respectively and the distribution of χ^2 changes to non-central chi-square distribution $\chi_2^2(\lambda)$ of two degrees of freedom with the non-centrality parameter λ, where

$$\lambda = \underline{\delta}_1' \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}^{-1} \underline{\delta}_1, \tag{4}$$

where, $\underline{\delta}_1 = (\delta_{1x}, \delta_{1y})'$. The out of control *ARL* (i.e. ARL_1) is given by,

$$ARL_1 = (1 - Pr\,[\,\chi_2^2(\lambda) \le UCL])^{-1} \tag{5}$$

Wu et al., (2001) introduced the term ‘Average Time to Signal’ (*ATS*) and used the *ATS* criterion. For fixed sample size *n*, *ATS* is the average number of units inspected by the time the process has gone out of control’. Under the assumption that the number of units produced are per unit of time, *ATS* = *n*(*ARL*). Further, ATS_0 and ATS_1 are the respective *ATSs* when the process is running smoothly and when the process has gone out of control. The *ATS* criterion is,

$$\left.\begin{array}{ll} \text{Minimize} & ATS_1 \\ \text{subject to} & ATS_0 \ge \tau \end{array}\right\} \tag{6}$$

where, τ is the minimum required value of ATS_0.

2.2. ‘Alternated Charting Statistic’ (*ACS*) Chart

This charting method was proposed by Leoni and Costa (2017) for dealing with bivariate (and trivariate) processes. This is an alternative to the Hotelling χ^2 chart, in which the Shewhart $\overline{X}$ chart for sample means is combined with an ‘Alternating the Charting Statistic’ strategy. It is assumed that the quality characteristics follow a normal distribution with known parameters, when it is in control. For example, in a bivariate distribution case, a sample of size *2n* is taken from $N_2(\underline{\mu}, \Sigma)$ distribution. Stepwise procedure of the operation of *ACS* chart is as follows.

Step-1: Initialize i to 1.

Step-2: Take i^{th} sample of size $2n$ from $N_2(\underline{\mu}, \Sigma)$ distribution. Measure X observations only in the i^{th} sample and compute the sample mean $\overline{X}$. If $\overline{X}$ is within the control limits for X sample, move to the next step; otherwise go to **Step-4**.

Step-3: Let $j = i+1$. Measure Y observations only in the j^{th} sample, and compute $\overline{Y}$. If the j^{th} sample mean $\overline{Y}$ is within the control limits for Y sample, add 2 to i (i.e. $i = i+2$) and then go back to **Step-2**; otherwise go to **Step-4**.

Step-4: The process has gone out of control. Identify the assignable causes and take a corrective action before restarting the process. Then go back to **Step-1**.

Consequently, the entire sequence of X observations are independent and as well for Y observations. For no correlation or small to moderate correlation between the two quality characteristics, Leoni and Costa (2017) have illustrated that the *ACS* chart is easier to operate and it is faster than Hotelling χ^2 chart in signaling changes in the mean vector of a bivariate process.

3. Weighted Alternated Charting Statistic Control Chart

For bivariate processes, let X and Y be the two continuous random variables having $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ respectively. Let ρ_{xy} be the correlation coefficient between X and Y.

3.1 Notations

Following are some basic notations used in bivariate WACS chart.

1. $\underline{\mu_0} = (\mu_{0x}, \mu_{0y})'$: In control values of process mean vector
2. $\underline{\mu_1} = (\mu_{1x}, \mu_{1y})'$: Out of control values of process mean vector
3. $\underline{\sigma} = (\sigma_x, \sigma_y)$: The process variability
4. $\underline{\delta}' = (\delta_x, \delta_y) = \left(\dfrac{\mu_x - \mu_{0x}}{\sigma_{\bar{x}}}, \dfrac{\mu_y - \mu_{0y}}{\sigma_{\bar{y}}} \right)$, where $\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{2n}}, \sigma_{\bar{Y}} = \frac{\sigma_y}{\sqrt{2n}}$

   are the process variability of $\overline{X}$ and $\overline{Y}$ respectively and it is a shift in the standardized mean vector

5. $\underline{\delta}_1^{/} = (\delta_{1x}, \ \delta_{1y}) = \left(\dfrac{\mu_{1x} - \mu_{0x}}{\sigma_{\bar{x}}}, \dfrac{\mu_{1y} - \mu_{0y}}{\sigma_{\bar{y}}} \right)$: The input parameter
6. $\underline{D} = (D_x, \ D_y)^{/}$: The vector of weights of the respective quality characteristics.
7. $D_x = \delta_{1x}$ and $D_y = \delta_{1y}$ are the related weighs.
8. $\underline{\mu}_1 = (\mu_{1x}, \ \mu_{1y})^{/} = \left(\mu_{0x} + \left(\delta_{1x} \ \ \sigma_{\bar{x}}\right), \ \mu_{0y} + \left(\delta_{1y} \ \ \sigma_{\bar{y}}\right)\right)$: Out of control values of process mean vector.
9. $2n$: Sample size
10. k_x, k_y: The coefficients used in the control limits of the sub chart.
11. $ARL_0 = ARL\ (\underline{0})$: In control ARL value; $ATS_0 = ATS\ (0)$: In control *ATS* value.
12. $ARL_1 = ARL(\underline{\delta}_1)$: Out of control *ARL* value; $ATS_1 = ATS\ (\underline{\delta}_1)$: Out of control *ATS* value.
13. τ: Minimum required value of ATS_0
14. p_x, p_y: $P(d_x), P(d_y)$: *P*(getting signal of the *ACS* chart with *X* sample means, given that shift in the standardized mean of the *X* quality characteristic is δ_x) and similar meaning for P(δ_y). Here, $p_x = \Phi(-k_x + \delta_x\sqrt{2n}) + \Phi(-k_x - \delta_x\sqrt{2n})$ and similarly for p_y. Also, $q_x = 1 - p_x$ and $q_y = 1 - p_y$.
15. p_{1x}, p_{1y}: P(δ_{1x}), P(δ_{1y}): Power of the *ACS* chart with *X* sample means and *Y* sample means respectively. Here, $p_{1x} = \Phi(-k_x + \delta_{1x}\sqrt{2n}) + \Phi(-k_x - \delta_{1x}\sqrt{2n})$ and similarly for p_{1y}.
16. p_{0x}, p_{0y}: P($\delta_x = 0$), P($\delta_y = 0$): Probabilities of Type-I error (getting signal under the assumption that the process is running smoothly) of the *ACS* chart with *X* sample means and *Y* sample means respectively. Here, $p_{0x} = 2\Phi(-k_x)$ and $p_{0y} = 2\Phi(-k_y)$; $q_{0x} = 1 - p_{0x}$ and $q_{0y} = 1 - p_{0y}$.

3.2 The operation

Leoni and Costa (2017) developed the *ACS* chart for bivariate and trivariate processes. The operation of *ACS* chart is described in sub-section (2.2). The operation of the proposed chart is distinct from that of the *ACS* chart. As mentioned in *ACS* chart by Leoni and Costa (2017), only one of the two quality characteristics is inspected alternately per sample.

For a bivariate process, let *X* and *Y* be the two quality characteristics respectively. As mentioned in sub-section 3.1, D_x and D_y are the weights corresponding to the respective variables. There are three possible situations related to the weights. These are $D_x > D_y$, $D_x < D_y$ and $D_x = D_y$.

Case-1: $D_x > D_y$ is an indicative of *X* observations are moving away from the center line (μ_{0x}) and reaching close the *UCL* of *X* inspection. To have a conformation about status of the process, it is essential to have one more successive *X* inspection. As $D_x > D_y$, for *Y* inspection no need to carry out two successive *Y* inspections. If $D_x = D_y$, the operation of the proposed chart is same as that of ACS chart given in sub-section 2.2.

To understand the derivation of *ATS* corresponding to the *X* inspection, the stepwise procedure of the operation of *WACS* chart is as follows. Let the term 'Counter' corresponding to the quality characteristics *X* and *Y* be respectively abbreviated as CNT_x *and* CNT_y.

Part-I: Start with *X* inspection.

- **Step-1:** Initialize *i* to 0.
- **Step-2:** Initialize *CNTx* to 0.
- **Step-3:** Add *i* by unity. Take i^{th} sample of size *2n* from $N_2(\underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$ distribution. Inspect every unit in the i^{th} sample corresponding to *X* inspection only and compute sample mean $\bar{X}$. If $\bar{X}$ is within the control limits of *X* inspection, move to the next step; otherwise go to **Step-6**.
- **Step-4:** Add CNT_x by unity. If $CNT_x < 2$, go back to **Step-3**; otherwise go to the next step.
- **Step-5:** Add *i* to unity. Inspect every unit in the i^{th} sample corresponding to *Y* inspection only and compute sample mean $\bar{Y}$ in the i^{th} sample. If $\bar{Y}$ is within the control limits of *Y* inspection, go back to **Step-2**; otherwise go to **Step-6**.
- **Step-6:** The process has gone out of control. Identify the assignable causes and take a corrective action before restarting the process. Then go back to **Step-1**.

Part-II: Start with Y inspection.

Step-1: Initialize i to 0.

Step-2: Add i by unity. Take i^{th} sample of size $2n$ from N_2 ($\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$) distribution. Inspect every unit in the i^{th} sample corresponding to Y inspection only and compute sample mean $\bar{Y}$. If $\bar{Y}$ is within the control limits of Y inspection, initialize CNT_x to zero and move to the next step; otherwise go to **Step-6**.

Step-3: Initialize $CNTx$ to 0.

Step-4: Add i by unity. Inspect every unit in the i^{th} sample corresponding to X inspection only and compute $\bar{X}$. If $\bar{X}$ is within the control limits of X inspection, move to the next step; otherwise go to **Step-6**.

Step-5: Add CNT_x by unity. If $CNT_x<2$, go back to **Step-3**; otherwise go to **Step-2.**

Step-6: The process has gone out of control. Identify the assignable causes and take a corrective action before restarting the process. Then go back to **Step-1**.

Case-2: Dx < Dy is an indicative of Y observations are moving away from the center line (μ_{0y}) and reaching close to the UCL of Y inspection.

Remark-1: The stepwise procedure of the operation of *WACS* chart of Case-2 is exactly same as that of Case-1 by replacing X by Y and Y by X in Part-I and Part-II both.

Remark-2: For bivariate processes, in *ACS* chart, though the weights of the two quality characteristics are significantly different, only one of the two quality characteristics is inspected alternately. Also, the *ARL* of the *ACS* chart is just the average of the *ARLs* corresponding to two quality characteristics, without considering the weights related to the respective quality characteristics. In *WACS* chart, comparison of the weights is considered in the operation as well in the derivation of the *ATS* expression.

The constants (n, k_x, k_y) govern the implementation of the *WACS* chart are the design parameters of the proposed chart and their optimal values can be obtained with the help of *ATS* criterion and *ATS* expression of the proposed chart as discussed in the next section.

4. DESIGN OF THE *WACS* CHART

As mentioned in sub-section 3.2, we consider three cases, namely, $D_x>D_y$, $D_x<D_y$ and $D_x = D_y$.

Case-1: Derivation of *ATS* of *WACS* control chart:

Suppose the data are drawn from $N_2\,(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let X and Y be the first and the second quality characteristic respectively. As mentioned in sub-section 3.1, p_x and p_y are probabilities of a unit being defective corresponding to the X and Y respectively. Consider the following Table 1:

Table 1: Relationship between Sample No., Inspection type and P (Getting Signal corresponding to the Sample No.)

Sample No.	Inspection type	Probability of getting signal
1	X	p_x
2	X	$q_x p_x$
3	Y	$q_x^2 p_y$
4	X	$q_x^2 q_y\, p_x$
5	X	$q_x^3 q_y\, p_x$
6	Y	$q_x^4 q_y\, p_y$
…	…	---
3*i*-2	X	$\left(q_x^2 q_y\right)^{i-1} p_x$
3*i*-1	X	$q_x \left(q_x^2 q_y\right)^{i-1} p_x$
3*i*	Y	$\left(q_x^2 p_y\right)\ \left(q_x^2 q_y\right)^{i-1}$
….	….	…
…	….	…

Now, '*ARL* of the first quality characteristic (X, Say)'(ARL_x) is,

$$ARL_x = p_x \left\{ \sum_{i=1}^{\infty} (3i-2) \left(q_x^2\, q_y\right)^{(i-1)} + q_x \sum_{i=1}^{\infty} (3i-1) \left(q_x^2\, q_y\right)^{(i-1)} \right\} + 3\, \left(q_x^2\ \ p_y\right) \sum_{i=1}^{\infty} i\ \left(q_x^2\ \ q_y\right)^{(i-1)}$$

$$ARL_x \frac{p_x \ \{1+2q_x^2 \ q_y + q_x \ (2+q_x^2 \ q_y \ )\} + 3 \ q_x^2 \ p_y}{(1 - q_x^2 \ q_y)^2} \tag{7}$$

and that of the second quality characteristic (Y, Say) is,

$$ARL_y = p_y \sum_{i=1}^{\infty} (3i-2)\left(q_x^2 \, q_y\right)^{(i-1)} + q_y \ p_x \sum_{i=1}^{\infty} (3i-1)\left(q_x^2 \, q_y\right)^{(i-1)} + 3 \, q_y \, q_x \, p_x \sum_{i=1}^{\infty} i\left(q_x^2 \, q_y\right)^{(i-1)}$$

$$= \frac{p_y \ \left(1 + 2 \ q_x^2 \ q_y\right) + q_y \ p_x \ \left(2 + q_x^2 \ q_y + 3q_x\right)}{\left(1 - q_x^2 \ q_y\right)^2}. \tag{8}$$

as $ATS = n \ (ARL)$, $ATS = \dfrac{n \left(ARL_x + ARL_y\right)}{2}$.

Remark-3: For **Case-2**, ARL_x and ARL_y expressions are,

$$ARL_x = \frac{p_x \ \left(1 + 2 \ q_y^2 \ q_x\right) + q_x \ p_y \ \left(2 + q_y^2 \ q_x + 3q_y\right)}{\left(1 - q_y^2 \ q_x\right)^2}. \tag{9}$$

$$ARL_y = \frac{p_y \{1+2q_y 2 \ q_x + q_y \ (2+q_y^2 \ q_x \ )\} + 3 \ q_y^2 \ p_x}{(1 - q_y^2 \ q_x)^2}. \tag{10}$$

Remark-4: For **Case-3**, ARL_x and ARL_y expressions are same as given in Leoni and Costa (2017). These are,

$$ARL_x = \frac{\left(p_x\left(1+q_x q_y\right)+2p_y q_x\right)}{\left(1-q_x q_y\right)^2}. \tag{11}$$

$$ARL_y = \frac{\left(p_y\left(1+q_y q_x\right)+2p_x q_y\right)}{\left(1-q_x q_y\right)^2}. \tag{12}$$

Additional Notations: Following are some additional notations of the χ^2, *ACS* and *WACS* charts.

1. ARL_{ch} = *ARL* of the χ^2chart,
2. P_{ch} = P (Getting Signal corresponding to the χ^2 chart), $ARL_{ch} = \dfrac{1}{P_{ch}}$.
3. ARL_a and ARL_{wa} are *ARLs* of the *ACS* and *WACS* charts. Similarly, P_a and P_{wa} are to be used.

To compare the performance of the proposed chart with the χ^2 chart and the *ACS* chart, some numerical illustrations are discussed in the next section.

5. Numerical Illustrations

5.1. Example 1

For the first example, input parameters are $(\delta_x, \delta_y) = (0, 0.5)$ and $\tau = 370$ are considered. For these input parameters, values of the design parameters for the related three charts along with respective ATS_1 values are as follows:

χ^2 chart (when $\rho = 0.7$): $n_{ch} = 15$, $k_{ch} = 6.42$, $ATS_{1ch} = 23.2314$

χ^2 chart (when $\rho = 0.5$): $n_{ch} = 19$, $k_{ch} = 5.94$, $ATS_{1ch} = 30.9906$

χ^2 chart (when $\rho = 0.3$): $n_{ch} = 22$, $k_{ch} = 5.65$, $ATS_{1ch} = 35.7164$

χ^2 chart (when $\rho = 0$): $n_{ch} = 23$, $k_{ch} = 5.56$, $ATS_{1ch} = 38.2393$

ACS chart: $n_a = 12$, $k_{ax} = 3.29$, $k_{ay} = 1.86$, $ATS_{1a} = 27.2111$

WACS chart: $n_{wa} = 11$, $k_{wax} = 3.12$, $k_{way} = 2.02$, $ATS_{1wa} = 24.7793$

This example shows that for $\rho < 0.7$, the ATS_{1wa} is less than ATS_1 of the other five related charts. For further comparison of the *WACS* chart with the χ^2charts for four values of ρ and *ACS* chart for various values of $\underline{\delta}$, the normalized *ATS* (normalized w.r.t. the χ^2 chart for $\rho = 0.5$) are computed. The *ATS* values and normalized *ATS* values are given in Tables 2 and 3.

The graphs of normalized *ATS* values against the normalized *ATS* values related to the data in Table 3 are given in Figure 1, from which it is observed that for all twenty-five entries, *WACS* chart performs better as compared to the *ACS* chart; also *WACS* chart performs better as compared to χ^2chart for small to moderate correlation coefficient ρ_{xy} (i.e. $\rho < 0.7$). For $\rho = 0.7$, when δ_y is reaching to δ_{1y} or large than δ_{1y}, *WACS* chart performs better as compared to the five charts.

Table 2: Comparative Study of the *ATS* values of the Six Charts

Sr. No.	(δ_x, δ_y)	*ACS*	*WACS*	$\chi^2 (\rho = 0.7)$	$\chi^2 (\rho = 0.5)$	$\chi^2 (\rho = 0.3)$	$\chi^2 (\rho = 0)$
1	(0, 0.024)	359.6111	359.3316	357.7862	361.8380	363.3739	364.0254
2	(0, 0.048)	331.5102	330.1180	323.5450	337.3781	342.8317	345.1819
3	(0, 0.072	293.0560	290.3676	278.3287	302.8188	313.0207	317.5179
4	(0,0.096	251.7401	247.9552	231.8375	264.2389	278.5881	285.0826
5	(0, 0.12)	212.6875	208.1569	189.8978	226.3554	243.5052	251.4810
6	(0, 0.144)	178.3791	173.4391	154.7171	191.9302	210.4394	219.2750
7	(0, 0.168)	149.4641	144.3721	126.3572	162.1071	180.7972	189.9365
8	(0, 0.192)	125.6346	120.5652	103.9598	137.0111	155.0659	164.0883
9	(0, 0.216)	106.2079	101.2706	86.4332	116.2501	133.1783	141.8036
10	(0, 0.24)	90.4325	85.6909	72.7540	99.2344	114.7887	122.8523
11	(00.264),	77.6202	73.1078	62.0664	85.3487	99.4457	106.8679
12	(0, 0.288)	67.1888	62.9303	53.6914	74.0308	86.6885	93.4466
13	(0, 0.312)	58.6651	54.6430	47.1042	64.7990	76.0928	82.1998
14	(0, 0.336)	51.6712	47.8910	41.9041	57.2546	67.2883	72.7781
15	(0, 0.36)	45.9085	42.3613	37.7864	51.0741	59.9623	64.8792
16	(0, 0.384)	41.1412	37.8159	34.5191	45.9980	53.8547	58.2472
17	(0, 0.408)	37.1834	34.0678	31.9246	41.8187	48.7523	52.6688
18	(0, 0.432)	33.8878	30.9689	29.8663	38.3708	44.4805	47.9676
19	(0, 0.456)	31.1371	28.4021	28.2377	35.5221	40.8975	43.9982
20	(0, 0.48)	28.8376	26.2734	26.9552	33.1667	37.8874	40.6409
21	(0, 0.504)	26.9138	24.5073	25.9521	31.2196	35.3559	37.7974
22	(0, 0.528)	25.3043	23.0427	25.1746	29.6116	33.2257	35.3868
23	(0, 0.552)	23.9590	21.8292	24.5785	28.2866	31.4335	33.3422
24	(0, 0.576)	22.8368	20.8257	24.1275	27.1983	29.9268	31.6082
25	(0, 0.6)	21.9031	19.9976	23.7915	26.3084	28.6623	30.1386

Table 3: Comparative Study of the Normalized *ATS* values of the Six Charts

Sr. No.	(δ_x, δ_y)	*ACS*	*WACS*	$\chi^2(\rho = 0.7)$	$\chi^2(\rho = 0.5)$	$\chi^2(\rho = 0.3)$	$\chi^2(\rho = 0)$
1	(0, 0.024)	0.993845588	0.993073143	0.988802171	1	1.004244717	1.006045247
2	(0, 0.048)	0.982607348	0.97848082	0.958998228	1	1.016164653	1.023130725
3	(0, 0.072	0.967760258	0.958882342	0.919126223	1	1.033689784	1.04854091
4	(0,0.096	0.952698865	0.938375084	0.877378388	1	1.054303889	1.078882027
5	(0, 0.12)	0.939617522	0.919602095	0.838936469	1	1.075764925	1.111000665
6	(0, 0.144)	0.929395687	0.903657163	0.806111284	1	1.096437142	1.142472628
7	(0, 0.168)	0.922008351	0.890597019	0.779467401	1	1.115294765	1.171672925
8	(0, 0.192)	0.916966582	0.87996666	0.75876918	1	1.131776185	1.197627783
9	(0, 0.216)	0.913615558	0.871144197	0.743510758	1	1.145618799	1.219814865
10	(0, 0.24)	0.911301928	0.86352011	0.73315302	1	1.156743025	1.238001137
11	(00.264),	0.909447947	0.85657778	0.72720967	1	1.165169475	1.252132721
12	(0, 0.288)	0.907579008	0.850055653	0.725257595	1	1.170978836	1.262266516
13	(0, 0.312)	0.905339589	0.843269186	0.726927885	1	1.174289727	1.268535008
14	(0, 0.336)	0.902481198	0.836456809	0.731890538	1	1.175247054	1.271131053
15	(0, 0.36)	0.898860675	0.829408644	0.739834867	1	1.174025582	1.270295512
16	(0, 0.384)	0.894412801	0.822120527	0.750447846	1	1.170805252	1.266298535
17	(0, 0.408)	0.889157243	0.814654688	0.763404888	1	1.165801424	1.259455698
18	(0, 0.432)	0.883166366	0.8070955	0.778360107	1	1.15922785	1.250106852
19	(0, 0.456)	0.876555722	0.7995614	0.794933295	1	1.151325513	1.238614834
20	(0, 0.48)	0.869474503	0.792162018	0.812718781	1	1.14233252	1.225352537
21	(0, 0.504)	0.862080232	0.784997245	0.831275865	1	1.132490487	1.210694564
22	(0, 0.528)	0.854540113	0.778164638	0.850160072	1	1.122050143	1.195031677
23	(0, 0.552)	0.847008831	0.771715229	0.86890966	1	1.111250557	1.178727737
24	(0, 0.576)	0.839640713	0.765698591	0.887095885	1	1.10031877	1.16213881
25	(0, 0.6)	0.83255158	0.760122242	0.904330936	1	1.089473324	1.145588481

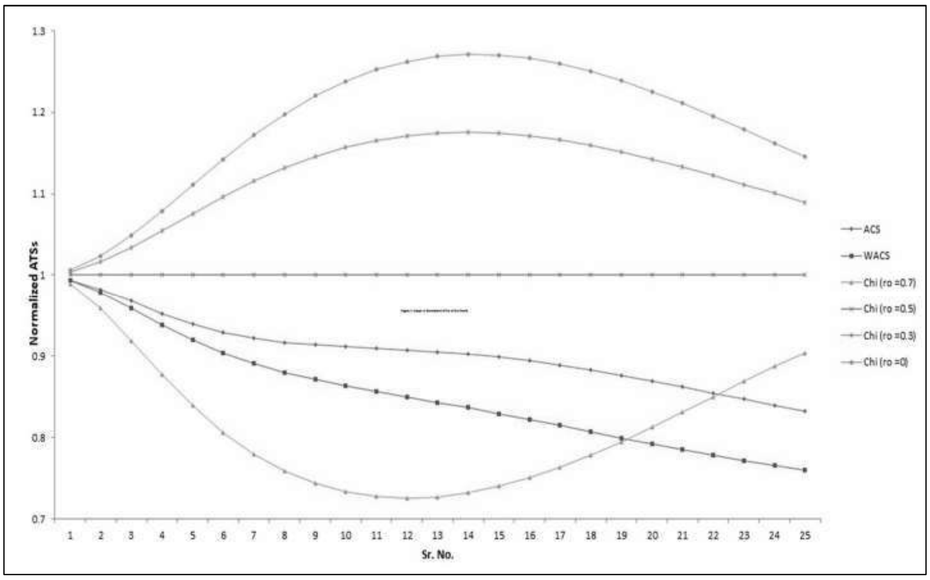

Figure 1: Graph of Normalized ATSs of Six Charts

5.2 Example 2

For $\tau = 370$, numerical illustrations are considered to carry out to compare the performance of the six control charts corresponding to 15 combinations of input parameters $(\delta_{lx}, \delta_{ly})$ and to study their effect on the performance of *WACS* chart.

Considering all possible 15 combinations of the input parameters, values of the design parameters along with respective ATS_1 values are computed for each of the six control charts and are given in Tables 4 and 5.

Results:

1. From Table-4, for $\underline{\delta_l \neq 0}$ and $\rho < 0.7$, $ATS_{1wa} < ATS_{1a} < ATS_{1ch}$. For $\delta_{lx} > 0$ and for four values of ρ, $ATS_{1wa} < ATS_{1a} < ATS_{1ch}$. Also for $\underline{\delta_l \neq 0}$ and for four values of ρ, $n_{wa} \leq n_a \leq n_{ch}$.
2. Also for $\underline{\delta_l \neq 0}$, and for four values of ρ, $n_{wa} \leq n_a \leq n_{ch}$.
3. From Table-5, it is observed that simulated ATS_1 is larger than related ATS_1; when $\delta_{lx} = \delta_{ly}$, simulated ATS_1 is significantly larger than related ATS_1.

Table 4: Optimum values of the design parameters and ATS_1 values of *ACS*, *WACS* and χ^2 (for four values of ρ) charts

$\underline{\delta}_1 = (\delta_{1x}, \delta_{1y})$	ACS				WACS				χ^2			χ^2			χ^2			χ^2		
									$\rho = 0.7$			$\rho = 0.5$			$\rho = 0.3$			$\rho = 0$		
	n	k_x	k_y	ATS_1	n	k_x	k_y	ATS_1	n	k	ATS_1	n	k	ATS_1	n	k	ATS_1	n	k	ATS_1
(0, 0)	6	2.34	2.48	370.0100	3	2.46	3	370.0067	1	7.04	370.0200	1	9.06	370.0100	2	7.44	370.0200	2	8.61	370.0100
(0, 0.50)	12	3.29	1.86	27.2111	11	3.12	2.02	24.7793	15	6.42	23.2314	19	5.94	30.9906	22	5.65	35.7164	23	5.56	38.2393
(0, 0.75)	6	3.42	2.15	14.8246	6	3.35	2.26	13.4207	8	7.67	12.3516	11	7.04	16.7419	13	6.70	19.4503	13	6.70	20.9144
(0, 1)	4	3.42	2.31	9.4583	4	3.52	2.41	8.5541	5	6.61	7.7828	7	7.94	10.6136	8	7.67	12.3891	9	7.44	13.3533
(0, 1.50)	2	3.61	2.56	4.9687	2	3.9	2.65	4.4442	3	9.63	3.9810	4	9.06	5.4827	4	9.06	6.4717	5	8.61	6.9472
(0.50, 0.50)	12	2.12	2.16	19.3067	12	2.12	2.16	19.3067	21	5.74	33.9704	19	5.94	30.9906	18	6.05	27.8786	14	6.55	22.8865
(0.50, 0.75)	7	2.46	2.26	13.3757	7	2.41	2.32	12.5512	13	6.70	20.8782	13	6.70	20.3251	12	6.86	18.7988	10	7.23	15.7359
(0.50, 1)	4	3.14	2.33	9.2816	4	2.97	2.44	8.4432	8	7.67	12.5848	9	7.44	13.3533	9	7.44	12.9230	8	7.67	11.2044
(0.50, 1.50)	2	3.61	2.56	4.9521	2	3.99	2.65	4.4402	4	9.06	5.7212	5	8.61	6.7608	5	8.61	6.9407	4	9.06	6.4049
(0.75, 0.75)	7	2.31	2.39	10.3563	7	2.31	2.39	10.3563	12	6.86	18.4397	11	7.04	16.7419	10	7.23	14.9678	8	7.67	12.1533
(0.75, 1)	5	2.53	2.42	8.0122	5	2.42	2.5	7.6807	9	7.44	13.3028	8	7.67	12.5371	8	7.67	11.3902	6	8.25	9.3578
(0.75, 1.50)	3	3.43	2.57	4.8965	2	3.45	2.66	4.4124	5	8.61	6.5748	5	8.61	6.9472	5	8.61	6.7281	4	9.06	5.7725
(1, 1)	4	2.52	2.58	6.5605	4	2.52	2.58	6.5605	8	7.67	11.7290	7	7.94	10.6136	6	8.25	9.4852	5	8.61	7.6490
(1, 1.50)	3	2.69	2.61	4.4918	2	3	2.71	4.2310	5	8.61	6.9357	5	8.61	6.7608	4	9.06	6.2165	4	9.06	5.1729
(1.50, 1.50)	2	2.74	2.83	3.4173	2	2.74	2.83	3.4173	4	9.06	6.0804	4	9.06	5.4827	3	9.63	4.9153	3	9.63	3.9235

Table 5: ATS_1 and Simulated ATS_1 values of ACS and WACS charts

$\underline{\delta}_1 = (\delta_{1x}, \delta_{1y})$	ACS		WACS	
	ATS_1	Simulated ATS_1	ATS_1	Simulated ATS_1
(0, 0)	370.0100	369.9872	370.0067	369.9872
(0, 0.50)	27.2111	27.2450	24.7793	24.8824
(0, 0.75)	14.8246	14.8393	13.4207	13.5111
(0, 1)	9.4583	9.4331	8.5541	8.4725
(0, 1.50)	4.9687	4.9661	4.4442	4.4191
(0.50, 0.50)	19.3067	23.2723	19.3067	23.2723
(0.50, 0.75)	13.3757	15.5588	12.5512	14.4647
(0.50, 1)	9.2816	9.5550	8.4432	8.7353
(0.50, 1.50)	4.9521	4.9773	4.4402	4.4303
(0.75, 0.75)	10.3563	12.4402	10.3563	12.4402
(0.75, 1)	8.0122	9.5984	7.6807	9.5065
(0.75, 1.50)	4.8965	5.4862	4.4124	4.4887
(1, 1)	6.5605	7.8958	6.5605	7.8958
(1, 1.50)	4.4918	5.3547	4.2310	4.6416
(1.50, 1.50)	3.4173	5.4799	3.4173	5.4799

5.3. A Real Life Example

This example is given to illustrate the use of the proposed chart and compare it with the χ^2 chart and *ACS* chart. The data set is collected by M.Sc. Statistics students (Miss Vidya Sawant and Miss Rutuja Sanap) for their project under my guidance. The data are from most important part, caliper of the brake system that measured the Lug-hole CD which is distance from two bottom holes of the caliper (X) with the specification 142.05 ± 0.75 mm and diameter which is the distance of center hole (Y) with the specification of 51.07 ± 0.15 mm for 800 observations. According to historical information about this type of Caliper, the in-control mean vector and covariance matrix were taken as $\underline{\mu}_0 = \begin{bmatrix} 51.07 \\ 142.05 \end{bmatrix}$.

Here, Σ0 is taken as $\Sigma_0 = \begin{bmatrix} 1 & .5 \\ 0.5 & 1 \end{bmatrix}$. Assuming that the in-control process has a $N_2(\underline{\boldsymbol{\mu}}_0, \Sigma_0)$ distribution, the process is stable with respect to its mean vector. As per the specifications, $\underline{\mu}_1$ is taken as $\underline{\mu}_1 = \begin{bmatrix} 51.22 \\ 142.8 \end{bmatrix}$. Now $\underline{\delta}_1 = (\delta_{1x},$

$\delta_{1y})'$ is computed by using the relationship between $\underline{\mu}_0$, $\underline{\mu}_1$ and $\underline{\delta}_1$. Here $\underline{\delta}_1$ = (0.4743, 2.3717)$'$ and ρ = 0.5 is considered. Let τ = 370. By using *ATS* criterion, following are the design parameters of the three related control charts along with the ATS_1 values.

1. χ^2 Chart: n_{ch} = 4, k_{ch} = 9.055, ATS_{1ch}= 5.65
2. *ACS* Chart: n_a = 1, k_{ax}= 3.84, k_{ay} = 2.79, ATS_{1a}= 2.3012
3. *WACS* Chart: n_{wa} = 1, k_{wax} = 3.8, k_{way} =2.88, ATS_{1wa} = 2.0771

For this real life example, we observed that $n_{ch} < n_a = n_{wa}$ and $ATS_{1wa} < ATS_{1a} < ATS_{1ch}$. This example also shows that *WACS* chart performs better as compared to both the related two control charts.

6. Conclusion

A procedure of the proposed *WACS* control chart is operationally easier and efficient as compared to the χ^2 chart. Though the operation of *WACS* chart is different from *ACS* chart, for all combinations of $\underline{\boldsymbol{\delta}}_1$ *WACS* chart performs better as compared to the *ACS* chart. Looking to the real life example also, a statistician may prefer to *WACS* chart as compared to the related two control charts.

Acknowledgement

The Author would like to thank the reviewers, the editors and Professor R. N. Rattihalli for their constructive comments, which helped me to improve this manuscript. Also, I would like to thank Miss Vidya Sawant and Miss Rutuja Sanap (my M. Sc. Project Students in the Department of Statistics, SPPU) for giving me a data set to study the *ATS* performance of the related charts.

References

1. Wu, Z. and Spedding, T.A. (2000): A Synthetic Control Chart for Detecting Small Shifts in the Process Mean, *Journal of Quality Technology*, 32(1), 32-38.
2. Hotelling, H. (1947): Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights; In: Eisenhart, C., Hastay, M.W. and Wallis, W.A., Eds., Techniques of Statistical Analysis, McGraw Hill, New York, 111-184.

3. Ghute, V.B. and Shirke, D.T. (2008): A Multivariate Synthetic Control Chart for Monitoring Process Mean Vector; *Commun. Stat. Theo. Meth.*, 37, 2136-2148.
4. Wu, Z., Yeo, S.H. and Spedding, T.A. (2001): A Synthetic Control Chart for Detecting Fraction Nonconforming Increases, *J. Qual. Tech.,* 33(1), 104-111.
5. Gadre, M.P. and Kakade, V.C. (2016): Some Group Runs Based Multivariate Control Charts for Monitoring the Process Mean Vector, *Open Journal of Statistics*, 6(6), 1098-1109.
6. Gadre, M.P. and Kakade, V.C. (2019): Multivariate Unit and Group Runs Control Chartto Detect Shifts in the Mean Vector; *Ind. Assoc. Prod. Qual. Reliab. Trans.,* 44(1), 29-42.
7. Woodall, W.H. and. Ncube, M.M. (1985): Multivariate CUSUM quality-control procedures, *Technometrics*, 27, 285-292.
8. Hamed, M.S., Mansour, M.M. and Elrazik, E.M.A. (2016): MCUSUM Control Chart Procedure: Monitoring the Process Mean with application, *J. Stat. Adv. Theo. Appl.*, 16(1), 105-132.
9. Xia, B., Jian, Z., Liu, L. and Li, L. (2018): An effective multivariate control chart for detecting small mean shifts using support vector data description, *Adv Mech. Engg,*, 10(11), 1-18.
10. Reynolds (Jr)., M.R. and Cho, G.Y. (2011): Multivariate Control Charts for Monitoring the Mean Vector and Covariance Matrix with Variable Sampling Intervals, *Sequential Analysis,* 30, 1-40.
11. Kang C.W., Lee, M.S., Seong, Y.J. and Howkins D.M. (2007): A control chart for coefficient of variation, *J. Qual. Tech.*, 39 (2), 151-158.
12. Melo, M.S., Ho, L.L. and Medeiros, P.G. (2017): Max D: An attribute control chart to monitor a bivariate process means, *Inter. J. Adv. Manuf. Technol.*, 90, 489-498.
13. Leoni, R.C. and Costa, A.F.B. (2017): Monitoring bivariate and trivariate mean vectors with a Shewhart chart, *Qual. Reliab. Engg. Inter*, 33, 2035-2042.
14. Montgomery, D.C. (2006): *Introduction to Statistical Quality Control*. John Wiley and Sons, New York.
15. Gadre, M.P. and Padappa, Nisha. (2022): Hybrid synthetic and group runs charts using alternating the charting statistic to monitor the mean vector of a multivariate process, *Commun. Stat. Theo. Meth.,* 51(21), 7607-7630.

IAPQR Transactions
Vol. 47, No. 1-2, 2022–2023

RELIABILITY AND AVAILABILITY IMPROVEMENT OF RAW MATERIAL CHARGING STATION IN A STEEL INDUSTRY

SUBRATA RATH[1*], ASHIS KUMAR CHAKRABORTY[2] AND SUBHASHIS CHATTERJEE[3]
[1*]*SQC & OR Unit, Indian Statistical Institute, Pune, India*
[2]*SQC & OR Unit, Indian Statistical Institute, Kolkata, India*
[3]*Mathematics Department, IIT(ISM) Dhanbad, India*

***Abstract*:** Raw Material Charging Station (RMCS) is one of the most important infrastructures in steel industry and is used to load and carry crushed iron ore to furnace. Breakdown behaviour of the RMCS was observed for a certain period, and its future behaviour was predicted using a reliability model. For improvement of availability however, a Six sigma approach was used. Initial focus was to reduce the repair time per month in order to make RMCS more available for use and save the financial loss due to unavailability of the furnaces to produce steel. This study discovers a way to simultaneously improve reliability, reduce repair time and thereby improve availability of the RMCS. The study also generalises an approach for improving reliability and availability for any mechanised repairable system using six sigma methodologies.

***Keywords and Phrases*:** Six Sigma, Define-Measure-Analyse-Improve-Control (DMAIC), Supplier-Input-Process-Customer-Output (SIPOC), Mean Time Between Failure (MTBF), Mean Time to Repair (MTTR), Raw Material Charging Station (RMCS), Critical to Quality (CTQ), Reliability, Availability.

1. INTRODUCTION

In a steel industry three bucket elevators connected with a RMCS for supply of iron ore in five furnaces. Breakdown of the station poses serious threat to the

---

* Corresponding author email: srath02@yahoo.com
DOI-https://doi.org/10.32381/IAPQRT.2023.47.01-02.4

production run resulting in loss of production. A typical RMCS involves a station with motors, bucket elevators, and iron ore charging point. Crushed iron ore are poured in running bucket elevators to carry to the respective furnaces. Mechanical, electrical and instrumentation related malfunction of the system leads to the breakdown and disrupts the production run. A typical RMCS with bucket elevator is shown in Figure 1.

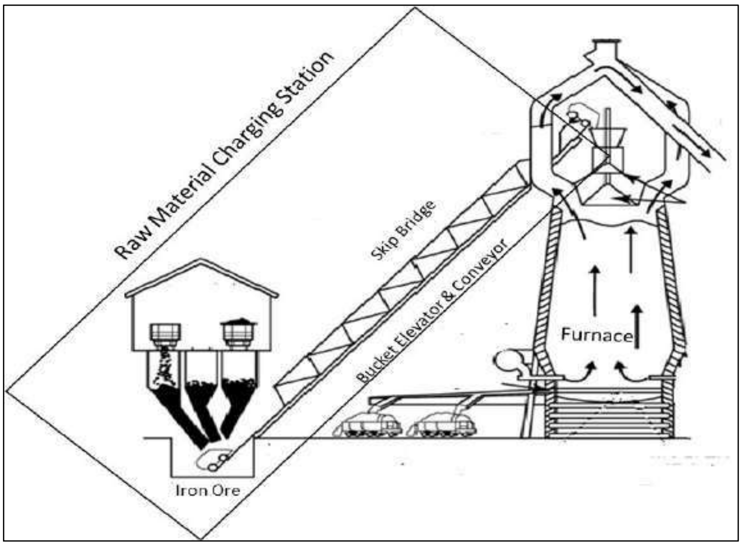

Figure 1: Raw Material Charging Station

Management of the organization felt the need for making the system available to a maximum extent in order to boost productivity of steel. Initial assessment of the performance of the station in terms of reliability and availability led the management to go in for an in-depth study to investigate the reasons for the disruption of the station and evolving solutions subsequently. Six Sigma-DMAIC approach was thought to be of use to improve the reliability and availability of the system.

The approach adopted in this article starts with an initial assessment of reliability and availability of the station at different time points. We then developed a solution methodology to improve and sustain the reliability and availability of the RMCS using Six Sigma.

This article is structured as follows. Section 2 briefly describes literature survey, while in section 3 the development of the methodologies to improve both reliability and availability is provided. We then concluded the article in section 4.

Literature in the field of reliability improvement using Six Sigma approach is not that large, however, there is a vast literature on Six Sigma and its applications in process development. Six Sigma, as it started in Motorola is given by Mikel ([1],1989). Since then, Six Sigma expanded all over the world like wildfire. Books by Mikel ([2], 2000) and Breyfogle ([3], 2003), promoted the spread of Six Sigma knowledge and practices in a mass scale across the globe. Industries of the types such as manufacturing, chemicals, electronics and even service industries were benefited using Six Sigma methodologies. However, interestingly, applications of Six Sigma methodologies for reliability or availability improvement even in these types of industries were rare. Leitch ([4], 1995), Pham ([5], 2003), Kumar et al., ([6], 2006), Al-Mishari and Suliman ([7], 2008) and Smith et al., ([8], 2011) illustrated some applications of six sigma for reliability improvement. Application of Six Sigma for reliability improvement of a HTDC compressor is also vividly described in Rath and Chakraborty ([9], 2019).

2. Reliability and Availability Improvement using Six Sigma

Reliability and Six Sigma are two different subjects developed almost independently. Reliability, or in particular, hardware reliability is a subject developed to ensure that products do not fail unnecessarily, whereas, Six Sigma is a quality initiative developed to improve processes. In this study, reliability and availability, two very important metrics are taken up for improvement using Six Sigma approach. Six Sigma approach have got five phases (Breyfogle, [3], 2003) and the article provides brief details of each of the phases.

2.1 Define Phase

Typically, define phase of Six Sigma has four tollgate steps to be accomplished.

- Project charter which includes why the project is important as per the top management priority.
- Listening to the stakeholders including customers in connection with the problems or and deriving representative metrics, CTQ characteristics for the respective stakeholders.
- If required, project CTQs, called Big Ys are drilled down to make the problem more focused and manageable to solve.
- Mapping the processes at macro level to identify all the platforms which essentially are the sources for potential causes for the problems.

Project charter through the business case illustrated the consequences of the poor reliability and availability of the RMCS. A loss in terms of 0.12 million of rupees of annualized recurring loss was estimated on account of only repair time of the system in case of breakdown. The estimate is conservative as it considered only the production loss. It did not consider the spares used, time engaged by the team of fitters, operators and engineering team for the job. Unlike typical one metric oriented six sigma projects, the problems are characterized as Mean Time Between Failure (MTBF), Reliability at the targeted time to survive (i.e., at 500 hours), Mean Time to Repair (MTTR) and Availability of the system for production. All the metrics have got their own merits though causes for the problems of the metrics may or may not be different. Twelve months data were considered for estimation of all the metrics and targets were set to achieve. However, considering that the scope of the problem is so huge to go unmanageable, only electrical breakdown was considered, to keep it more focused and appropriate, given the considerations, prevailing that time. One team was formed with a project leader who is or holds a trained Six Sigma Black Belt with another five cross functional team members. Eight months of targeted duration were set to complete the study with the split targets for five phases viz. Define, Measure, Analyze, Improve, and Control.

Having obtained the approval of the top management to go ahead with the study, next task was to identify the stakeholders in connection with the problem and listen to them through a structured process for deriving the requirements or concerns of the stakeholders from their respective voices, called Voice of the Stake holders (VoS) or Voice of the Customers (VoC) and ultimately deriving the CTQ characteristics. Such VoS template is made available for RMCS in Table 1.

Table 1: Voice of the Stakeholders Template

Stakeholders/ Customers	Sample Voices	Requirements/ Concerns	CTQ	Code
Production (Customer)	Looking for zero production loss due to breakdown	Less Breakdowns & repair	Time Between	Y1
			Time to Repair	Y2
			Reliability at the targetted time	Y3
			Availability	Y4
Management/ Owner	How much are we loosing out of the breakdowns	Opportunities of savings on the event of breakdown reduction	Savings per month	Z1

	Less losses of energy during running of furnace without charge	Energy savings opportunities on the event of breakdown reduction	Energy savings per month	Z2
Engineering team	Time will be spared for preventive maintenance of the equipment	Availability of time to the engineering team on the event of breakdown reduction	Time savings per month	Z3
Customer	On time delivery has been a chronic deficiency	Ontime delivery to the end customer is an issue of concern	Ontime Delivery percentage	Z4

Unlike usual Six Sigma projects, here multiple project CTQs are noted in terms of Y1, Y2, Y3 and Y4. Consequent variables, denoted as Z1, Z2, Z3 and Z4 are commercial benefits and soft benefits which would be accrued if the problem is solved. Problems envisaged in the project charter perfectly match with the voice of the stakeholders.

Project CTQs, selected for the study were focused enough and did not require further CTQ drilling down.

Processes for maintenance and use of RMCS were then investigated with a view to map it in a macro level, following SIPOC model (Rath and Chakraborty ([9], 2018)) and is available in Figure 2.

These processes are considered the main areas where the causes for the reliability and availability problems could be lying.

2.2. Measure Phase

Measure phase primarily deals with three tollgate steps as under:

- Plan to collect data
- Ensure that the measurement uncertainty is known and suitable following measurement system analysis processes
- Collect baseline data and establish baseline process capability.

Data were planned to be collected on Time Between Failures (TBF) and Time to Repair (TTR) for 12 months period and a sample of data are shown in Table 2. Target for time to repair was in the monthly level, and not in the specific instances of breakdown repair. Also, the management wanted to get a good estimate of breakdowns of RMCS. However, other measure such as TTR, TBF, Availability percentage were considered for the significant process improvement and it was made sure that such improvement will enable the monthly targets to be accomplished.

Supplier	Input	Code	Processes	Resources	Output	Customer
Supplier	Raw material	P	Use and Maintan the RMCS	RMCS	Uninterrupted operation of RMCS	Production
Energy Supplier	Energy					
		P1	Review the RMCS engineering and operational configurations	Engineering team		
		P2	Use the system	RMCS and Iron ore storage		
		P3	Maintain the RMCS			
		P31	Preventive maintenance on RMCS	Preventive Maintenance schedule		
		P32	Breakdown maintenance of RMCS			
		P321	Indentify the breakdown instance			
		P322	Communicate to furnace Control Room	Furnace Control room		
		P323	Communicate to Electrical Control Room	Electrical Control Rool		
		P324	Assign a Person for repair by Shift Incharge	Maintenance Crew and Tool Room		
		P325	Assigned Person attends the breakdown			
		P326	Check for functioning of equipment, if minor problem in control/Power circuit than immediately resolved the problem.			
		P327	In case spare parts required person will go check spare availability in Shift rack/Shop floor	Store/Supply Chain Management for Spares		
		P328	If Spare part not available in above mentioned place than person will go to central store	Central Store		
		P329	Issue of spares			
		P330	Replace the spares and check for functioning of equipment.			
		P331	Communication to furnace Control Room for their service			
		P4	Periodic assessment of RMCS			

Figure 2: Process details for Raw Material Charging Station

Table 2: Data on baseline TBF, TTR and Availability

Month	TTR-monthly	No. of incidences of breakdown	Instances of Breakdown	Repair time	TBF	Availability
Unit of measure	Hours	No.		Hours	Hours	%
Target	<= 3.5					
M1	9.25	4	1	2.8		
			2	2	74.55	97.39
			3	2.4	235.15	98.99
			4	2.05	150.62	98.66
M2	2.41	1	5	2.41	460.21	99.48
M3	5.5	2	6	2.6	509.19	99.49
			7	2.9	435.10	99.34
M4	3.84	3	8	1.4	422.52	99.67
			9	1.3	216.50	99.40
			10	1.14	306.20	99.63
M5	2.76	2	11	1.36	297.14	99.54
			12	1.4	202.14	99.31
M6	3.67	2	13	1.9	461.00	99.59
			14	1.77	387.85	99.55
M7	11.9	4	15	3.1	204.75	98.51
			16	2.9	179.33	98.41
			17	3.4	152.03	97.81
			18	2.5	296.03	99.16
M8	14.74	5	19	3	103.65	97.19
			20	3.8	125.15	97.05
			21	2.2	145.37	98.51
			22	2	185.53	98.93
			23	3.74	186.37	98.03
M9	4.84	2	24	2.87	425.38	99.33
			25	1.97	244.13	99.20
M10	3.25	2	26	2.1	305.66	99.32
			27	1.15	247.43	99.54
M11	4.81	2	28	2.57	441.97	99.42
			29	2.24	483.75	99.54
M12	7.47	3	30	2.15	229.71	99.07
			31	3.11	203.33	98.49
			32	2.21	302.34	99.27

Measurement system analysis was not required to be done in this case as the instruments used to measure the time duration and values pose no challenge to measurement system accuracy and precision.

Targets were set by the management on time to repair as less than or equal to 3.5 hours per month in order to make the RMCS more dependable and as a consequence a sum of 0.12 million rupee recurring annualised savings would be accrued.

Process capability analysis was made on this process of maintenance. Baseline monthly repair time data did not show to come from normal population. However, Johnson transformation of the baseline data could transform the data to normal and the Johnson transformation and the distribution status is/are shown in Figure 3.

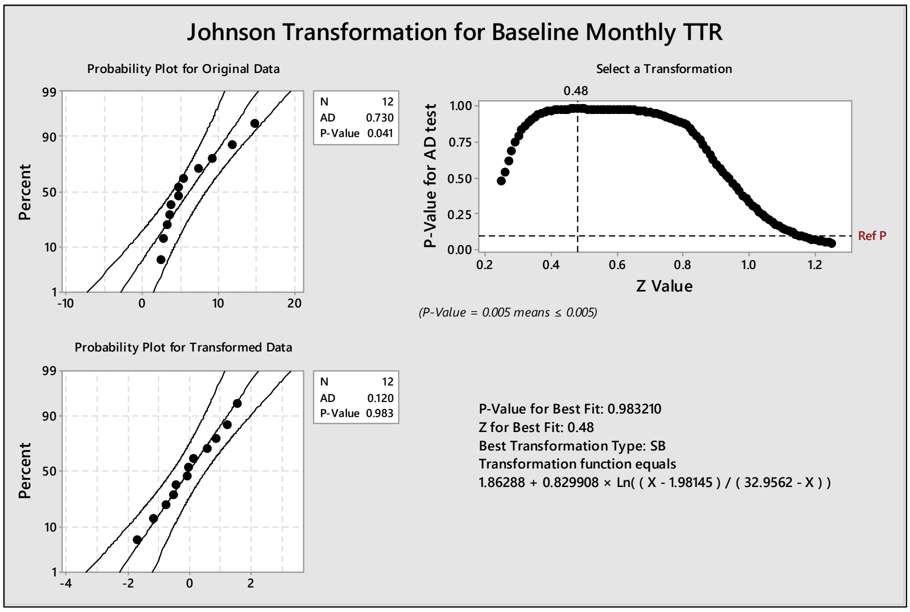

Figure 3: Johnson Transformation Status

Process capability analysis was done with transformed data and the capability of the process is shown in Figure 4.

Average of the monthly time to repair was observed as 6.2 hours with standard deviation 3.9 hours which is way beyond the target of the system to go below 3.5 hour per month. This poor status is also reflected with the process capability index, Ppk at -0.020 and defects per million opportunity (dpmo) 7,26,173.02. Process baseline sigma level of short-term with this dpmo is estimated as 0.90 and long-term sigma level at -0.60 (i.e., 0.9-1.5).

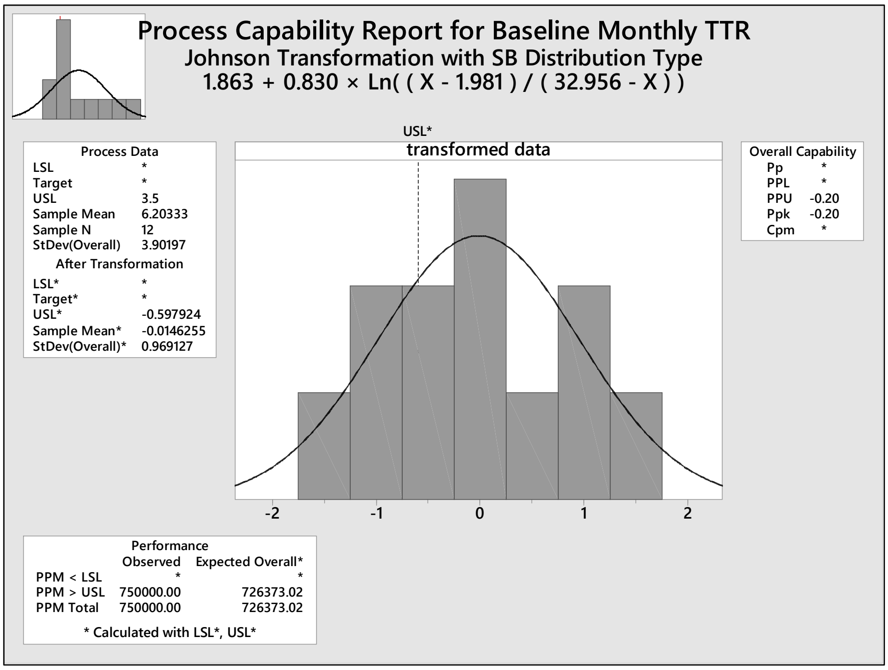

Figure 4: Process Capability Analysis of Baseline Monthly TTR

Such level of process capability is way below the desirable status for the management. Detailed reliability analysis and micro-level investigations were planned which are part of the Analyse phase.

2.3. Analyse Phase

Following sequence of analysis was adopted to investigate the problem of breakdown, time to repair, and availability issue of RMCS and this sequence is typical of tollgate steps of Analyse phase of Six Sigma methodology.

- Problem analysis and for this study in-depth reliability and availability analysis to ascertain the predictability of the system status
- Identification of all potential causes of the problem
- Confirm real causes or root causes or high-risk causes.

Data on time between failures, time to repair, percentage availability were collected for each breakdown instances and is already shown in Table 2. The data thus obtained were tested for appropriate statistical distributions.

Lognormal distribution turned out to be the best for the TBF data set with the highest p-value. Maximum likelihood estimates of the distribution parameters are as under:

Distribution	Location parameter	Scale parameter
Lognormal	5.51995	0.49020

Reliability estimates of the data using lognormal distribution was tried and the estimates and the survival plot is made and shown in Table 3 and Figure 5 respectively.

Table 3: Reliability Estimates of the RMCS

					95% Confidence Interval	
Percent	Proportion	Reliability	TBF-Baseline	Standard error	Lower	Upper
1	0.01	0.99	81.30	13.56	58.64	112.72
2	0.02	0.98	92.72	14.16	68.74	125.07
3	0.03	0.97	100.78	14.52	75.98	133.68
4	0.04	0.96	107.31	14.79	81.91	140.59
5	0.05	0.95	112.93	15.00	87.04	146.52
6	0.06	0.94	117.94	15.18	91.64	151.79
7	0.07	0.93	122.52	15.34	95.86	156.59
8	0.08	0.92	126.77	15.48	99.79	161.05
9	0.09	0.91	130.77	15.61	103.49	165.23
10	0.10	0.90	134.55	15.73	107.00	169.19
20	0.20	0.80	166.35	16.77	136.53	202.68
30	0.30	0.70	193.85	17.91	161.75	232.32
40	0.40	0.60	220.92	19.44	185.92	262.50
50	0.50	0.50	249.62	21.62	210.65	295.81
60	0.60	0.40	282.06	24.82	237.38	335.15
70	0.70	0.30	321.45	29.69	268.21	385.24
80	0.80	0.20	374.58	37.75	307.43	456.39
90	0.90	0.10	463.10	54.13	368.29	582.33
91	0.91	0.09	476.51	56.87	377.13	602.09
92	0.92	0.08	491.52	60.01	386.92	624.41
93	0.93	0.07	508.58	63.66	397.93	649.99
94	0.94	0.06	528.32	68.00	410.52	679.92

95	0.95	0.05	551.77	73.30	425.29	715.88
96	0.96	0.04	580.67	80.03	443.21	760.76
97	0.97	0.03	618.26	89.10	466.13	820.06
98	0.98	0.02	672.04	102.63	498.20	906.54
99	0.99	0.01	766.45	127.79	552.79	1062.69

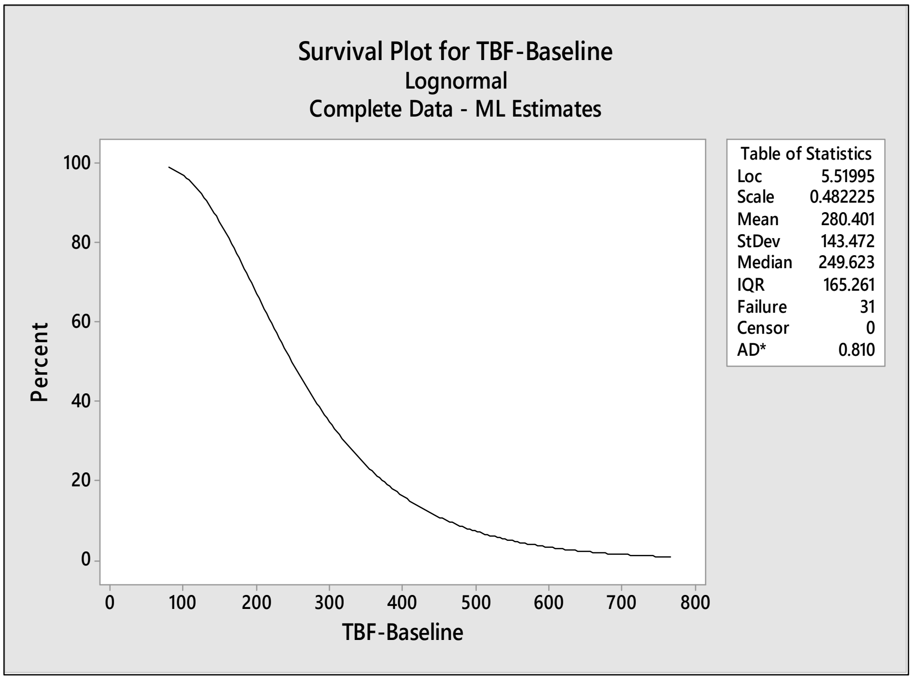

Figure 5: Survival Plot of RMCS at Different Time Points

Poor reliability is reflected throughout the baseline duration of the study.

Availability summaries estimated are as under:

Variable	No. of data points	Mean	SE Mean	St Dev	Minimum	Q1	Median	Q3	Maximum
Availability	31	98.93	0.135	0.749	97.053	98.509	99.274	99.492	99.67

TBF, TTR are predictably with poor performance status and as a consequence availability of RMCS suffers with 98.93% availability.

Macro level process map, available in Figure 6, is detailed out in terms of a deployment flow chart or swim lane chart to make it amenable to identify potential causes for the problem and this chart is shown in Figure 6.

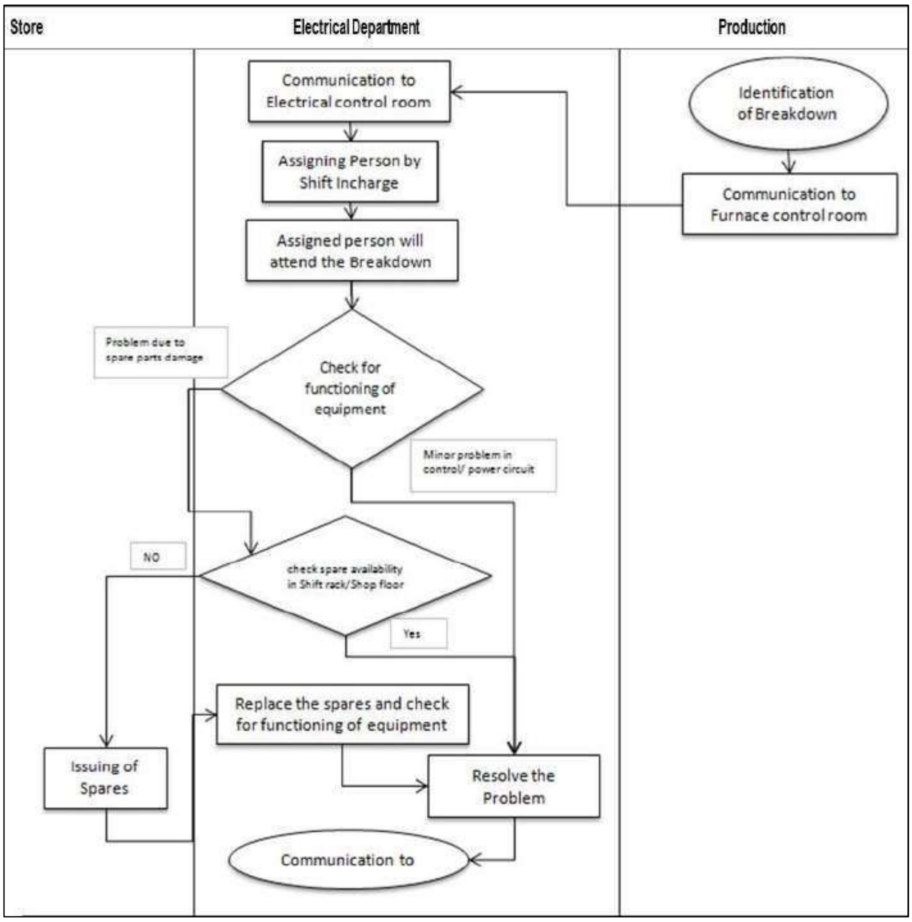

Figure 6: Deployment Flow Chart of Maintenance Process

Consultations with the maintenance crew and the production engineers using SIPOC and the Deployment flow chart arrived at five major drivers to the problems.

They are vibration feeder, programmable logic controller (PLC) and automation, Skip, Conveyor System and Telfer. Structured brainstorming with the engineering team detailed out with the potential causes of the reliability and availability problems and the same are available in a Cause-and-Effect Diagram in Figure 7.

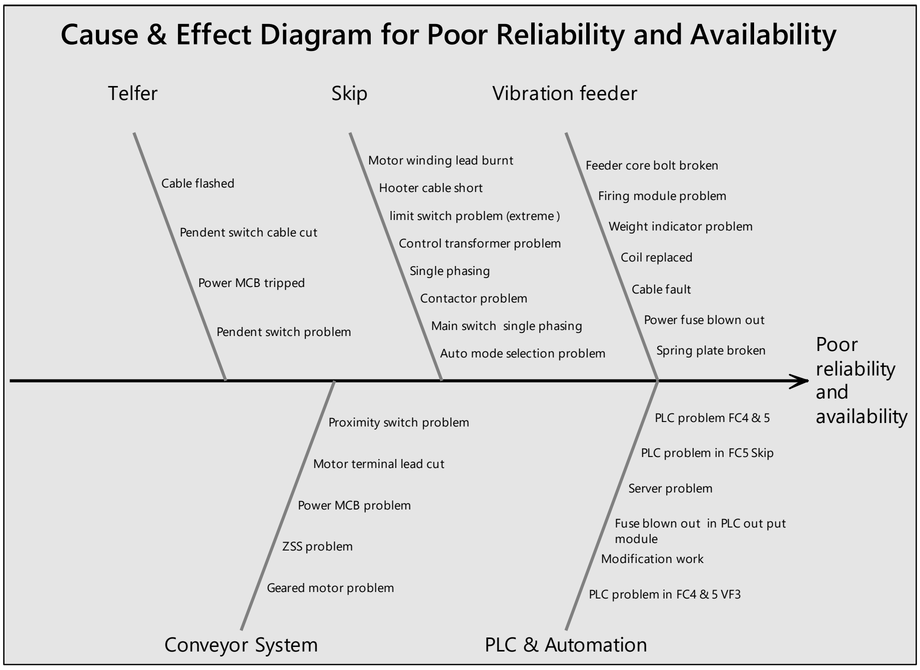

Figure 7: Cause and Effect Diagram for poor reliability and availability

Duration of breakdowns was noted for each cause for each contributing factor and Pareto analyses were made for each of the driver with respect to the breakdown duration to isolate vital few causes from trivial many and is available in Table 4.

Grey shaded causes in Table 4 were considered for further root cause analysis using Why-Why Analysis and is presented in Table 5.

Study now goes to improvement phase for acting on the identified root causes.

2.4. Improve Phase

Solutioning team with the specialist engineers were formed to evolve solutions on the identified root causes. The team came out with the solutions with a comprehensive implementation plan. The action plan as indicated in Table 6 is implemented.

Table 4: Pareto Analysis for Contribution Drivers

Contribution Drivers and Causes	Duration of breakdown	%ge Contribution	Cumulative %ge Contribution	Contribution Drivers and Causes	Duration of breakdown	%ge Contribution	Cumulative %ge Contribution
Vibration Feeder				**Conveyor System**			
Feeder core bolt broken	11.25	31.1	31.1	Geared motor problem	3.42	49.6	49.6
Firing module problem	7	19.4	50.5	ZSS problem	1	14.5	64.2
Weight indicator problem	3.17	8.8	59.3	Power MCB problem	0.84	12.2	76.3
Coil replaced	3	8.3	67.6	Motor terminal lead cut	0.58	8.4	84.8
Cable fault	2.58	7.1	74.7	Proximity switch problem	0.42	6.1	90.9
Power fuse blown out	1.91	5.3	80.0	Others	0.63	9.1	100.0
Spring plate broken	1.75	4.8	84.8	Total	6.89		
Others	5.49	15.2	100.0	**Telfer**			
Total	36.15			Cable flashed	3.25	48.1	48.1
PLC and Automation				Pendent switch cable cut	1.67	24.7	72.9
PLC problem in FC4 & 5 VF3	3.83	30.9	30.9	Power MCB tripped	1.08	16.0	88.9
Modification work	2.58	20.8	51.8	Pendent switch problem	0.42	6.2	95.1
Fuse blown out in PLC out put module	1.09	8.8	60.6	Others	0.33	4.9	100.0
Server problem	1	8.1	68.7	Total	6.75		
PLC problem in FC5 Skip	0.83	6.7	75.4				
PLC problem FC4 & 5	0.83	6.7	82.1				
Others	2.22	17.9	100.0				
Total	12.38						
Skip							
Motor winding lead burnt	2.83	23.1	23.1				
Hooter cable short	2.75	22.4	45.5				
limit switch problem (extreme)	2.17	17.7	63.2				
Control transformer problem	1.17	9.5	72.8				
Single phasing	0.67	5.5	78.2				
Contactor problem	0.58	4.7	83.0				
Main switch single phasing	0.5	4.1	87.0				
Auto mode selection problem	0.42	3.4	90.5				
Others	1.17	9.5	100.0				
Total	12.26						

Table 5: Why-Why Analysis

Contribution Drivers and Respective Vital Causes	Why 1	Why 2	Why 3
Vib Feeder			
Feeder core bolt broken	Feeder is running more than design capacity 50- 60 tph v/s 30 tph design capacity	Because we have increased the furnace capacity	
Firing module problem	internal transformer blown out	Feeder is running beyond capacity	Because we have increased the furnace capacity
Weight indicator problem	Due to operating in high ambient temp (more than required temp)	Due to lack of awareness	
Cable fault	Feeder cable insulation damage	Fedder cable came in contact with structure during vibration.	No extra insulation provision
Spring plate broken	Feeder is running more than design capacity 50- 60 tph v/s 30 tph design capacity	Because we have increased the furnace capacity	
PLC and Automation			
PLC problem in FC4 & 5 VF3	All feedback of the charging system drives are not shown in scada pages	During design phase not incorporated in the system	
Modification work	Shutdown not incorporated in Business Plan	Not communicated	
Skip			
Motor winding lead burnt	Due to overload operation	Overload potection not operating.	Slow response time thermal overload relay
Hooter cable short	Cable got burnt	Directly affected by Chimeny hot temp.	
limit switch problem (extreme)	Limit switch displaced/Stuckup	No schedule for limit switch condition checking	Limit switch not incorporated in Preventive maintenance
Conveyor System			
Geared motor problem	Due to motor insulation failed	Due to motor end shield problem	Aging Problem
Telfer			
Cable flashed	Normal PVC Cable used	Due to poor operating condition	
Pendent switch problem	Scheduled replacement of Pendent was not done	Std. notdefined forreplacement	

Table 6: Improvement Plan and Actions

S. No.	Problem	Potential Cause(s)	Root Cause	Intermediate Action	Person Responsible	Permanent Action	Person Responsible	Target Month of Completion	Status
Vibration Feeder									
1	Feeder core bolt broken	Feeder is running more than design capacity 50-60 tph v/s 30 tph design capacity	Because we have increased the furnace capacity	Feeder is to be cleaned shift wise and regular monitoring	FC4 and FC5 Shift Incharge	Electromagnetic vibrating feeder replaced with electro motorized feeder	BB*	Month 14	Completed
2	Firing module problem	Internal transformer blown out	Because we have increased the furnace capacity	Spare firing module must be kept with Shift Crew.	Electrical shift Incharge	Electromagnetic vibrating feeder replaced with electro motorized feeder	BB	Month 14	Completed
3	Weight indicator problem	Due to operating in high ambient temp (more than required temp)	Due to lack of awareness	Awareness training to control room operator	All Furnace Incharge	Cooling fan installed in weight indicator	YB**	Month 14	Completed
5	Cable fault	Feeder cable insulation damage	No extra insulation provision	Cable repaired	Electrical shift Incharge	Steel Wire braided PVC flexible conduit provided for covering the cable for extra protection.	BB	Month 15	Completed
6	Spring plate broken	Feeder is running more than design capacity 50-60 tph v/s 30 tph design capacity	Because we have increased the furnace capacity	Spring plate replaced	Electrical shift Incharge	Electromagnetic vibrating feeder replaced with electro motorized feeder	Champion	Month 14	Completed
PLC and Automation									
7	PLC problem in FC4 & 5 VF3	All feedback of the charging system drives are not shown in scada pages	During design phase not incorporated in the system	Not Applicable		Design upgradation of Scada Pages	YB	Month 15	Completed
8	Modification work	Shutdown not incorporated in Business Plan	Not communicated	Not Applicable		Now started incorporating modification work in business plan/Communicate with operation team.	YB	Ongoing	Completed

9	Motor winding lead burnt	Due to overload operation	Slow response time thermal overload relay	Motor & Overload relay replaced	Electrical Shift Incharge	Advance motor protection relay to be installed	YB	Month 14	Completed
10	Hooter cable short	Cable got burnt	Directly affected by Chimney hot temp.	Hooter cable replaced	Electrical Shift Incharge	Hooter cable route to be changed	YB	Month 15	Completed
11	limit switch problem (extreme)	Limit switch displaced/Stuckup	Limit switch not incorporated in Preventive maintenance	Limit switch fixed at it's position	Electrical Shift Incharge	1. Limit switch has been incorporated in Preventive maintenance schedule.	YB	1. Month 13	Completed
						2. Another rotary limit switch provided for extra protection & reducing breakdown	YB	2. Month 15	Completed
Conveyor System									
12	Geared motor problem	Due to motor insulation failed	Aging Problem			Phase wise motor replacement for which aging is more than 5 years	YB	Month 15	Completed
Telfer									
13	Cable flashed	Normal PVC Cable used	Due to poor operating condition	Cable repaired	Electrical Shift Incharge	EPR cable Started using	BB	Month 14	Completed
14	Pendent switch problem	Scheduled replacement of Pendent was not done	Std. not defined for replacement	Pendent repaired	Electrical Shift Incharge	Frequency of Pendent replacement to be standardised	BB	Month 15	Completed

*BB: Black-Belt, **YB- Yellow Belts

Improvement in repair time was observed as the actions were being taken during Analyse and Improvement phase and is evident in Figure 8.

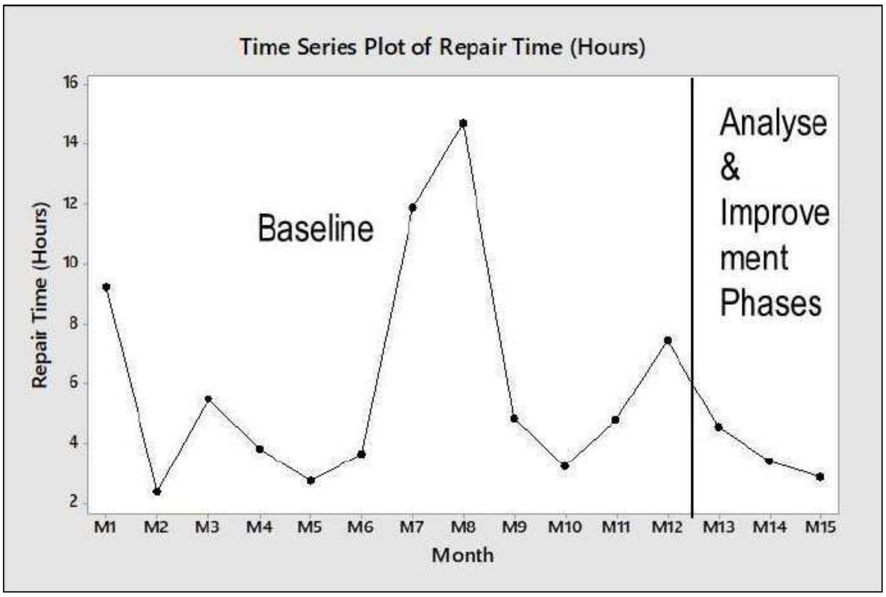

Figure 8: Time Series Plot of Repair Time for Baseline vs. Analyse and Improve Phases

Analyse and Improvement Phase took three months of time to complete and during these stages even when actions were being implemented, monthly repair time showed declining trend as evident from Figure 6. Average repair time came down to 4.8 hours per month. However, real impact of the actions of impact of improve phase is expected to show stabilised improved status in the control phase. Improved actions encouraged the team to proceed towards control phase by tightening the loose ends and relevant changes in the processes.

2.5 Control Phase

The team felt necessary to observe the effect of the improvement actions for 5 months. Control plans were in place. Control phase starts after the completion of all the actions in the improvement phase. Analyse and improvement phase data are expected to be vulnerable as a few experiments takes place during this stage. That is why data during analyse and improvement phases are usually not considered for comparison against baseline. Data of repair and TBF were collected and are shown in Table 7.

Table 7: Data of TTR, TBF and Availability during Analyse, Improve and Control Phases

Period	Month	TTR-monthly	No. of Incidences of Breakdown	Instances of Breakdown	Repair time	TBF	Availa-bility
	Unit of measure	Hours	No.		Hours	Hours	%
	Target	<= 3.5					
Analyse and Improve Phase	M13	4.57	2	33	2.15	398.22	99.46
				34	2.42	291.33	99.18
	M14	3.42	3	35	1.05	271.23	99.61
				36	1.10	232.18	99.53
				37	1.27	234.82	99.46
	M15	2.90	2	38	1.12	284.16	99.61
				39	1.78	272.18	99.35
Control Phase	M16	2.17	1	40	2.17	563.06	99.62
	M17	2.72	2	41	1.3	607.08	99.79
				42	1.42	410.75	99.66
	M18	1.96	1	43	1.96	612.11	99.68
	M19	2.44	1	44	2.44	651.51	99.63
	M20	2.26	1	45	2.66	592.09	99.55

Monthly TTR data at control stage, were verified if the improvements have been significant over base line period and against the target of 3.5 hours.

Mann Whitney Test were done to test for significance of control phase monthly repair status over its baseline status and the test results are made available as under:

Mann-Whitney Test and CI: Monthly TTR - Baseline, Monthly TTR - Control

Number Median

Monthly TTR – Baseline 12 4.825

Monthly TTR – Control 5 2.260

Point estimate for η1 - η2 is 2.565

96.0 Percent CI for η1 - η2 is (0.800,7.290)

W = 136.0

Test of η1 = η2 vs η1 > η2 is significant at 0.0019

Thus, with p-value at 0.0019, it is confirmed that the monthly TTR of 2.26 hours is significantly lower to that of baseline status.

Monthly TTR data at control phase follows normal distribution with p-value at 0.883. One sample t-test was done to test the significance of the average at control phase was significantly lower to the target of monthly 3.5 hours or not and the test summary is as under:

One-Sample T: Monthly TTR - Control

Test of $\mu = 3.5$ vs < 3.5

Variable	No.	Mean	St Dev	SE Mean	95% Upper Bound	t-value	p-value
Monthly TTR-Control	5	2.310	0.287	0.128	2.584	-9.27	<0.001

Thus, it is concluded with p-value at <0.001 that the average monthly TTR have been significantly brought below the target value.

Process capability analysis on monthly time to repair data at control phase (after improvement) were done and the analysis is available in Figure 9.

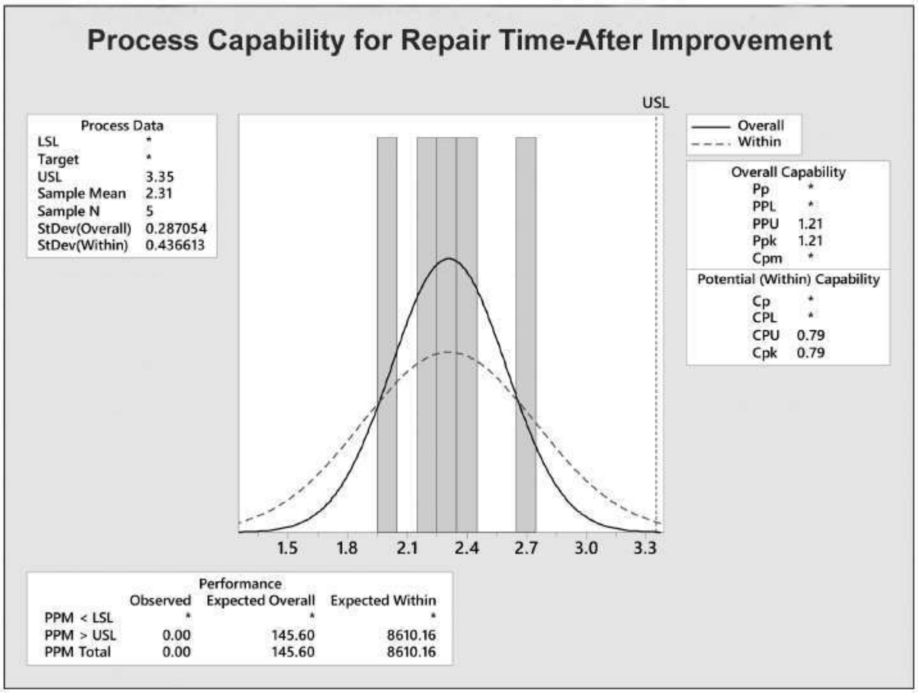

Figure 9: Process Capability Analysis at Control Phase

It is evident that drastic improvement in monthly repair time against the target of 3.5 hours maximum per month showed decreased dpmo of 145.6 which is translated to 5.12 sigma level in the short term and 3.62 in the long term.

I-MR chart was made over phase to understand the level of improvement of monthly repair time and is shown in Figure 10.

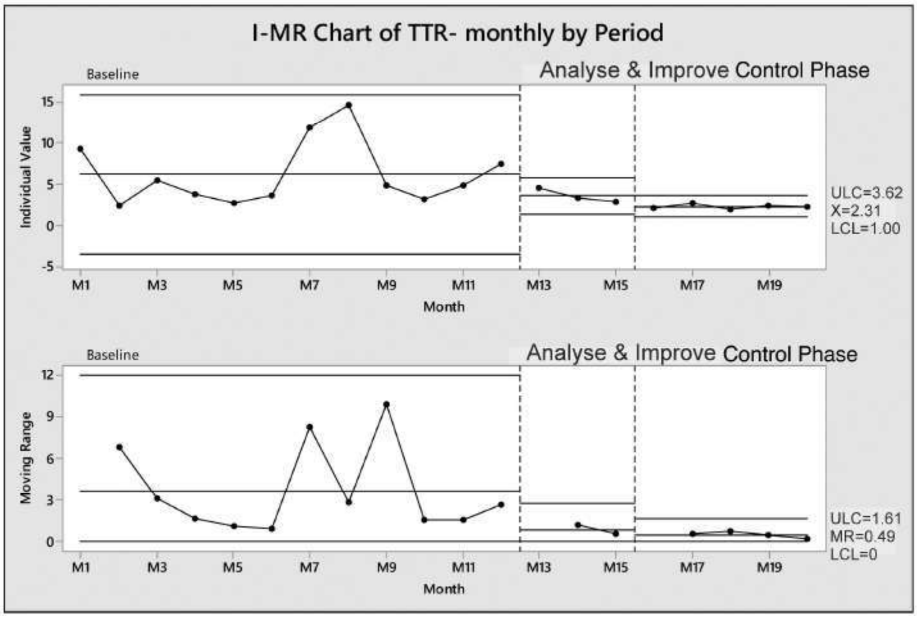

Figure 10: Control Chart over phases

Control phase also dealt with assessing all other reliability and availability parameters as was done in the measure phase to assess the extent of improvement from its respective baseline status.

First TBF data were considered for identifying which probability distribution fits the most and is available in Table 8.

Table 8: Individual Probability Distribution of TBF-Control Phase

Probability Distribution	Anderson Darling Test Statistic	p-Value
Normal	0.608	0.06
Lognormal	0.726	0.027
3-Parameter Lognormal	0.644	*
Exponential	2.151	0.004
2-Parameter Exponential	1.206	0.011
Weibull	0.478	0.21

3-Parameter Weibull	0.377	0.293
Smallest Extreme Value	0.377	>0.250
Largest Extreme Value	0.81	0.025
Gamma	0.728	0.06
3-Parameter Gamma	4.455	*
Logistic	0.515	0.128
Loglogistic	0.625	0.055
3-Parameter Loglogistic	0.516	*

Three-Parameter Weibull Distribution fitted the best with the highest p-value at 0.293. Reliability estimates are made using Three Parameter Weibull Distribution over different time points and is available in Table 9.

Table 9: Reliability and Availability at Control Phase

Percent	Proportion	Reliability	TBF-Control	Std. Error	Lower	Upper
1	0.01	0.99	380.615	85.6491	212.746	548.484
2	0.02	0.98	414.656	74.2618	269.106	560.206
3	0.03	0.97	434.676	67.6235	302.136	567.216
4	0.04	0.96	448.956	62.9249	325.625	572.287
5	0.05	0.95	460.092	59.2873	343.891	576.293
6	0.06	0.94	469.239	56.3199	358.854	579.624
7	0.07	0.93	477.015	53.8146	371.541	582.49
8	0.08	0.92	483.789	51.6473	382.562	585.015
9	0.09	0.91	489.796	49.7381	392.311	587.281
10	0.1	0.90	495.2	48.0325	401.058	589.342
20	0.2	0.80	531.807	36.8934	459.498	604.117
30	0.3	0.70	554.69	30.5351	494.842	614.538
40	0.4	0.60	572.217	26.2418	520.784	623.649
50	0.5	0.50	587.11	23.2207	541.598	632.622
60	0.6	0.40	600.73	21.1889	559.2	642.259
70	0.7	0.30	614.056	20.0904	574.679	653.432
80	0.8	0.20	628.222	20.0497	588.925	667.519
90	0.9	0.10	645.703	21.5905	603.387	688.02
91	0.91	0.09	647.887	21.893	604.978	690.797
92	0.92	0.08	650.218	22.2397	606.629	693.807
93	0.93	0.07	652.733	22.6399	608.36	697.107
94	0.94	0.06	655.484	23.1073	610.194	700.773
95	0.95	0.05	658.549	23.6623	612.172	704.926
96	0.96	0.04	662.056	24.3385	614.353	709.759
97	0.97	0.03	666.235	25.1969	616.85	715.62
98	0.98	0.02	671.577	26.369	619.894	723.259
99	0.99	0.01	679.54	28.2508	624.17	734.911

Reliability shows sharp improvement at Control Phase and is reflected in the survival plot shown in Figure 11.

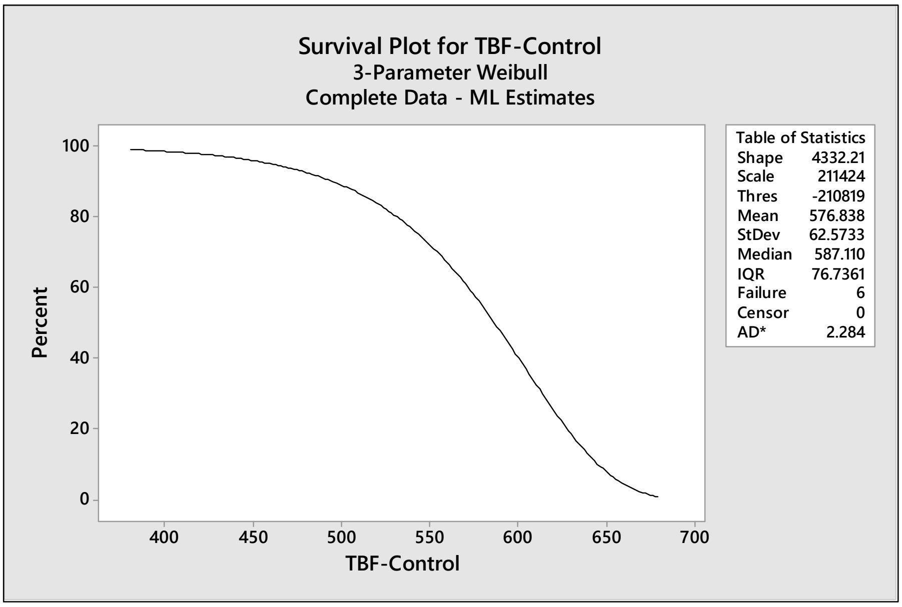

Figure 11: Survival Plot at Control Phase

Availability Summaries are also made with the Control phase data which is as under:

No. of Obser-vations	Mean	Standard Error of Mean	Standard deviation	Minimum	Q1	Median	Q3	Maximum
6	99.653	0.032	0.0783	99.553	99.6	99.641	99.707	99.786

3. CONCLUSION

Six Sigma methodology has been used to reduce the monthly downtime of RMCS to a good extent. While doing so, the reliability and availability of the system also have been comprehensively improved, as can be seen in Table 10.

All round development in maintenance process of RMCS could get all round improvement in the reliability and availability of the system using six sigma methodology. The company also gained an estimated annual recurring cost savings of 15 Lakhs INR in the process.

Table 10: Comparison of important maintenance and reliability parameters at baseline and after improvement

Sl. No.	Maintenance and Relaibility Parameters	Unit of Measure	Baseline Period	After Improvement
1	**Monthly Performance:**			
1.1	Average number of breakdown incidences per month	Number	2.7	1.2
1.2	Average time to repair	Hours	6.2	2.31
1.3	Standard deviation of time to repair	Hours	3.9	0.29
1.4	Defects per million opportunity	Unit free	726373.02	145.6
1.5	Sigma Level- Short-term		0.9	5.12
1.6	Sigma Level- Long term		-0.6	3.62
2	**Reliability and Availability**			
2.1	MTTR	Hours	2.33	1.96
2.2	MTBF	Hours	278.84	572.77
2.3	Reliability at 300 hours of uninterrupted run of RMCS	Probability	0.354228	0.998
2.4	Average Availability	Percentage	98.93	99.85

The approach can be adopted for any repairable mechanised system to improve reliability and availability of the system.

ACKNOWLEDGEMENT

The authors thank the reviewer for his comments, which helped to improve the presentation.

REFERENCES

Harry, Mikel J. (1989): The Nature of six sigma quality. Rolling Meadows, Illinois: Motorola University Press. p. 25. ISBN 978-1-56946-009-2.

Schroeder, Richard A. and Harry Mikel. (2000): *Six Sigma*. New York: Currency, c2000. ISBN 978-0-38549-437-3.

Breyfogle, Forrest W. (2003): *Implementing Six Sigma: Smarter Solutions using Statistical Methods*. Productivity Press.

Leitch, Roger D. (1995): *Reliability Analysis for Engineers*: *An Introduction*. Oxford University Press.

Pham, Hoang. (2003): *Handbook of Reliability Engineering*. Springer.

Kumar, U. D., Crocker, J., Chitra, T., and Saranga, H. (2006): *Reliability and six sigma*. Springer Science & Business Media.

Al-Mishari, S. T., and Suliman, S. (2008): Integrating Six-Sigma with other reliability improvement methods in equipment reliability and maintenance applications. *J. Qual. Maint. Eng.*, 14(1), 59-70.

Smith, Chad, Wood, Suzanne and Beauvais, Bradley. (2011): Thinking Lean: Implementing DMAIC Methods to Improve Efficiency within a Cystic Fibrosis Clinic, *Journal of Health Care Quality*, 33(2), 37-46.

Rath, Subrata and Chakraborty, Ashis Kumar. (2019): Improving Reliability of HTDC Compressor in a Petrochemical Industry using Six Sigma Methodologies, *IAPQR Transactions*, 43(2), 117-134. DOI-https://doi.org/10.32381/IAPQRT.2019.43.02.2

IAPQR Transactions
Vol. 47, No. 1-2, 2022–2023

INFERENCE AND PREDICTION OF ORDER STATISTICS BASED ON K-RECORD VALUES FROM A WEIBULL DISTRIBUTION

LAJI MURALEEDHARAN* AND MANOJ CHACKO

Department of Statistics, University of Kerala,

Trivandrum, India

Abstract: In this paper, the problem of estimation of parameters and prediction of order statistics from a future sample based on upper *k*-record values arising from a two parameter Weibull distribution is considered. The maximum likelihood estimators and the Bayes estimators are obtained based on observed upper *k*-record values. Bayesian point predictors of the order statistics from a future sample and its two-sided equi-tailed prediction intervals are also derived. Finally, a numerical illustration is considered for exemplifying the proposed inferential procedures developed in this paper.

Keyword and Phrases: Weibull distribution, Order statistics, Upper *k*-records, Bayesian estimation, Bayesian prediction.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n arising from a population with absolutely continuous cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. By arranging the random sample in a non-decreasing order of magnitude as $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$, the order statistics of the sample can be obtained. The rth order statistic of the sample $X_1, X_2, \ldots, X_n$ is then $X_{r:n}$. Order statistics have wide range of applications in many fields including industry, reliability analysis and material strength. For more

* Correspondence author email: lajikmr@gmail.com
DOI-https://doi.org/10.32381/IAPQRT.2023.47.01-02.5

discussion regarding the order statistics, one may refer to Arnold et al. (1992) and David and Nagaraja (2003). One major application of order statistics in the study of reliability of systems is the following. A system is called a $k - out - of - m$ system if it consists of m components and the system functions satisfactorily if at least k ($\leq m$) components function. If the lifetimes of the components are independently distributed, then the lifetime of the system coincides with that of the $(m - k + 1)th$ order statistic of the lifetime of the components. Thus, order statistics play a key role in studying the lifetimes of such systems.

The cdf of the *rth* order statistic $X_{r:n}$ based on a random sample of size n from an absolutely continuous population with cdf $F(x)$ and pdf $f(x)$ is given by (see, Arnold et al., 1992).

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} [F(x)]^{i} [1 - F(x)]^{n-i}, -\infty < x < \infty. \quad (1)$$

The pdf corresponding to the cdf (1) is given b

$$f_{r:n}(x) = \frac{1}{B(r,n-r+1)} [F(x)]^{r-1} [1 - F(x)]^{n-r} \mathrm{f}(\mathrm{x}), -\infty < x < \infty, \quad (2)$$

where $B(r,n\text{-}r+1)$ denotes the complete beta function.

Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed (iid) random variables having an absolutely continuous cdf $F(x)$ and pdf $f(x)$. An observation X_j is called an upper record value if it exceeds all of its previous observations. That is, $X_j > X_i$, for every $i < j$. Thus, X_1 is the first upper record value. An analogous definition deals with lower record values. Many authors have studied the record values of iid random variables as well as their features in the literature. Arnold et al. (1998), Ahsanullah (1995) and the literature referenced therein can be used to have a more in-depth look at this topic.

Since Chandler (1952) brought up the idea of record values for the first time in the literature, there has been a significant growth in the study of record values. Record values have many statistical applications, such as modelling and inference involving data pertaining to mining, sports, industry, seismology, life testing and so on. Interested Surveys are given in Glick (1978), Gulati and Padgett (1994), Ahsanullah (1995), Arnold et al. (1998), Nagaraja (1988) and the literature cited therein.

One of the challenges in dealing with problems involving inference with record data is that the expected waiting time for consecutive records after the first is infinite. Such an issue does not arise if we use the k-records proposed by Dziubdziela and Kopocinski (1976). We use the following formal definition of *k*-record value given by Arnold et al., (1998).

For a fixed positive integer k, the upper k-record times $\tau_{n(k)}$ and the upper k-record values $R_{n(k)}$ are defined as follows.

Define $\tau_{1(k)} = k$ and $R_{1(k)} = X_{1:k}$ then for $n>1$,

$$\tau_{n(k)} = \min\left\{i;\ i > \tau_{n-1(k)}, X_i > X_{\tau_{n-1(k)}-k+1:\tau_{n-1(k)}}\right\}.$$

Then the sequence of upper k- record values $\left\{R_{n(k)}, n \geq 1\right\}$ is defined as

$$R_{n(k)} = X_{\tau_{n(k)}-k+1:\tau_{n(k)}}. \qquad (3)$$

In an analogous way, one can define the lower k-record values. The classical record values are contained in the k-record values as a special case when k=1.

There are two types of k-record values in the literature, called Type 1 k-record values and Type 2 k-record values. Arnold et al. (1998) refers (3) as Type 2 k-record sequence. The integer parameter k involved in the upper k-record value can be chosen in such a manner that the record data generated discard away the specified number of outliers which are feared to be crewed into the data. For example, if some initial scrutiny of the data revels that there is a possibility of occurrence of only one outlier in terms of its largeness in the data and the if general interest is with upper record values, then it is enough to consider upper *2*- record values as the desirable record data that may be used for further analysis and storage of it for future purposes. Inventing more and more characterization results based on the distributional properties of the statistics arising from a distribution makes the model mathematically tractable for developing statistical methods to analyze the data arising from it. Arnold et al. (1998) stated that all distributional properties of k-records arising from the cdf $F(x)$ can be studied from those of classical records arising from $1-(1-F)^k$.

The pdf of the nth upper k- record value $R_{n(k)}$, for $n \geq 1$ is given by

$$f_{n(k)}(x) = \frac{k^n}{(n-1)!}\{-\log[1-F(x)]\}^{n-1}[1-F(x)]^{k-1}f(x), \quad -\infty < x < \infty. \qquad (4)$$

Since $\left\{R_{n(k)}, n \geq 1\right\}$ is identical in distribution to $\{R_n, n \geq 1\}$ of upper record values arising from a population with cdf $1-(1-F)^k$, then the joint pdf of the first n upper k-record values $R_{1(k)}, R_{2(k)}, \ldots, R_{n(k)}$ is given by (see, Arnold et al. 1998).

$$f(r_1, r_2, \ldots, r_n) = k^n[1-F(r_n)]^k \prod_{i=1}^{n} \frac{f(r_i)}{1-F(r_i)}, -\infty < r_1 < r_2 < \ldots < r_n < \infty. \qquad (5)$$

Recently, the *k*-records data has shown an increased trend in problems involving statistical inference and future event prediction. Chacko and Muraleedharan (2018) obtained the Bayesian and maximum likelihood estimators (MLE) for the parameters of a generalized exponential distribution based on lower *k*-record values. The same problem has discussed by Muraleedharan and Chacko (2019) for a Gompertz distribution. The recurrence relation for the single and product moments of Gompertz distribution and its characterization based on *k*-records were studied by Minimol and Thomas (2014). The Bayesian estimation of the parameters for a Gumbel distribution and the prediction of future *k*-record values under the Bayesian frame work were studied by Malinowska and Szynal (2004). The best linear unbiased predictor (BLUP) for future *k*-record value based on *k*-records arising from a normal distribution was discussed by Chacko and Mary (2013) whereas the same problem for a generalized Pareto distribution was discussed by Muraleedharan and Chacko (2022). Paul and Thomas (2015) established some properties of upper *k*-record values which characterize the Weibull distribution and derived the BLUP for the model. Deheuvels and Nevzorov (1994) studied the limiting behavior of *k*-record values such as strong laws of large numbers, central limit theorems, functional laws of the iterated logarithm and strong invariance principles.

In statistical inference, predicting future events based on the current knowledge is a fundamental problem. It can be expressed in a variety of ways and in various settings. There are two different sorts of prediction problems. The one sample prediction problem is that the event to be predicted comes from the same sequence of events, whereas the two-sample prediction problem is when the event to be predicted comes from a different independent sequence of events. The two-sample prediction problem is the prime focus of this paper.

Prediction of order statistics from a future sample based on the observed *k*-records became a problem of considerable attention in the literature. Ahmadi et al. (2011) considered the two-sample prediction of future order statistics based on the observed lower k-record values arising from an exponential distribution under the Bayesian frame work. Recently, Vidovic (2019) used the same type of problem for a single parameter (shape) generalized exponential distribution. In these two works, the authors are considered the assumption that the future sample size is fixed. But Shafy et al. (2017) considered the prediction of future order statistics by using the assumption that the future sample size is random based on the observed upper *k*-records arising from a Pareto distribution.

In this paper, we discuss the problem of estimation of parameters and the prediction of order statistics from a future sample based on observed upper *k*-

records under the Bayesian approach, where the k-records are arising from a Weibull distribution with pdf and cdf are respectively given by

$$f(y) = \alpha\beta y^{\beta-1} exp\left(-\alpha y^{\beta}\right),\ y>0,\ \alpha,\ \beta > 0 \tag{6}$$

and $$F(y) = 1\text{-}exp\left(-\alpha y^{\beta}\right),\ y>0,\ \alpha,\ \beta > 0. \tag{7}$$

The Weibull distribution is a well-known distribution that has widespread application for lifetime models. It has varieties of shapes and it has increasing and decreasing failure rates. Therefore, this model is used for many applications, for example in hydrology, industrial engineering, weather forecasting and insurance. Throughout this paper, we use the notation W (α, β) to denote the Weibull distribution with scale parameter α and shape parameter β.

The rest of the paper is organized as follows. In Section 2, we derive the maximum likelihood estimators of the parameters of W (α, β). In Section 3, we consider the Bayes estimation of α and β. In Section 4, we obtain the predictive distribution of order statistics from a future sample under the Bayesian frame work. The Bayesian point predictor of future order statistics based on upper k-records is considered in Section 5. The corresponding Bayesian predictive intervals are derived in Section 6. A numerical study is considered in Section 7 and finally, some concluding remarks are included in Section 8.

2. MAXIMUM LIKELIHOOD ESTIMATION

In this section, we obtain the MLE of the parameters of W (α, β) based on observed upper k-record values. Let $R_{i(k)}$, i = 1, 2..., n be the first n upper k-record values arising from W (α, β) with pdf and cdf are respectively given in (6) and (7). Let $\boldsymbol{R}_{n(k)} = (R_{1(k)}, R_{2(k)}, \ldots, R_{n(k)})$ and let $\boldsymbol{r}_{n(k)} = (r_1, r_2, \ldots, r_n)$ be the observed realization of $\boldsymbol{R}_{n(k)}$. Then by using (5), the likelihood function of (α, β) is obtained as given below.

$$L(\alpha, \beta \mid \mathbf{r}_{n(k)}) = (\alpha\beta)^n\ exp\left(-k\alpha {r_n}^{\beta}\right)\left[\eta(\boldsymbol{r}_{n(k)})\right]^{\beta-1}, \tag{8}$$

where $\eta(\boldsymbol{r}_{n(k)}) = \prod_{i=1}^{n} r_i$. The logarithm of the likelihood function is given by

$$\text{L}=\log L(\alpha, \beta \mid \mathbf{r}_{n(k)}) = \text{n} \log \alpha + \text{n} \log \beta - k\alpha {r_n}^{\beta} + (\beta\text{-}1) \log (\eta(\boldsymbol{r}_{n(k)}))\ .$$

Then, $$\frac{\partial \text{L}}{\partial \alpha} = \frac{n}{\alpha} - k{r_n}^{\beta}$$

and $$\frac{\partial L}{\partial \beta} = \frac{n}{\beta} - k\alpha {r_n}^{\beta} log\ r_n + \log (\eta(\boldsymbol{r}_{n(k)})).$$

Then the corresponding normal equations are obtained as

$$\alpha k r_n{}^{\beta} = n \tag{9}$$

and
$$\frac{n}{\beta} - k\alpha r_n{}^{\beta} \log r_n + \log(\eta(\boldsymbol{r_{n(k)}})) = 0. \tag{10}$$

Using (9), (10) can be expressed as

$$\frac{n}{\beta} - n\, log\, r_n + \log(\eta(\boldsymbol{r_{n(k)}})) = 0. \tag{11}$$

From (11), the MLE of β, denoted by $\hat{\beta}_{ML}$ can be obtained as

$$\hat{\beta}_{ML} = n\left[\log\left(\frac{r_n^n}{\eta(\boldsymbol{r_{n(k)}})}\right)\right]^{-1}. \tag{12}$$

By using (12), the MLE of α, denoted by $\hat{\alpha}_{ML}$ can be computed as

$$\hat{\alpha}_{ML} = n\left[k r_n{}^{\hat{\beta}_{ML}}\right]^{-1}. \tag{13}$$

3. Bayesian Estimation

In this section, we describe the Bayesian approach for the problem of estimation of parameters and define the structure of prior and posterior distributions. For the Bayesian estimation, we consider the squared error loss (SEL) function which is defined as

$$\mathrm{L}(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2, \tag{14}$$

where $\hat{\theta}$ is an estimate of the parameter θ. The Bayes estimator of θ under the SEL function is defined as

$$\hat{\theta}_{BS} = E_{\theta}(\theta), \tag{15}$$

provided $E_{\theta}(.)$ exists. We prefer this estimator for the Bayesian context over others as it provides smaller mean squared error (MSE) when compared to others which is achieved by small variance and small bias. This small MSE of an estimator result into a high probability that the estimator is too close to true value of the parameter by Chebyshev's inequality.

Let $\boldsymbol{R_{n(k)}} = (R_{1(k)}, R_{2(k)}, \ldots, R_{n(k)})$ be a vector of first n upper k- record values arising from W (α, β) and let $\boldsymbol{r_{n(k)}} = (r_1, r_2, \ldots, r_n)$ be the corresponding vector of observed realization of $\boldsymbol{R_{n(k)}}$. We take the bivariate prior distribution of (α, β) as considered by Nigm (1989) and A1-Hussaini (1999) and is given by

$$\pi(\alpha, \beta) \propto \alpha^{c+a-1}\beta^{a-1} exp[-\alpha(d + b\beta)],\ \beta > 0, \tag{16}$$

where *a, b, c* and *d* are hyper parameters which are assumed to be known and non-negative. Then the posterior density function of *(α, β)* given $\boldsymbol{r_{n(k)}}$ is defined as

$$\pi^*(\alpha,\beta|\, r_{n(k)}) = \frac{L(\alpha,\beta|\boldsymbol{r_{n(k)}})\pi\,(\alpha,\beta)}{\int_{\alpha=0}^{\infty}\int_{\beta=0}^{\infty} L(\alpha,\beta|\boldsymbol{r_{n(k)}})\pi\,(\alpha,\beta)d\beta d\alpha}. \tag{17}$$

By using (8) and (16), the posterior density function (17) can be obtained as

$$\pi^*(\alpha,\beta|\, \mathbf{r_{n(k)}}) = \frac{\alpha^{n+c+a-1}\beta^{n+a-1}exp\left[-\alpha(kr_n^{\beta}+d+b\beta)\right]\left[\eta(\boldsymbol{r_{n(k)}})\right]^{\beta-1}}{\Gamma(n+c+a)\phi_1(1,-1,a,c,1,1)}, \tag{18}$$

where $\Gamma(\cdot)$ denotes the complete gamma function and

$$\phi_s(z,p,a,c,i,j) = \int_{t=0}^{\infty}\frac{t^{n+a-1}\left[z^s\eta(\boldsymbol{r_{n(k)}})\right]^{t-1}}{\left[z^t(p+i-j+1)+kr_n^t+d+bt\right]^{n+c+a}}dt, s = 0,1,2,\ldots. \tag{19}$$

Then by using (15), the Bayesian estimator of α and β under the SEL function is the mean of the posterior density function (18) and are respectively obtained as

$$\hat{\alpha}_{BS} = (n+a+c)\frac{\phi_1(1,-1,a,c+1,1,1)}{\phi_1(1,-1,a,c,1,1)} \tag{20}$$

and $$\hat{\beta}_{BS} = \frac{\phi_1(1,-1,a+1,c-1,1,1)}{\phi_1(1,-1,a,c,1,1)}, \tag{21}$$

where ϕ_1 is defined in (19).

4. PREDICTIVE DENSITY FUNCTION OF THE FUTURE ORDER STATISTICS

Let $Y_1, Y_2, \ldots, Y_m$ be a future sample of size *m* from W (*α, β*) and let $Y_{1:m} \leq Y_{2:m} \leq \ldots \leq Y_{m:m}$ be the corresponding order statistics obtained from the future sample. Then by using (6) and (7) in (2), the pdf of $Y_{j:m}$ is obtained as

$$f_{j:m}(y|\alpha,\beta) = \frac{\alpha\beta y^{\beta-1}}{B(j,m-j+1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}exp\left[-\alpha y^{\beta}(m+i-j+1)\right],\ y > 0, \alpha, \beta > 0. \tag{22}$$

The Bayesian predictive density function of $Y_{j:m}$ given the observed k-records $\boldsymbol{r_{n(k)}}$ is obtained by integrating the product of the posterior density function (18) and the density function (22) with respect to α and β. Let $\hat{f}_{j:m}(y|\alpha,\beta)$ be the predictive density function of $Y_{j:m}$. Then

$$\hat{f}_{j:m}(y|\boldsymbol{r_{n(k)}}) = \int_{\alpha=0}^{\infty}\int_{\beta=0}^{\infty} f_{j:m}(y|\alpha,\beta)\pi^*(\alpha,\beta|\,\boldsymbol{r_{n(k)}})d\alpha d\beta$$

$$= \frac{\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\Gamma(n+c+a+1)}{B(j,m-j+1)\Gamma(n+c+a)\phi_1(1,-1,a,c,1,1)} \times \int_{\beta=0}^{\infty} \frac{\beta^{n+a}[y\eta(\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})})]^{\beta-1}}{\left[y^\beta(m+i-j+1)+kr_n^\beta+d+b\beta\right]^{n+c+a+1}} d\beta$$

$$= \frac{\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}(n+c+a+1)\phi_1(y,m,a-1,c,i,j)}{B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}$$

$$= \sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\frac{(n+c+a)}{B(j,m-j+1)}\frac{\phi_1(y,m,a+1,c,i,j)}{\phi_1(1,-1,a,c,1,1)}, \quad (23)$$

where ϕ_1 is defined in (19).

The predictive density function corresponds to the smallest and largest order statistics are respectively obtained as

$$\hat{f}_{1:m}(y|\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})}) = m(n+c+a)\left[\frac{\phi_1(y,m,a+1,c,0,1)}{\phi_1(1,-1,a,c,1,1)}\right] \quad (24)$$

and

$$\hat{f}_{m:m}(y|\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})}) = m(n+c+a)\sum_{i=0}^{m-1}(-1)^i\binom{m-1}{i}\left[\frac{\phi_1(y,m,a+1,c,i,m)}{\phi_1(1,-1,a,c,1,1)}\right], \quad (25)$$

where ϕ_1 is defined in (19).

5. Bayesian Point Prediction of Future Order Statistics

In this section, we discuss the Bayesian point prediction of order statistics from a future sample based on observed upper *k*-record values under SEL function. The Bayesian point predictor of $Y_{j:m}$, under the SEL function is the mean of the predictive density function (23), which is obtained as given below.

$$\hat{Y}_{j:m} = E_{\hat{f}_{j:m}}[Y_{j:m}|\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})}]$$

$$= \frac{(n+c+a)}{B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}\int_{y=0}^{\infty} y\,\phi_1(y,m,a+1,c,i,j)\mathrm{d}y \quad (26)$$

Using (19) in (26) and integrating with respect to *y*, we obtain $\hat{Y}_{j:m}$ as given below

$$\hat{Y}_{j:m} = \frac{(n+c+a)}{B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}\sum_{i=0}^{j-1}(-1)^i\binom{j-1}{i}$$

$$\times \int_{\beta=0}^{\infty} \frac{\beta^{n+a-1}[\eta(\boldsymbol{r}_{\boldsymbol{n}(k)})]^{\beta-1}B(\frac{1}{\beta}+1,n+a+c-\frac{1}{\beta})}{(m+i-j+1)^{\frac{1}{\beta}+1}\left[kr_n^\beta+d+b\beta\right]^{n+a+c-\frac{1}{\beta}}} d\beta. \quad (27)$$

If we denote

$$\psi(m,a,c,i,j) = \int_{\alpha=0}^{\infty} \frac{\beta^{n+a-1}\left[\eta(\boldsymbol{r_{n(k)}})\right]^{\beta-1} B(\frac{1}{\beta}+1, n+a+c-\frac{1}{\beta})}{(m+i-j+1)^{\frac{1}{\beta}+1}\left[kr_n^{\beta}+d+b\beta\right]^{n+a+c-\frac{1}{\beta}}} d\beta, \tag{28}$$

then the Bayesian point predictor of $Y_{j:m}$ can be expressed as

$$\hat{Y}_{j:m} = \sum_{i=0}^{j-1}(-1)^i \binom{j-1}{i} \left[\frac{(n+c+a)\psi(m,a,c,i,j)}{B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}\right]. \tag{29}$$

For the special cases $j = 1$ and $j = m$, (29) reduces to the point predictor for smallest and largest future order statistics, which are given by

$$\hat{Y}_{1:m} = \frac{m(n+c+a)\psi(m,a,c,0,1)}{\phi_1(1,-1,a,c,1,1)} \tag{30}$$

and $$\hat{Y}_{m:m} = \sum_{i=0}^{m-1}(-1)^i \binom{m-1}{i} \frac{m(n+c+a)\psi(0,a,c,i,0)}{\phi_1(1,-1,a,c,1,1)}. \tag{31}$$

6. BAYESIAN INTERVALS OF ORDER STATISTICS FROM A FUTURE SAMPLE

In this section, we construct the $100(1-\gamma)$ % prediction intervals for $Y_{j:m}$. The Bayesian predictive survival function of $Y_{j:m}$ from a future sample of size m based on observed upper k- record values $\boldsymbol{r_{n(k)}}$ is given by

$$\hat{\bar{F}}_{j:m}(z|\boldsymbol{r_{n(k)}}) = \int_{y=z}^{\infty} \hat{f}_{j:m}(y|\boldsymbol{r_{n(k)}})dy$$

$$= \frac{(n+c+a)}{B(j,m-j+1)\phi_1(1,-1,a,c,1,1)} \sum_{i=0}^{j-1}(-1)^i \binom{j-1}{i}$$

$$\times \int_{y=z}^{\infty}\int_{\beta=0}^{\infty} \frac{\beta^{n+a}\left[y\eta(\boldsymbol{r_{n(k)}})\right]^{\beta-1}}{\left[y^{\beta}(m+i-j+1)+kr_n^{\beta}+d+b\beta)\right]^{n+c+a+1}} dy d\beta$$

$$= \sum_{i=0}^{j-1}(-1)^i \binom{j-1}{i} \frac{1}{(m-j+i+1)B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}$$

$$\times \int_{\beta=0}^{\infty} \frac{\beta^{n+a}\left[\eta(\boldsymbol{r_{n(k)}})\right]^{\beta-1}}{\left[z^{\beta}(m+i-j+1)+kr_n^{\beta}+d+b\beta)\right]^{n+c+a}} d\beta$$

$$= \sum_{i=0}^{j-1}(-1)^i \binom{j-1}{i} \frac{\phi_0(z,m,a,c,i,j)}{(m-j+i+1)B(j,m-j+1)\phi_1(1,-1,a,c,1,1)}. \tag{32}$$

where ϕ_0 and ϕ_1 are defined in (19).

Now the two-sided equi-tailed interval for $Y_{j:m}$ is defined as *(L, U)*, where *L* and *U* are obtained by solving the following non-linear equations.

$$\hat{\hat{F}}_{j:m}(L|\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})}) = 1 - \frac{\gamma}{2} \text{ and } \hat{\hat{F}}_{j:m}(U|\boldsymbol{r}_{\boldsymbol{n}(\boldsymbol{k})}) = \frac{\gamma}{2}. \quad (33)$$

Since it is not possible to obtain the solutions analytically, we need to apply a suitable numerical method for solving the equations given in (33).

7. Numerical Illustrations

Example 1: *Simulated Data*

In this section, we carry out a simulation study for illustrating the estimation procedures developed in the previous sections. The numerical procedures were conducted by using the R Programme. We have obtained the bias and MSE of the MLE and Bayes estimators of α and β for different values of *n* based on *k*-records under different combinations of (α, β). Table 1 demonstrates the bias and MSE of the MLEs and Table 2 provides that for the Bayes estimator under the SEL function for $k = 2$ and $k = 3$. We choose the hyper parameters as (*a, b*): (2.5, 2.5) and (*c, d*) : (2.5, 2.5) for the Bayesian estimation. From the tables, one can observe that the MSE decreases when the sample size increases as one would expect. Also, from the Tables 1 and 2, one can see that the bias and MSE of Bayes estimators are smaller than bias and MSE of MLEs, therefore we conclude that Bayes estimators perform better than MLE in terms of bias and MSE. For the simulation studies for Bayes point predictors as well as the two-sided equi-tailed Bayesian prediction intervals, we take (α, β) = (1.5, 2) and the hyper parameters for the prior distributions of α and β as (*a, b*): (2.5, 2.5), (2, 2), (1, 1) and for (*c, d*) : (2.5, 2.5), (2, 2), (1, 1), respectively. With these setting, we have obtained the point predictors of future order statistics, 95% Bayesian two sided equi-tailed prediction intervals *(L, U)* and average interval length (AIL) along with the mean square predictive error (MSPE) from a future sample of size $m = 5, 7, 9$ based on upper *k*-records, for $k = 2$ and $k = 3$ with *n* is fixed as 10 and the values are presented in Tables 3 to 5, respectively. The predictive intervals with shortest width are attained at the least value of k. The results in these tables reveal that the width of the Bayesian intervals decreases as the sample size m increases. Moreover, the prediction intervals get shorter as the hyper parameters (*a, b*) and (*c, d*) decreases.

Example 2: *Real-life Data*

In this section, we illustrate the inferential procedures developed in the previous sections using the following data set (see, Nelson, 1982) representing the time to failure of certain 19 electrical insulating fluid at constant voltage 34KV has a Weibull distribution.

0.96, 4.15, 1.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89.

A method suggested by engineering considerations is that for a fixed voltage level, time to break has a Weibull distribution (see, Zakerzadeh and Jafari, 2015). The *k*- records $R_{n(k)}$ (k = 2, 3) extracted from the above data set are as follows.

n	1	2	3	4	5	6	7	8	9
$R_{n(2)}$	0.96	1.19	4.15	8.01	8.27	31.75	32.52	33.91	36.71
$R_{n(3)}$	0.96	1.19	4.15	7.35	8.01	8.27	31.75	32.52	33.91

For this illustration, we choose the hyper parameters for the prior distributions of α and β as (*a, b*): (1, 1) and for (*c, d*): (1, 1). With this setting, we have obtained the point predictors for future order statistics and constructed 95% Bayesian two sided equi-tailed prediction intervals (*L, U*) from a future sample of size m = 5, 7, 9 based on upper *k*-records, for k = 2 and k = 3 with n is fixed as 10. The results are obtained and are presented in Table 6.

Table 1: The bias and MSE of MLEs of α and β for *k* = 2 and *k*= 3

n	α	β	***k*=2**				***k*=3**			
			$\hat{\alpha}_{ML}$		$\hat{\beta}_{ML}$		$\hat{\alpha}_{ML}$		$\hat{\beta}_{ML}$	
			Bias	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
	1.5	2	0.7174	0.1413	-0.9430	1.8782	0.3467	0.1297	-1.6390	0.6957
6	1.5	2.5	0.9058	0.1337	-1.2690	1.0209	0.3763	0.1170	-2.0230	0.9608
	2	2	0.5495	0.1560	-1.4070	1.1986	0.1102	0.1842	-1.7840	0.3262
	2	3	0.6577	0.0747	-2.3200	1.3592	0.0196	0.0794	-2.8130	0.2472
	1.5	2	1.0397	0.1102	-0.3050	1.1025	0.5999	0.1079	-1.1170	0.5744
8	1.5	2.5	1.2642	0.0558	-0.7760	1.0932	0.6206	0.0755	-1.5300	0.7839
	2	2	0.7427	0.0913	-0.8560	1.9572	0.4352	0.1476	-1.8420	0.1665
	2	3	0.9538	0.0344	-2.0010	0.8171	0.4506	0.0785	-2.7550	0.4035
	1.5	2	1.1657	0.0882	0.1590	1.0491	0.8676	0.0816	-0.5940	0.4079
10	1.5	2.5	1.4301	0.0142	-0.4280	1.0485	0.7433	0.0117	-0.6728	0.4850
	2	2	0.8322	0.0132	-0.5681	0.1540	0.3152	0.1017	-0.8144	0.0150
	2	3	0.9014	0.0108	-0.3700	0.4520	0.4506	0.0653	-0.7550	0.0351

Table 2: The bias and MSE of Bayes estimators of α and β for *k* = 2 and *k* = 3

n	α	β	k = 2				k = 3			
			$\hat{\alpha}_{BS}$		$\hat{\beta}_{BS}$		$\hat{\alpha}_{BS}$		$\hat{\beta}_{BS}$	
			Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
6	1.5	2	0.4182	0.1068	-0.1130	0.9862	0.1599	0.0924	-0.9340	0.4746
	1.5	2.5	0.4131	0.1032	-0.6059	0.5826	0.1653	0.1093	-0.5070	0.7284
	2	2	-0.0460	0.0114	-0.1658	0.3695	-0.2324	0.0735	-0.1546	0.2815
	2	3	-0.1627	0.0293	-0.0378	0.1567	-0.1694	0.0294	-0.1990	0.1368
8	1.5	2	0.4440	0.0835	-0.1581	0.9350	0.3621	0.0850	-0.2373	0.2637
	1.5	2.5	0.5116	0.0163	-0.1528	0.3879	0.3423	0.0178	-0.7346	0.1010
	2	2	-0.0472	0.0069	-0.1513	0.3391	-0.1534	0.0258	-0.2093	0.0653
	2	3	-0.1263	0.0183	-0.0515	0.1105	-0.1719	0.0103	-0.1809	0.1583
10	1.5	2	0.4033	0.0707	-0.1238	0.7412	0.3777	0.1758	-0.2562	0.0751
	1.5	2.5	0.4411	0.0116	-0.6619	0.2829	0.3796	0.0107	-0.7751	0.0519
	2	2	-0.0109	0.0046	-0.2077	0.0103	-0.0845	0.0125	-0.2903	0.0146
	2	3	-0.0458	0.0045	-0.1563	0.1024	-0.1602	0.0101	-0.2287	0.0175

Table 3: The 95% Bayesian prediction interval and point predictor for $Y_{j:m}$ from a future sample of size m from a Weibull distribution when (a, b) = (2.5, 2.5) and (c, d) = (2.5, 2.5)

		k=2					*k=3*				
m	j	L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$	L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$
10	1	0.0031	0.0339	0.0308	0.0106	0.0128	0.0094	0.0434	0.0340	0.0127	0.3806
	3	0.0240	0.0715	0.0475	0.0423	0.4845	0.0276	0.0623	0.0347	0.0547	0.8438
	5	0.0683	0.0752	0.0069	0.0712	0.0812	0.0888	0.2014	0.1126	0.1519	0.8941
	10	0.8586	1.9684	1.1098	1.4070	1.4524	0.8854	1.9753	1.0899	1.7031	2.3870
15	1	0.0002	0.0016	0.0014	0.0015	0.0157	0.0011	0.0017	0.0006	0.0016	0.1518
	4	0.0211	0.0250	0.0039	0.0242	0.0471	0.0222	0.0286	0.0064	0.0247	0.7092
	10	0.1598	0.2006	0.0408	0.1955	0.4498	0.1819	0.2027	0.0207	0.2005	0.8128
	15	1.0035	1.2180	0.2145	1.1547	1.2544	1.1032	1.2050	0.1018	1.1581	1.9445
20	1	0.0006	0.0012	0.0006	0.0009	0.0112	0.0007	0.0011	0.0004	0.0011	0.1361
	5	0.0201	0.0232	0.0031	0.0237	0.4256	0.0207	0.0246	0.0040	0.0240	0.6562
	10	0.1613	0.2101	0.0488	0.2003	0.3250	0.1809	0.2213	0.0405	0.2093	0.6925
	15	0.2234	0.2663	0.0429	0.2331	0.4126	0.2724	0.2969	0.0245	0.2898	0.9126
	20	1.2014	1.3937	0.1923	1.2547	2.5125	1.2156	1.3009	0.0853	1.5578	2.9513

Table 4: The 95% Bayesian prediction interval and point predictor for $Y_{j:m}$ from a future sample of size m from a Weibull distribution when (a,b) = (2, 2) and (c,d) = (2, 2)

m	j	k=2					K = 3				
		L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$	L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$
10	1	0.0031	0.0333	0.0319	0.0034	0.1024	0.0074	0.0411	0.0337	0.0111	0.3911
10	1	0.0031	0.0333	0.0302	0.0104	0.10236	0.0073	0.0411	0.0337	0.0110	0.3910
10	3	0.0239	0.0751	0.0512	0.0401	0.20063	0.0250	0.0699	0.0448	0.0503	0.9525
10	5	0.0617	0.0702	0.0085	0.0620	0.19190	0.0622	0.0862	0.0241	0.0840	1.0154
10	10	0.8102	1.8381	1.0280	1.1155	1.69552	0.8761	1.8382	0.9621	1.5115	2.5562
15	1	0.0011	0.0015	0.0005	0.0040	0.40103	0.0011	0.0016	0.0004	0.0013	0.1733
15	4	0.0204	0.0250	0.0046	0.0218	0.61814	0.0213	0.0275	0.0062	0.0243	0.8263
15	10	0.1427	0.1909	0.0482	0.1824	0.88753	0.1693	0.1884	0.0191	0.1724	1.0472
15	15	0.9711	1.2008	0.2297	1.1955	1.95498	0.9721	1.1803	0.2082	1.1155	2.1014
20	1	0.0006	0.0012	0.0006	0.0006	0.54681	0.0006	0.0009	0.0003	0.0009	0.5915
20	5	0.0170	0.0204	0.0034	0.0179	0.64383	0.0170	0.0204	0.0034	0.0009	0.9438
20	10	0.1581	0.2085	0.0505	0.2004	0.80366	0.1780	0.2085	0.0305	0.2013	0.8269
20	15	0.2096	0.2344	0.0248	0.2103	0.35103	0.2133	0.2734	0.0602	0.2503	1.2026
20	20	1.3603	1.8665	0.4637	1.5283	2.93121	1.1990	2.9654	1.7665	1.2833	3.6935

Table 5: The 95% Bayesian prediction interval and point predictor for $Y_{j:m}$ from a future sample of size m from a Weibull distribution when (*a*, *b*) = (1,1) and (*c*, *d*) = (1,1)

m	j	k=2					K = 3				
		L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$	L	U	AIL	$\hat{Y}_{j:m}$	MSPE $(\hat{Y}_{j:m})$
10	1	0.0014	0.0017	0.0003	0.0013	0.2580	0.0064	0.0174	0.0110	0.0104	0.9061
10	3	0.0211	0.0710	0.0499	0.0314	0.3246	0.0207	0.0608	0.0401	0.0450	0.9983
10	5	0.0520	0.0616	0.0096	0.0581	0.2651	0.0580	0.0741	0.0161	0.0640	1.1547
10	10	0.8003	1.0137	0.2134	0.9116	1.9460	0.8123	1.5137	0.7014	1.3513	2.5235
15	1	0.0002	0.0004	0.0002	0.0003	0.2401	0.0009	0.0010	0.0001	0.0010	0.7332
15	4	0.0207	0.0221	0.0014	0.0212	0.8269	0.0207	0.0211	0.0004	0.0203	0.8318
15	10	0.1185	0.1825	0.0640	0.1518	0.9527	0.1447	0.1806	0.0359	0.1590	2.7173
15	15	0.8002	1.1983	0.3981	1.1498	2.8926	0.9024	1.2328	0.3304	1.0151	3.1416
20	1	0.0001	0.0022	0.0021	0.0005	0.4468	0.0008	0.0082	0.0074	0.0007	0.9155
20	5	0.0103	0.0182	0.0079	0.0105	0.4589	0.0162	0.0211	0.0049	0.0174	0.9828
20	10	0.0249	0.0289	0.0040	0.0294	0.6561	0.0249	0.0290	0.0041	0.0254	0.8937
20	15	0.0170	0.2020	0.1850	0.0201	0.9822	0.2017	0.2602	0.0585	0.2426	2.2593
20	10	0.1580	0.2084	0.0504	0.2004	0.8037	0.1780	0.2085	0.0305	0.2014	0.8269
20	15	0.2096	0.2344	0.0248	0.2103	0.3510	0.2133	0.2734	0.0601	0.2503	1.2026
20	20	1.3603	1.8965	0.5362	1.5283	2.9312	1.1989	2.9654	1.7665	1.2833	2.6935

Table 6: The 95% Bayesian prediction interval and point predictor for $Y_{j:m}$ from a future sample of size m from a Weibull distribution based on the real data set when (*a, b*) = (1,1) and (*c,d*) = (1,1)

m	j	k=2			k=3		
		L	U	$\hat{Y}_{j:m}$	L	U	$\hat{Y}_{j:m}$
5	1	0.8215	15.8268	2.6189	0.9222	18.8268	3.0189
	2	2.6801	24.1056	6.2025	3.6401	29.3265	7.8015
	4	7.5921	29.2056	8.3624	8.9592	31.0254	9.5210
	5	10.2347	39.2058	12.5689	12.347	44.2580	13.689
7	1	0.8933	14.2539	3.6210	1.3089	25.2352	3.9625
	3	1.3028	19.3257	8.9632	1.6935	32.0216	9.3258
	5	4.2206	25.2363	11.0258	4.9806	26.2630	12.0258
	7	4.9254	26.8897	15.3698	5.0925	32.8897	18.6980
9	1	1.0256	14.9637	3.6985	1.2588	19.2670	4.0126
	3	2.2479	18.3260	11.0276	2.9852	18.6545	11.8950
	5	4.8956	26.3259	16.3289	5.0258	29.3289	16.9858
	7	6.3259	32.5890	19.9635	7.3699	32.9889	20.3688
	9	10.3289	52.0055	32.0146	10.968	55.0255	35.2325

8. CONCLUSION

Serious difficulties for the statistical inference based on records arise due to the fact that the occurrences of record data are very rare in practical situations and the expected waiting time is infinite for every record after the first. In this work we considered k-record values arising from a Weibull distribution. We obtained the MLEs and Bayes estimates of the parameters α and β. It can be seen that the bias and MSE of Bayes estimators are smaller than thatof Bayes estimators. Therefore, Bayes estimators perform better than MLE in terms of biasand MSE. The prediction interval for future order statistics and the point predictor were obtained based on upper k-record values.

ACKNOWLEDGEMENTS

The authors would like to thank the editor and the anonymous reviewers for their helpful suggestions which substantially improved the earlier version of the manuscript.

References

1. Ahmadi, J., Mir Mostafaee, S.M.T.K. and Balakrishnan, N., (2011): Bayesian prediction of order statistics based on k-record values from exponential distribution, *Statistics*, 45, 375-387.
2. Ahsanullah, M. (1995): *Record Statistics*, Nova Science Publishers, New York.
3. Al-Hussaini, E.K., (1999): Predicting observable from a general class of distributions, *Journal of Statistical Planning and Inference*, 79, 79-91.
4. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., (1992): A First Course in Order Statistics, Classics in Applied Mathematics, 54, USA.
5. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., (1998): *Records*. John Wiley and Sons, New York.
6. Chacko, M., and Mary, S.M., (2013): Estimation and prediction based on k-record values from normal distribution, *Statistica*, 54, 21-46.
7. Chacko, M. and Muraleedharan, L., (2018): Inference based on k- record values from generalized exponential distribution, *Statistica*, 54, 21-46.
8. Chandler, K.N., (1952): The distribution and frequency of record values, *Journal of the Royal Statistical Society*, Series B, 14, 220-228.
9. David, H.A. and Nagaraja, H.N., (2003): *Order Statistics*. John Wiley and Sons, New York.
10. Deheuvels, P. and Nevzorov, V. B., (1994): Limit laws for k-record times, *Journal of Statistical Planning and Inference*, 38(3), 279-307.
11. Dziubdziela, W. and Kopocinski, B., (1976): Limiting properties of the k-th record values, *Zastosowania Mat*., 15, 187-190.
12. Glick, N., (1978): Breaking records and breaking boards, *The American Mathematical Monthly*, 85(1), 2-26.
13. Gulati, S. and Padgett, W. J., (1994): Smooth nonparametric estimation of the distribution and density functions from record-breaking data, *Communications in Statistics-Theory and Methods*, 23(5), 1259-1274.
14. Malinowska, I. and Szynal, D., (2004): On a family of Bayesian estimators and predictors for Gumbel model based on the k-th lower records, *Applied Mathematics*, 31, 107-115.
15. Minimol, S. and Thomas, P. Y., (2014): On characterization of Gompertz distribution by properties of generalized record values, *Journal of Statistical Theory and Applications*, 13(1), 38-45.
16. Muraleedharan, L. and Chacko, M., (2019): Inference on Gompertz distribution based upperk- record values, *Journal of the Kerala Statistical Association*, 30, 47-63.

17. Muraleedharan, L. and Chacko, M., (2022): Estimation and prediction based on *k*-record values from a generalized Pareto distribution, *International Journal of Mathematics and Statistics*, 23(1), 59-75.

18. Nagaraja, H.N., (1988): Record values and related statistics – a review, *Communications in statistics-Theory and Methods*, 17, 2223-2238.

19. Nelson, W.B., (1982): Applied Life Data Analysis, John Wiley & Sons, New York.

20. Nigm, A. M., (1989): An informative Bayesian prediction for the Weibull lifetime distribution. *Communications in Statistics-Theory and Methods*, 18(3), 897-911.

21. Paul, J. and Thomas, P.Y., (2015): On generalized upper (*k*) record values from Weibull distribution, *Statistica*, 75, 313-330.

22. Shafay, A.R., Balakrishnan, N. and Ahmadi, J., (2017): Bayesian prediction of order statistics with fixed and random sample sizes based on *k*-record values from pareto distribution, *Communications in Statistics- Theory and Methods*, 46, 721-735.

23. Vidovic, Z., (2019): Bayesian prediction of order statistics based on *k*-record values from a generalized exponential distribution, *Stats*, 2, 447-456.13.

24. Zakerzadeh, H and Jafari, A. A., (2015): Inference on the parameters of two Weibull distributions based on record values, *Statistical Methods and Applications*, 24(1), 25-40.

NEWS CORNER

Golden Jubilee Celebration Conference of IAPQR

The Indian Association for Productivity, Quality and Reliability (IAPQR) has concluded its yearlong Golden Jubilee Celebrations. Founded in 1973 as a Registered Society with full support from the Department of Statistics, University of Calcutta, the IAPQR landed in 2022-23 in its fiftieth year. The round-the-year celebrations included a few on-line programmes in collaboration with several universities, departments and institutions within the country on a variety of topics besides an event with the Tampere University, Finland abroad. The climax of the landmark year was Golden Jubilee Celebration Conference (GJCC) on the theme "Productivity, Quality and Reliability: New Vistas" that was organized during January 7-8, 2023 in collaboration with the Sister Nivedita University (SNU), New Town, Kolkata within its premises. The conference had three broad subthemes such as (1) Reliability and Survival Analysis, (2) Improving Quality in Services, Productivity Improvement and on (3) Data Analysis in Emerging Domains. Besides theoretical issues under the subthemes there were large number of contributions from the allied disciplines such as Information Technology, Management Science, Agriculture, Finance & Banking and Safety and Health issues in the area of Industrial Productivity.

The inaugural function of the conference was chaired by IAPQR's mentor Prof S P Mukherjee. Chief Guest Mr Satyam Roy Choudhury, Founder-cum-Managing Director of Techno India Group and Chancellor of Sister Nivedita University could not make it to the inaugural session but extended his wishes through his message. Prof Dhrubajyoti Chattopadhyay, the Vice-Chancellor, SNU in his welcome address, greeted the participants and gave a brief outline of the Sister Nivedita University. He looked forward to a fruitful interaction with IAPQR. From the IAPQR side Prof Asis Chattopadhyay, General Secretary and presently Vice-Chancellor, University of Calcutta welcomed the audience and gave a brief outline of the growth and evolution of IAPQR on this momentous occasion. Prof Anupam Basu, the newly appointed Pro Vice-Chancellor also spoke on the occasion and shared his experience on the issues of productivity, quality and reliability in the field of computer technology. An important item of this session was the release of the conference souvenir. Dr Biswsanath Das, Former Chairman of IAPQR who accomplished the difficult task of compilation of souvenir with large number of pictures and historical anecdotes of IAPQR introduced the souvenir in a brief talk. The souvenir was then released by the Guest-in-Honour of the session Mr D P S Negi, Member (Finance), Prasar Bharati, and Government of India who provided his greetings and wishes to the organizers and hoped that the IAPQR will prosper further. Prof Srijib Bhusan Bagchi, Chairman, IAPQR in his remarks spoke on the contribution of IAPQR to the quality movement in the country and beyond ever since the inception. Prof Bagchi also handed over mementoes to the dignitaries on the Dias. Dr Goutam Banerjee, Convener of the GJCC provided the Vote of Thanks. Prof Sharmistha Banerjee acted as the anchor of the inaugural session while Dr Anindita Ghoshal and her colleagues of SNU coordinated the session. Prof S P Mukherjee who chaired the session, later spoke to the media persons and his impressions were later telecast from the Door Dharshan Channel of Kolkata.

The inaugural session was followed by Professor P K Bose Memorial Lecture Session in which two overseas participants gave lectures. The first was by Prof Malay Ghosh from Florida State University who spoke on Multivariate Global-Local Priors for Small Area Estimation. The second lecture was delivered by Prof Partha Lahiri, University of Maryland, USA. His topic was Multidimensional Poverty Mapping. Both papers dwelt on the small area estimation techniques at length. The Professor P K Bose Memorial Lecture Session, a yearly programme of IAPQR, was a sponsored session by L B Specialities, Mumbai. The plenary lecture of the day was delivered by Prof T V Ramanathan, S V Pune University who presented a Review on Some of the Recent Developments in the field of Credit Risk, which is a topic of currency, especially in the finance and banking sector.

The events in the post lunch session ran in eight parallel sessions namely Design of Experiment for Improving Quality, Quality Improvement in Higher Education, Probability Models, Information Technology, Data Analysis, Advanced Statistical Quality Control, Life Distributions and Productivity in Services.

The second day of the conference started with a Special Lecture by Prof Partha Pratim Majumdar who spoke on the current limitations of statistical data analysis techniques in handling voluminous genomic data. He suggested the need for development of newer statistical techniques to crunch on large data sets that genomic research generates. The plenary session of the day had two lectures on the Role of Innovation in Productivity Improvement. While Ms Deepanwita Chattopadhyay, Chairman & CEO, IKP Knowledge Park, Hyderabad addressed the issues faced during the translation of innovation into productivity enhancement. Mr Sanjay Sikdar, Associate Partner - Data

Modernization, IBM India, Kolkata addressed similar issues in the context of Information Technology sector. Both sessions were chaired by Prof Dhrubajyoti Chattopadhyay, Vice-Chancellor of SNU.

As in the first day, the lectures on the second day were divided into six parallel technical sessions namely Productivity in Agriculture, Data Analysis-New Applications, Inference I, SHE in Industrial Productivity, Quality Improvement in Financial Services and Data Analysis II respectively.

The sessions on Agricultural Productivity, Quality in Higher Education, SHE in Industrial Productivity and Quality Improvement in Financial services were deliberated by hardcore professional of the respective domains. Mr R K Elangovan, Director General, Factory Advisory Service & Labour Institute (FASLI), Ministry of Labour & Employment, Government of India, Mumbai chaired the session on Safety & Health issues in Industrial Productivity while Mr Bhaskar Sen, Retired Chairman & Managing Director, United Bank of India, Kolkata chaired the session on Quality in Financial Services. In this session, improving quality of services in the three major sectors: asset management, banking and insurance were deliberated. The session on quality improvement in higher education was chaired by Prof Dhrubajyoti Chattopadhyay, Vice-Chancellor of SNU where in the quality of education in the digital age was also discussed.

On the application side, most papers highlighted the concerns of healthcare industry and the issues of job satisfaction, occupational stress and psychological wellbeing in different work settings and the productivity issues in the public domain. In the Statistics sessions, there were quite a few presentations on theoretical aspects from areas of Advanced SQC, Life distributions and Inference. From the Data Analysis side, a review paper was presented that explored the application of directional data in management science and how data analysis can be employed in new applications such as the railway wagon wheel inspection problem.

The conference received financial support from the Tata Steel, Jamshedpur, L B Specialities, Mumbai and Gloster Limited, Kolkata. The programme was attended by about 120 participants from the premier research and educational institutions such as Indian Statistical Institute, Kolkata; Indian Institute of Technology, both Indore and Kharagpur; ISER, Mohanpur, West Bengal; IIEST, Sibpur, West Bengal; DST-NATMO, Kolkata; IIT-ISM, Dhanbad; Indian Institute of Management- Ahmedabad, Kozhikode and Indore; Viswabharati University, Shantiniketan, University of Calcutta, Kolkata, University of Burdwan, Purba Bardhaman, Aliah University, Kolkata, Raiganj

University, Raiganj, Midnapur Collage (Autonomous), Medinipur all belonging to West Bengal and SV Pune University, Pune. Apart from the collaborating Institution of SNU, there were participants from privately funded universities namely Alliance University, Bengaluru; Ravenshaw University, Bhubaneshwar and a few healthcare companies such as the multinationals IQVIA and Inference Clinical Research Services Pvt. Ltd, Kolkata.

The conference came to close after a valedictory function in which participants were encouraged to round up their impressions about the events - besides the leaders from both the collaborators: Sister Nivedita University and the IAPQR. A group photo of the participants was snapped at the end of the valedictory session to commemorate the milestone that the IAPQR touched in its 50th year of journey.

ORSI-2023 and ICBAI - 2023

The Operational Research Society of India (ORSI) Karnataka, Department of Management Studies, Indian Institute of Science (IISc), Bangalore and the Analytics Society of India, DCAL, IIM Bangalore are jointly organising the 56th Annual Convention of ORSI (2023) and the 10th International Conference on Business Analytics and Intelligence (ICBAI-2023), and the same are scheduled to be held during December 18-20, 2023 at J.N. Tata Auditorium, IISc Bangalore, INDIA. The aim of the joint conference is to provide a platform for distinguished practitioners, academicians, and researchers to share their knowledge on current utilities of Operations Research, Business Analytics and Business Intelligence.

Intending participants are invited to submit abstracts of their papers by July 15, 2023. All accepted papers will be published as a proceedings volume. Three best papers from each of industry, academic and student presentations will be awarded.

For detailed information, visit the conference website: www.mgmt.iisc.ac.in/orsi-ka

Conference on Statistical Foundations of Data Science and their Applications

International Conference on Statistical Foundations of Data Science and their Applications is being organised at Princeton University during May 8-10, 2023. The conference intends to provide an excellent forum for scientific communications and promote collaborations among researchers in statistics and data science. It will cover a wide range of topics in relation to recent developments and the state of the art in a variety of modern research topics on statistics and data science, as well as their applications. For further information, visit the website https://fan60.princeton.edu/ or contact the local organizing committee chair: Matias D. Cattaneo (cattaneo@princeton.edu), or the program chair: Runze Li (rzli@psu.edu).

IISA-2023

The annual conference of the International Indian Statistical Association (IISA) will take place at the Colorado School of Mines, Golden, Colorado, USA during June 1-4, 2023. The planery talks at the conference will be delivered by Professors John Abowd, Paul Albert, Amarjit Budhiraja, Doug Nychka and Aarti Singh. There will be several special invited talks, invited and contributed sessions, poster sessions, and three short courses on record linkage, machine learning and networks. Detailed information can be obtained from www.intindstat.org/conference2023/index

ACKNOWLEDGEMENT TO REVIEWERS

The Editorial Board of IAPQR Transactions thanks the following referees for reviewing papers for the journal during 2022-23.

Dr. Durba Bhattacharya	:	St. Xavier's College, Kolkata
Dr. Moumita Chatterjee	:	Aliah University, Kolkata
Dr. Nandita Das	:	Indian Statistical Institute, Kolkata
Dr. Minerva Mukhopadhyay	:	Indian Institute of Technology, Kanpur
Dr. Sujoy K. Mukhuti	:	Indian Institute of Management, Indore
Dr. T.J. Rao	:	Retd. Professor, Indian Statistical Institute, Kolkata
Prof. Sanjeev Sabnis	:	Indian Institute of Technology, Bombay
Dr. Bikas K. Sinha	:	Retd. Professor, Indian Statistical Institute, Kolkata
Dr. Gajendra K. Viswakarma	:	Indian Institute of Technology (ISM), Dhanbad

BOARD OF GOVERNORS

INDIAN ASSOCIATION FOR PRODUCTIVITY, QUALITY AND RELIABILITY

Board of Governors (2020-2022)

Founder Members	
Professor B. R. Dey	Professor S.P. Mukherjee
Mentor	
Professor S.P. Mukherjee Former Centenary Professor of Statistics University of Calcutta	
President	
Mr. Rakesh Himatsingka Chairman & Managing Director India Carbon Ltd.	
Vice-Presidents	
Professor H.S. Ray Former Emeritus Scientist (CGCRI), Kolkata	Dr. Bikas C. Sanyal
General Secretary	
Professor Asis Kumar Chattopadhyay Department of Statistics Calcutta University	
Joint Secretaries	
Dr. Sharmistha Banerjee Department of Business Management Calcutta University	Dr. Krishnendu Das ICAR-National Bureau of Soil Survey & Land Use Planning
Treasurer	
Dr. Lakshmikanta Dutta	
Ex-Officio Members	
Indian Statistical Institute Kolkata	Indian Institute of Social Welfare and Business Management, Kolkata
Life Members	
Dr. Surojit Mookherjee	Dr. S. Panja
Dr. Gautam Banerjee	Dr. K.N. Singh
Dr. Ayan Chandra	
Institutional Members	
CSIR-Central Glass and Ceramic Research Institute, Kolkata	Assam Carbon
Defence Metallurgical Research Laboratory, Hyderabad	Gloster Ltd. Kolkata
Corporate Members	
India Carbon Ltd.	Urban Engineering Association Pvt. Ltd.
Permanent Invitees	
Professor B. Das	Professor Bikas K. Sinha
Professor Manisha Pal	Professor R.N. Chakraborty
Mr. Subrata Ghosh	Dr. Ambar Ghosh