

INTERNATIONAL RESEARCH JOURNAL OF SCIENCE ENGINEERING AND TECHNOLOGY



ISSN 2454-3195

An Internationally Indexed Peer Reviewed & Refereed Journal

WWW.RJSET.COM
www.isarasolutions.com

Published by iSaRa Solutions

ENVIRONMENTAL SUSTAINABILITY OF FUTURE SMART CITIES IN INDIA.....	5
J.Bhattacharjee	5
A DETERMINATION OF THE RELATIVE PRODUCTIVITY OF AUTOMOBILE DEALERS IN KERALA THROUGH DATA ENVELOPMENT ANALYSIS WITH SPECIAL REFERENCE TO THE SMALLEST DISTRICT ALAPPUZHA	12
Dr. Jeemon Joseph	12
For the effective conduction of the study we refer to both PRIMARY and SECONDARY data.	13
A Generalized Double Sampling Estimator of Population Variance Using Auxiliary Information in the form of Mean and Variance Both	21
Peeyush Misra.....	21
Converting weakness into the Strength: the Israel Story of Water	32
Ankit Saraswat	32
Managing CLOUD computing vulnerabilities with Managed services Technology.....	36
By Monisha.....	36
Different techniques of Analysis of Data.....	41
By ARUN PRATAP SINGH	41
Process of Data Analysis	41
Net Neutrality and It's Aspects.....	48
Parental Challenges of Mentally Retarded Children	52
VINNETH.V	52
A REVIEW: HEURISTIC ALGORITHMS.....	59
Jugmendra Singh.....	59
Study of friction stir welding and its process parameters for joining of dissimilar metals	68
Amit Kumar	68
1. INTRODUCTION	68
3. LITERATURE SURVEY	73
Design of a 4 th Order Low-Pass Butterworth Filter	79
Sukumar Chandra*, Tapas Misra.....	79
Wall shear stress: time periodic change during blood flow through a porous medium.....	83
S CHANDRA.....	83
Bioremediation of contaminated soil :principle,challenges and future perspectives.....	88

Jyoti Pathak,	88
Importance of Remote Sensing Technology in Indian Agriculture Improvement.....	95
Mukul Kumar	95
Article	102

ENVIRONMENTAL SUSTAINABILITY OF FUTURE SMART CITIES IN INDIA**J.Bhattacharjee**

Prof. & Advisor, Civil Engineering Deptt.,
Amity University UP, Noida, Sector-125,
Former Chief Engineer & Jt. Director General (MES/MOD)

ABSTRACT

Urbanization in India has significant implications for the development of the Country. By 2030, India's population is likely to touch about 600 millions or nearly twice that of US. This will exert tremendous pressure on urban infrastructure and services. It is therefore imperative that we find innovative solutions for urban challenges of growth and sustainability. Green growth strategies can build environmentally sustainable cities. There is evidence that human activity has caused unprecedented change, and population will soon stress the world's natural resources to the breaking point. Global warming, air pollution, land degradation, declining per-capita availability of fresh water, food shortages, and reduced biodiversity are some of the starkest challenges. Top priorities for smart cities include sustaining water, energy, and food supplies, managing waste (about 95% of cities still dump raw sewage into their waters), and reducing green house emissions. All these issues will be discussed in the paper in great details to address future smart cities being planned in India. We can take lessons learnt from advanced countries while planning our smart cities.

INTRODUCTION

The transformation to smarter cities in India will require innovation in planning, management, and operations. Several ongoing projects around the world illustrate the opportunities and challenges of this transformation. Cities must get smarter to address an array of emerging urbanization challenges particularly the problem of environmental sustainability. The number of cities worldwide pursuing smarter transformation is growing rapidly. Cities are experiencing unprecedented socio-economic crises. Urban growth and migration are putting significant stress on city infrastructures, as demand outpaces supply of water, energy, transportation, healthcare, education and safety, apart from major problem of environmental sustainability. To reduce costs, improve efficiencies, and deliver quality of life, citizen expects while balancing budget, cities are increasingly looking to information and communication technology and new working practices. There is an urgent need for cities to become smarter in how they manage their infrastructure and resources to the existing and future needs of their citizen. Concurrent trends in urbanization, economic growth, technological progress, and environmental sustainability are the drivers for the new found urgency in India, in line with other advanced countries.

The vision of “Smart cities” is the urban center of the future, made safe, secure, environmentally green, and efficient, because all structures, whether for power, water, transportation, etc. are designed, constructed, and maintained making use of advanced, integrated materials, sensors, electronics, and networks, which are interfaced with computerized systems comprised of database, tracking, and decision making algorithms.

ENVIRONMENTAL SUSTAINABILITY

There is evidence that human activity has caused unprecedented environmental change, and population growth will soon stress the world’s natural resources to the breaking point. Global warming, air pollution, and degradation, declining per capita availability of fresh water, food shortage, and reduced biodiversity are some of the starkest challenges. So, top priorities for the smart cities will include sustaining water, energy, food supplies, managing wastes, since about more than 90% of cities in India still dump raw sewage into their waters, and reducing greenhouse gas emissions. The various processes of smart cities can be broadly divided into planning, management, and operations.

Planning

A smart city provides urban planners with tools to exploit various sources of information about human behavior to aid in the allocation of resources like water, transportation and so on, as the city evolves. Holistic modeling of the city’s eco system provides quantitative support for strategy development, performance evaluation, identification of emerging best practices, and integration of initiatives. Analyzing data from other comparables cities can help planners calibrate urban dynamics models and compare their relative progress. Difficulties encountered while planning/ implementation of scheme and lessons learnt can be gainfully utilized, while planning in India’s smart cities. Special attention need to be given in transportation system of the future cities, since almost all the big cities in India is almost choking due to mismanagement of traffic on roads.

Management

A smarter city can coordinate infrastructure management activities like the creation and maintenance of roads, equipment and other assets, by providing cross-agency visibility of planned interventions. As an example, the electrical utilities replacement of a cable under a street intersection might offer traffic managers an opportunity to save money by replacing a signal in the same location. By providing a time dimensions, smarter city data can reveal historical views of each domain and enable managers to project its evaluation.

Operations A smart city integrates multiple data sources to represent the interdependence of urban domains in real time. For example, electrical utilities can combine sophisticated models of near-term demand based on historical usage pattern (day of week, holidays, local weather, major events, and so on) with real time traffic information that could impact future demand. Thus, awareness of a major delay in outbound commuter traffic in the early evening could let the utility

project a delay in demand, because those commuters will arrive home late. The utility could likewise use real time weather data to predict the location of cables damaged in a rainstorm. In the long run Smart City's vision, systems and structures will monitor their own conditions and carry out self-repair, as needed. The physical environment, air, water, and surrounding green spaces will be monitored in non-obtrusive ways for optimal quality, thus creating an enhanced living and working environment that is clean, efficient, and secure and that offers these advantages within the framework of the most effective use of all resources.

EXAMPLES OF SMART CITIES IN THE WORLD

There are number of Smart Cities of various sizes, geography, and economy illustrates various aspects of Smart city transformation. The name of those smart cities is brought out below:

- Arlington Country, Virginia, USA
- Columbus, Ohio, USA
- Ipswich, Queen island, Australia
- Mitchell, South Dakota, USA
- New Taipei, Taiwan
- Rio de Janeiro Brazil
- Surrey, British Columbia, Canada
- Dubuque, Iowa
- Bornholm, Denmark
- Songdo JBD, Korea
- Stockholm, Sweden

The first seven cities were declared by the ICF as "World top 7 Cities" in 2015, from 5 different nations.

INNOVATION CHALLENGES OF SMART CITIES

The ideal of smart city is a closed loop of interconnected city systems. These can be characterized by function: sensing, information management, analysis, and modeling, and influencing outcomes. To optimize key metrics and performance indicators, all of these systems must be tightly integrated.

Sensing a City and its Inhabitants

A city is full of sensors of smart water and electric meters, mobile phones, GPS devices, traffic sensors, parking meters, pipe sensors, weather sensors, building sensors, and so on. Even people can be sensors-using crowdsensing to gather intelligence on city operations is an emerging research area. The main innovation challenges in sensing a city and its inhabitants are trading off cost with quality and dual usage, and ensuring privacy and security.

Cost Versus Quality

Cheap, ubiquitous sensors can be used in large numbers, but their noisy, low quality signals impose nontrivial burden on analytics systems and might also require frequent calibration and diagnostic evaluation. In contrast, high cost sensors with embedded intelligence can make analysis simpler and more accurate, and may be self-calibrating and diagnosing, but cannot be installed in the quantities needed to cover large areas.

Managing Information across all City Systems

The main research challenges in managing smart city data are the need for common information models and the ability to safely share information across multiple agencies within a city and among multiple cities in metropolitan region.

WHY INTIATIVE FOR SMART CITIES

The idea of smart city came into formulation owing to the need to accommodate rapid urbanization of the age. Interest in smart cities continues to grow, driven by a range of socio-economic and technological developments across the globe. It is due to the increasing number of smart cities that established supplier from energy, transport, buildings, and government sectors are moving into smart city market, while setups are addressing a range of emerging opportunities in the same field.

According to Research paper, the global smart city technology market is expected to grow from \$ 8.8 billion to annually in 2014 to more than \$ 27.5 billion by 2023. Cities are seeking partners and suppliers to collaborate on ambitious programs for sustainability, innovation in public services, and economic development that depends on significant technology investments. The leading players in this market not only have the capacity to provide leadership on large scale projects spanning multiple city requirements, but also delivering smart infrastructure, IT, and communication solutions to cities, supporting cities across multiple operational and infrastructure issues, and have established a global presence.

The report examines the strategy and execution of 16 leading smart city suppliers with the capacity to provide leadership on large scale smart city projects spanning over multiples operational and service areas. These smart city suppliers are rated on 10 criteria i.e. Vision; go-to-market strategy; Partners; Product strategy; Geographic reach; Market share; Sales and marketing; Product performance and features; Product integration, and staying power. IBM and Cisco are some of the top suppliers in the global smart city market. They are ranked the highest in terms of strategy and execution.

SCOPE OF SMART CITIES IN INDIA

India is drawing on the development of smart cities at the global level. Honorable Prime Minister Narendra Modi's vision of "Digital India" has a plan to build 100 smart cities across the country. PM in his speech said "Cities in the past were built on river banks. They are now built along

highways. But in future, they will be built based on availability of optical fiber networks and next – generation infrastructure”

Digital India envisages making India a leader in digitally delivering services in the health, education, banking sectors. PM announced an investment of \$ 1.2 billion in smart cities with more funding coming from private sectors and abroad. As reported, a number of new cities are already under planning/execution, especially, in the corridor between Delhi and Mumbai. Many of the planned cities include Special Investment Regions or Special Economic Zones, which relax regulations, reduce taxes, and generally make it easier for foreign Countries to invest. The \$ 100 billion Delhi- Mumbai corridor effort has a 26% investment from Japan. Singapore Foreign Minister on his visit to India offered to build one smart city. Also, UK Govt. is giving help UK companies to invest in Indian infrastructure. Like this many other Countries are showing interest in developing PM’s vision of smart cities in India. The Government of India had signed an agreement with the US to develop Ajmer, Allahabad and Vishakhapatnam as smart cities during the visit of US President Barack Obama in 2014. Earlier, an agreement with Japan to develop Varanasi as a smart City was signed also in August 2014.

Cisco has already announced the ”Cisco Smart City” as a blue print for the future of smart and connected communities in India at Bangalore. The Cisco Smart City, with LEED platinum certified buildings, has also been designed to showcase the future of work- a future that is not just about when and where a people work, and companies can foster creativity, collaboration and productivity in the workplace. The Cisco Smart City allows the company’s employees, customers and partners to experience what’s possible when you realize the potential of internet of everything and digitize companies, cities and countries. Spread over a 2.6 million square feet, and designed as a campus-as-a-city for thousands of Cisco employees to work,play and learn. The Cisco Smart City is a spectacular showcase of how a pervasive physical network infrastructure can easily connect devices such as sensors, information access points and mobile devices and with a high degree of security. The Cisco Smart City also demonstrated how intelligent networks could enable digitally empowered citizens through the availability of government services in real time, online and on mobile platforms.

India’s first smart city is planned on the dusty banks of the Sabarmati River in Gujarat. The modern underground infrastructure and Office blocks has already been constructed. The plan is for a meticulously planned metropolis complete with gleaming towers, drinking water on tap, automated waste collection and a dedicated power supply, which are luxuries to many Indians. At a cost of about \$ 1 trillion, according to estimates from consultants KPMG, the plan is also crucial to PM’s ambition of attracting investment, while providing jobs for the million or more Indians who join the workforce every month. This grand scheme, still a nebulous concept involving quality communications and infrastructure, is beginning to take shape outside Gandhinagar, capital of Gujarat;with the first city the government hopes will provide a model for India’s urban future. To build smart cities, India allocated 60 billion rupees in its annual budget for the financial year 2015-16.

CHALLENGES IN ENERGY SYSTEMS FOR SMART CITIES

The Smart Cities strategies for environmental sustainability should include followings:

- Promoting distributed generation systems;
- Promoting cogeneration (combined heat and power systems) also for residential, tertiary and commercial buildings;
- The deployment of sustainable mobility through the use of electric vehicles;
- Promoting the rational use of electricity through actions aimed at curbing energy expenditure and the development of near-zero energy buildings.

In addition, further development needs to be undertaken in terms of the overall vision to propose new tools, aimed at optimizing the design and e-governance for energy systems by followings:

- The integration of new conceptual tools based on complexity science and risk analysis aimed at proposing a new generation of technological tools;
- The development of new technologies for service optimization and reconfiguration aimed at ensuring adequate levels of reliability and quality;
- The development of new technologies for reducing the infrastructure vulnerabilities;
- The analysis and control of interdependencies among the infrastructures and between them and the environment in which they are installed;
- The proposal advanced systems for on line” Sustainable Energy Microsystem (SEM) operation management;
- The study of the organizational problems and the human factor impacts on SEM and management.

The above activities are the general topics of design and construction of nearly zero energy buildings with the use of innovative technologies as home and building automation. The buildings integrated in “energy hub” with dispersed generation and urban mobility systems, constitute the “energy islands” of the future smart grid.

In the long term Smart Cities vision, systems and structures will monitor their own conditions and carry out self-repair, as needed. The physical environment, air, water, and surrounding green spaces will be monitored in non-obtrusive ways for optimal quality, thus creating an enhanced living and working environment that is clean, efficient, and secure and that offers these advantages within framework of the most effective use of all resources.

VISION OF SMART CITIES COUNCIL IN INDIA

Smart Cities Council-India (SCC-I) is a collaborative initiative of the Global Smart Cities Council and center for Sustainable Development (CSD) - a nonprofit organization engaged in promoting sustainability across sectors. Located in Bangalore, the IT and knowledge capital of India, SCC-I work in collaboration with the Global Council’s efforts to promote development of smart cities around the world. The mission of Council is to accelerate growth in the smart cities

sector by lowering barriers to adoption through leadership, outreach, tools and advocacy. The Council's partners include some of the world's leading technology companies with expertise in smart cities and infrastructures.

CONCLUSION

The smart city project needs to be designed wisely, considering the total population as the key point. As such the vision of smart cities is incredible and rational; its implementation in India at the given socio-economic condition needs to be cautiously implemented. However, as a fast developing country, the country needs to keep up with global standards. Hence planning and execution of this plan could make India take a major leap in the race of sustainable development, keeping the environmental aspect in mind. As our existing cities are choking with environmental problems, the future smart cities performers have to use innovative concepts for utilizing resources in such a way that an environmental sustainability will be ensured. All growth in cities is now mainly at the cost of environment. Thereby there is an urgent need to address the issue, while planning our future smart cities, based on experience of advanced countries while developing smart cities. A smart city is capable of joining competitiveness and sustainability by integrating different dimensions of development. The policy makers could consider sustainable, maintainable, affordable, reliable and technically feasible solutions to make cities truly smart and propel India into higher trajectory of growth. Therefore, planners could consider setting of satellite towns as smart cities, to begin with, as modifying existing infrastructure in deeply congested existing cities may be more difficult as well as expensive. The environmental sustainability will be the key factor for future smart cities in India.

REFERENCES

1. (2015) Smart Cities require Smart Planning by Singh Charan & Mittal Tarun in Internet
2. (2013) Smart Cities and their Innovation Challenges- published in 27th International Conference pp- 1361-1367
3. (2012) Challenges in energy systems for the smart cities of the future published in IEEE International issue Sept 12. By Brena M etc.

A DETERMINATION OF THE RELATIVE PRODUCTIVITY OF AUTOMOBILE DEALERS IN KERALA THROUGH DATA ENVELOPMENT ANALYSIS WITH SPECIAL REFERENCE TO THE SMALLEST DISTRICT ALAPPUZHA

Dr. Jeemon Joseph

INTRODUCTION

The main goal of the study is the determination of the relative productivity of automobile dealers at Alappuzha. Productivity can be defined as a ratio output produced to the input required for producing the output. The productivity analysis helps to make a perfect evaluation about employees of the company. As far as the automobile industry is concerned there is a huge opportunity for jobs. There is a firm presence of big automobile companies in Alappuzha. Since there exists a tough competition between the companies, dealers get the advantage of price reduction. This study helps to understand the productivity of these 5 companies and also it helps to understand the sales of the different automobile products.

The automobile industry in India is the tenth largest in the world with an annual production of approximately 2 million units. India is expected to overtake China as the world's fastest growing car market in terms of the number of units sold and the automotive industry is one of the fastest growing manufacturing sectors in India. Because of its large market (India has a population of 1.1 billion; the second largest in the world), a low base of car ownership (25 per 1,000 people) and a surging economy, India has become a huge attraction for car manufacturers around the world. Though several major foreign automakers, like Ford, Suzuki, GM, and Honda have their manufacturing bases in India, the Indian automobile market is dominated by domestic companies. Maruti Suzuki is the largest passenger vehicle company, Tata Motors is the largest commercial vehicle company while Hero Honda is the largest motorcycle company in India. Other major Indian automobile manufacturers include Mahindra & Mahindra, Ashok Leyland and Bajaj Auto.

OBJECTIVES

- “A determination of the relative productivity of automobile dealers at Alappuzha through data envelopment analysis”.
- To study on the sales of the different automobile products.

LIMITATIONS

- Some of the companies are not ready to reveal the official matters deeply.
- Tough competition exists between companies as they compete similar products to the same kind of customers.
- The present financial recession had affected the entire automobile industry thus affecting the productivity of the employees.

METHODOLOGY

Population:-

- Consist of all dealers in different automobile companies in Alappuzha.

Sample:-

- Consists of all the employee who are interview for the data collection.

Sampling Technique:-

- Probability sampling

Sample Size:-

- 50

Data Collection:-

For the effective conduction of the study we refer to both PRIMARY and SECONDARY data.

- Primary data collection through the direct interview with the managers.
- The secondary data was collected by referring to company brochures, and by surfing the internet and collecting the relevant data from the various company website etc and other related websites.
- **Analysis and Conclusion**

The respected study conducted by collecting data from different dealers and the respected data by analyzed using various techniques so as to find out the determination of the relative productivity of automobile dealers at Alappuzha through data envelopment analysis.

DATA ANALYSIS AND INTERPRETATION

The research study is conducted by preparing the questionnaire to understand the productivity of Automobile dealers in Alleppey. The prepared questionnaire is analyzed by conducting the survey with the dealers after making a visit to the company and by interviewing the manager to get an insight in to the details of the research process.

The study is aimed at collecting the data from five companies namely TVS Motors, TATA Motors, Hero Honda, Maruthi Suzuki and Bajaj Motors. The greater insight in to the data was analyzed by collecting adequate information from the internet. A referral was also made to the various magazines and trade journals like fast track and India Today and other English weeklies and dailies.

The automobile industry in India is the tenth largest in the world with an annual production of approximately 2 million units. India is expected to overtake China as the world's fastest growing car market in terms of the number of units sold and the automotive industry is one of the fastest growing manufacturing sectors in India. Because of its large market (India has a population of 1.1 billion; the second largest in the world), a low base of car ownership (25 per 1,000 people) and a surging economy, India has become a huge attraction for car manufacturers around the world. Though several major foreign automakers, like Ford, Suzuki, GM, and Honda have their manufacturing bases in India, Indian automobile market is dominated by domestic companies. Maruti Suzuki is the largest passenger vehicle company, Tata Motors is the largest commercial

vehicle company while Hero Honda is the largest motorcycle company in India. Other major Indian automobile manufacturers include Mahindra & Mahindra, Ashok Leyland and Bajaj Auto.

**TABLE NO.1
NUMBER OF EMPLOYEES WORKING IN THE ORGANIZATION**

Interpretation

Employees of the organization are divided into based on the department, basically automobile companies have sales department, front office, account section and also service.

TATA MOTORS	12
TVS	13
SUZUKI	12
HERO HONDA	42
BAJAJ AUTO	18

Only hero Honda has the servicing center. So the number of employees is more for Hero Honda than TVS, Suzuki, Tata and Bajaj.

**TABLE NO.2
BASIC CRITERIA FOR SELECTING THE EMPLOYEES TO THE
ORGANIZATION**

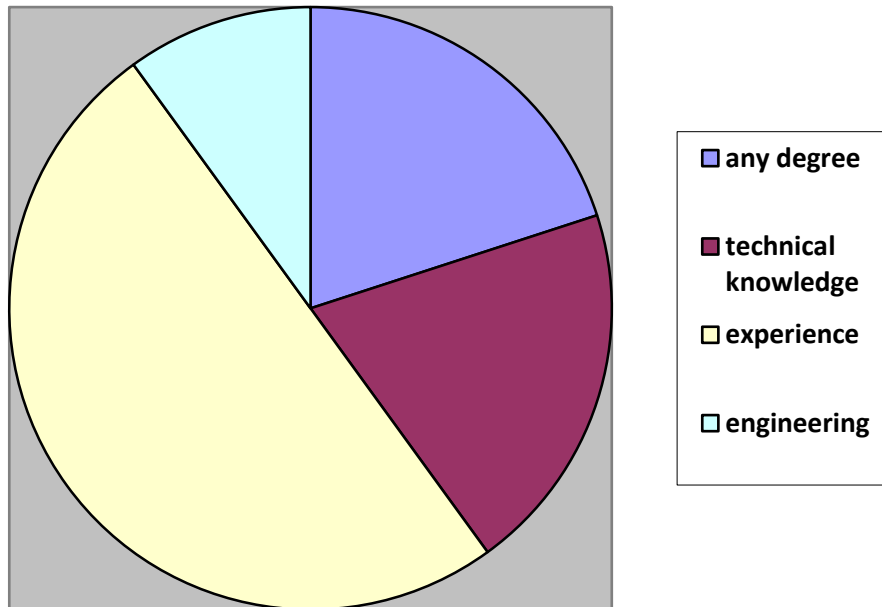
Interpretation

The basic criterion of selecting the employees into these showrooms is based on the experience. For the front office graduation is necessary than the experience. For the servicing department candidate should be from either Automobile (engineering or diploma) or mechanical (engineering or diploma) background. And also experience is required for the servicing job.

COMPANY	ACCOUNT SEC:	SALES	FRONT OFFICE	SERVICE
TATA	Tally, B com	Graduation, experience	graduation	-
TVS	tally, B com	Experience	Ms office, graduation	-
SUZUKI	B com	Experience	graduation	-
HERO HONDA	Tally, B com	Graduation, experience	Ms office, graduation	Automobile eng:, experience
BAJAJ	B com	Experience	Ms office, graduation	-

For the sales department graduation, communication skill and also experience are required. For the account section Technical skill (Ms Office, Tally) and also graduation in B com is required.

Chart No. 1



Interpretation

For recruitment of employees most of the automobile companies are looking experience, basic qualification, technical knowledge. Among this more weightage is given to experience of the candidate. And the second best preference is given to technical knowledge.

TABLE NO.3

BASIC METHODS OF PROVIDING MOTIVATION TO THE EMPLOYEES

Interpretation

The basic methods of providing motivation to the employees are, Tata motors provide regular meeting and incentives to the employees. TVS gives incentives to the employees on the

performance base. Suzuki gives incentives and TA (for the sales department only) to the employees. Hero Honda gives incentives, TA (for the sales department only), Bonus (for the festival season) to the employees. Bajaj gives incentives to the employees.

COMPANY	METHODS
TATA	Regular meeting, incentives
TVS	Incentives
SUZUKI	Incentives, TA
HERO HONDA	Incentives, TA, Bonus
BAJAJ	Incentives

TABLE NO.4
MODES OF EMPLOYEE EVALUATION

Interpretation

The mode using these companies for evaluating the employees through the sales, customer feedback. Every alternative two months Tata motors will evaluate their employee's performance. And in the case of TVS Motors and Suzuki every alternative three months and one month respectively for performance evaluation..

COMPANY	TIME	MODES
Tata	After 2 months	Through the sales, customer feedback
TVS	After 3 months	Through the sales, customer feedback
Suzuki	After 1 month	Through the sales, customer feedback
Hero Honda	After 3 months	Through the sales, customer feedback, behavior in the company
Bajaj	After 2 months	Through the sales, customer feedback

Every alternative three months Hero Honda motors will be evaluating their employee's performance. And in every alternative two months Bajaj motors will be evaluating their employee's performance.

TABLE NO.5
THE METHODS ADOPTED FOR COMPENSATION MANAGEMENT OF EMPLOYEES

Interpretation

The basic methods adopted for the compensation management of the employees are, Tata motors provides promotional offers to the employees. TVS gives ESI to the employees on the performance base. Suzuki gives incentives to the employees.

COMPANY	METHODS
TATA	Promotions
TVS	ESI
SUZUKI	Incentives
HERO HONDA	Incentives
BAJAJ	Incentives

Hero Honda and Bajaj provide incentives to the employees.

CHART NO.2

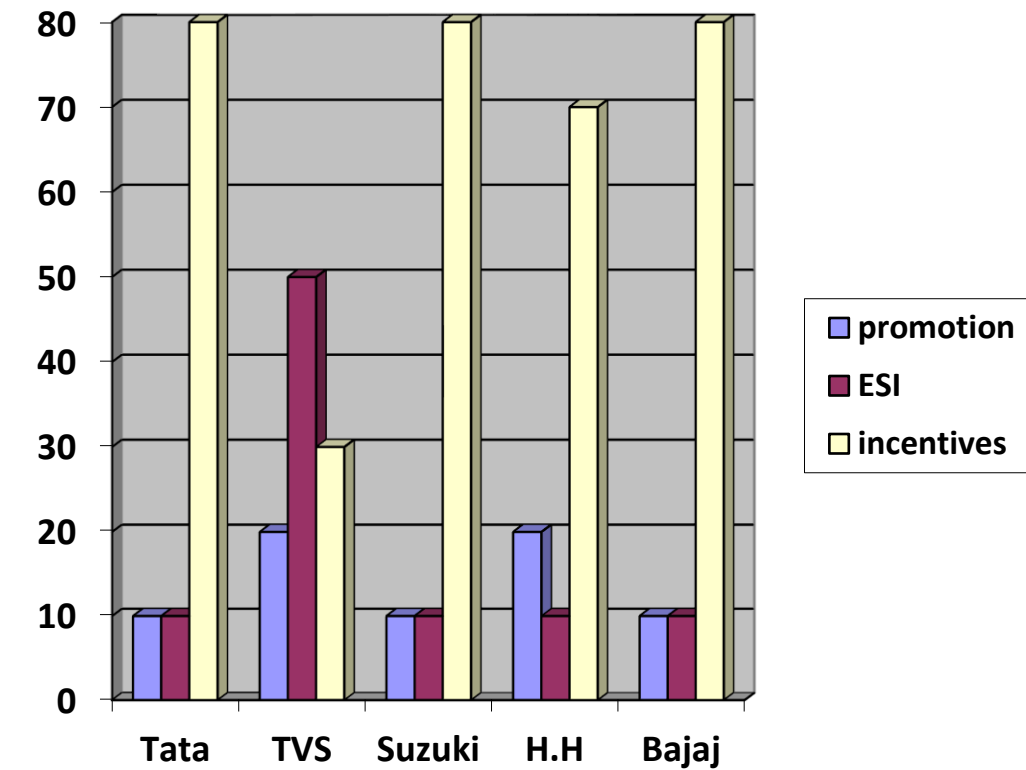


TABLE NO.6
ADDITIONAL BENIFITS

Interpretation

Employees get additional benefits in the form of DA, TA, BONUS etc...from the company. Tata gives DA, TA (for the sales department only) to the employees. TVS gives BONUS (for the festival season) to the employees. Suzuki gives TA and DA to the employees.

	ADDITIONAL BENIFITS
TATA	DA, TA (for the sales department only)
TVS	BONUS(for the festival season)
SUZUKI	TA, DA
HERO HONDA	TA (for the sales department only), Bonus (for the festival season)
BAJAJ	TA, BONUS

Hero Honda gives TA (for the sales department only), Bonus (for the festival season) to the employees. And Bajaj gives TA and BONUS to the employees.

CHART NO.3

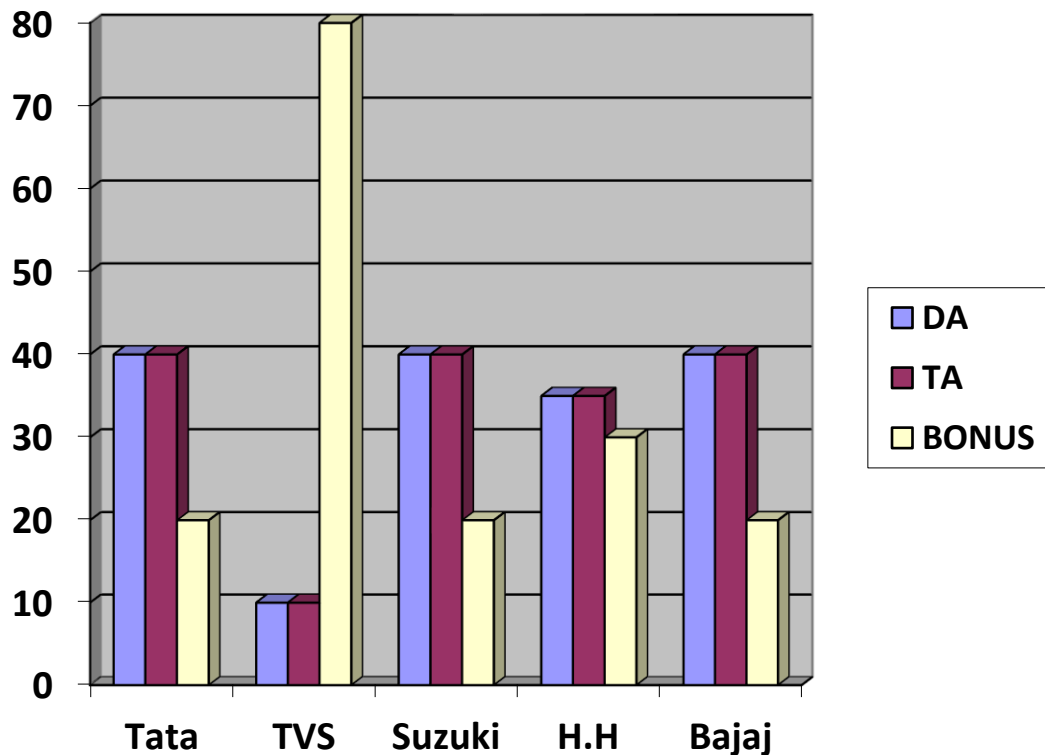


TABLE NO.7
THE SITUATION IN WHICH THE EMPLOYEES FELT OVERBURDEN OF THEIR TARGET

Interpretation

In the case of automobile industry most of the employees become over burden through huge unachievable target. In such a stressful situation company will ask the reason for unachievement and impose more instructions to the employees.

COMPANY	METHODS
TATA	Give instructions
TVS	Through the meeting(understand the employee

	situation & give instructions)
SUZUKI	Give instructions
HERO HONDA	Through the meeting(understand the employee situation & give instructions)
BAJAJ	Through the meeting(understand the employee situation & give instructions)

TABLE NO.8**PROVISION OF TRAINING AND DEVELOPMENT TO THE EMPLOYEES****Interpretation**

Companies are provides the formal training program to the employees. Tata gives training & development program after 6 months. TVS gives after 6 months. Suzuki gives after 1 year. Hero Honda gives after 6 months. And also Bajaj gives after 1 year.

	FORMAL TRAINING PROGRAM
TATA	YES(after 6 months)
TVS	YES(after 6 months)
SUZUKI	YES(after 1year)
HERO HONDA	YES(after 6 months)
BAJAJ	YES(after 1year)

FINDINGS

- 1) The analysis of the Relative Productivity helps to make a perfect evaluation about employees of the company.
- 2) As far as the automobile industry is concerned there is a huge opportunity for jobs.
- 3) Since there exists a tough competition between the companies, dealers get the advantage of price reduction
- 4) Maruti Suzuki is the largest passenger vehicle company.
- 5) Tata Motors is the largest commercial vehicle company.
- 6) Hero Honda is the largest motorcycle company in India.
- 7) Hero Honda providing the effective training program.
- 8) Suzuki motors are more concerned about the targets and impose more instructions to the employees.
- 9) The basic criteria of selecting the employees in these showrooms is based on their experience.

SUGGESTIONS

- 1) The firm has to take necessary steps to give the formal training program to the employees of firm.
- 2) The firm has to adopt only those projects which are profitable and which adds value to the society.

- 3) All the companies should start an effective grievance redressal wing.
- 4) The organization must adopt innovative methods of performance appraisal than going for the traditional method.
- 5) They should focus on effective reward management and should ensure that the reward goes to the deserved hands.
- 6) The compensation management should provide a basic framework to evaluate the level of performance of the employees.

CONCLUSION

The study revealed that the organization is adopting very extensive strategy for improving the productivity of the dealers. They are involved in the improvement of morale of the dealers by adopting performance appraisal methods and motivation techniques. The study has given an opportunity to set an insight into the human resource management activities of the firm.

A Generalized Double Sampling Estimator of Population Variance Using Auxiliary Information in the form of Mean and Variance Both

Peeyush Misra

Department of Statistics

D.A.V.(P.G.) College, Dehrdaun- 248001

Uttarakhand (India)

dr.pmisra.dav@gmail.com

Abstract: For the estimation of finite population variance, a generalized double sampling estimator representing a class of estimators using auxiliary information in the form of mean and variance both is proposed. The expression for bias and mean square error are derived. It is shown that the proposed generalized double sampling estimator has minimum variance in comparison with some of the estimators available in the literature. Efficiency comparison and an empirical study is also included.

Keywords: Auxiliary information, Bias, Mean square error and Taylor's Series Expansion.

1. Introduction

In sampling theory it is well known that the auxiliary information in sample surveys results in substantial improvement in the precision of the estimators of the population parameters and we know that sometimes parameters of the auxiliary variables are not known in advance then double or two phase sampling technique is used. In double sampling or two-phase sampling technique, we first take a preliminary large sample of size n' (called first phase sample) from a population of size N and then a sub-sample of size n (called second phase sample) is drawn from the first phase sample of size n' by simple random sampling without replacement scheme at both the phases. At first phase sample of size n' , only the auxiliary variable X be observed but at the second phase sample of size n , the study variable Y and the auxiliary variable X both are observed.

Let us denote by

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad = \quad \text{population mean of study variable}$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad = \quad \text{population mean of auxiliary variable}$$

$$S_Y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 \quad = \quad \text{population variance of study variable}$$

$$S_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \quad = \quad \text{population variance of auxiliary variable}$$

$$\bar{x}' = \frac{1}{n'} \sum_{i=1}^{n'} x_i \quad = \quad \text{sample mean of the first phase } n' \text{ sample values on auxiliary}$$

character X

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \text{sample mean of } y \text{ based on second phase sample of size } n$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \text{sample mean of auxiliary variable } X \text{ based on second phase sample of size } n$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \text{sample variance of auxiliary variable } X \text{ based on second phase sample of size } n$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \text{sample variance of study variable } Y \text{ based on second phase sample of size } n$$

$$s_x'^2 = \frac{1}{n'-1} \sum_{i=1}^{n'} (x'_i - \bar{x}')^2 = \text{sample variance of auxiliary variable } X \text{ based on second phase sample of size } n$$

$$\text{and } \mu_{rs} = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^r (X_i - \bar{X})^s .$$

For estimating finite population variance, a generalized double sampling estimator using auxiliary information in the form of mean and variance both is given by

$$d_g = \hat{\theta} - \bar{y} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \quad (1.1)$$

where $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i^2$ satisfying the validity conditions of Taylor's series expansion is a bounded function of $(\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X'^2)$ such that

$$(i) \quad g(\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X'^2) = \bar{Y} \quad (1.2)$$

(ii) first order partial differential coefficient of $g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2)$ with respect to \bar{y} at $T = (\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X'^2)$ is unity, that is

$$g_0 = \left(\frac{\partial}{\partial \bar{y}} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T = 1 \quad (1.3)$$

- (iii) second order partial differential coefficient of $g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2)$ with respect to \bar{y} at $T = (\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X^2)$ is zero, that is

$$g_{00} = \left(\frac{\partial^2}{\partial \bar{y}^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T = 0 \quad (1.4)$$

- (iv) $g_1 = -g_2$ (1.5)

for g_1 and g_2 being the first order partial derivatives of $g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2)$ with respect to \bar{x} and \bar{x}' respectively at the point $T = (\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X^2)$, that is

$$g_1 = \left(\frac{\partial}{\partial \bar{x}} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$$

$$g_2 = \left(\frac{\partial}{\partial \bar{x}'} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$$

- (v) $g_{01} = -g_{02}$ (1.6)

for $g_{01} = \left(\frac{\partial^2}{\partial \bar{y} \partial \bar{x}} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$

$$g_{02} = \left(\frac{\partial^2}{\partial \bar{y} \partial \bar{x}'} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$$

- (vi) $g_3 = -g_4$ (1.7)

for g_3 and g_4 being the first order partial derivatives of $g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2)$ with respect to s_x^2 and $s_x'^2$ respectively at the point $T = (\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X^2)$, that is

$$g_3 = \left(\frac{\partial}{\partial s_x^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$$

$$g_4 = \left(\frac{\partial}{\partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \text{ and}$$

- (vii) $g_{03} = -g_{04}$ (1.8)

for $g_{03} = \left(\frac{\partial^2}{\partial \bar{y} \partial s_x^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T$ and

$$g_{04} = \left(\frac{\partial^2}{\partial \bar{y} \partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T .$$

2. Bias and Mean Square Error of the Proposed Estimator

In order to obtain bias and mean square error of the proposed generalized double sampling estimator of population variance, let us denote by

$$\bar{y} = \bar{Y} + e_0$$

$$\bar{x} = \bar{X} + e_1$$

$$\bar{x}' = \bar{X}' + e_1'$$

$$s_x^2 = S_X^2 + e_2$$

$$s_x'^2 = S_X'^2 + e_2'$$

$$\hat{\theta} = \theta + e_3 \quad \text{where} \quad \theta = \frac{1}{N} \sum_{i=1}^N Y_i^2 \quad (2.1)$$

$$\text{with} \quad E(e_0) = E(e_1) = E(e_1') = E(e_2) = E(e_2') = E(e_3) = 0 \quad (2.2)$$

$$E(e_0^2) = \frac{\mu_{20}}{n}$$

$$E(e_1^2) = \frac{\mu_{02}}{n}$$

$$E(e_1'^2) = \frac{\mu_{02}}{n'}$$

$$E(e_2^2) = \frac{\mu_{02}^2}{n} (\beta_2 - 1)$$

$$E(e_2'^2) = \frac{\mu_{02}^2}{n'} (\beta_2 - 1)$$

$$E(e_3^2) = \frac{1}{n} (\mu_{40} + 4\bar{Y} \mu_{30} + 4\bar{Y}^2 \mu_{20} - \mu_{20}^2)$$

$$E(e_0 e_1) = \frac{\mu_{11}}{n}$$

$$E(e_0 e_1') = \frac{\mu_{11}}{n'}$$

$$E(e_0 e_2) = \frac{\mu_{12}}{n}$$

$$E(e_0 e'_2) = \frac{\mu_{12}}{n'}$$

$$E(e_1 e_2) = \frac{\mu_{03}}{n}$$

$$E(e_0 e_3) = \frac{1}{n}(\mu_{30} + 2\bar{Y} \mu_{20})$$

$$E(e_1 e'_1) = \frac{\mu_{02}}{n'}$$

$$E(e_1 e'_2) = \frac{\mu_{03}}{n'}$$

$$E(e_1 e_3) = \frac{1}{n}(\mu_{21} + 2\bar{Y} \mu_{11})$$

$$E(e'_1 e_2) = \frac{\mu_{03}}{n'}$$

$$E(e'_1 e'_2) = \frac{\mu_{03}}{n'}$$

$$E(e'_1 e_3) = \frac{1}{n'}(\mu_{21} + 2\bar{Y} \mu_{11})$$

$$E(e_2 e'_2) = \frac{\mu_{02}^2}{n'}(\beta_2 - 1)$$

$$E(e_2 e_3) = \frac{1}{n}(\mu_{22} + 2\bar{Y} \mu_{12} - \mu_{02} \mu_{20})$$

$$E(e'_2 e_3) = \frac{1}{n'}(\mu_{22} + 2\bar{Y} \mu_{12} - \mu_{02} \mu_{20}) \quad (2.3)$$

Now expanding $t = g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2)$ in (1.1) in the third order Taylor's series about the point

$T = (\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X^2)$, we have

$$t = g(\bar{Y}, \bar{X}, \bar{X}, S_X^2, S_X^2) + (\bar{y} - \bar{Y})g_0 + (\bar{x} - \bar{X})g_1 + (\bar{x}' - \bar{X})g_2 + (s_x^2 - S_X^2)g_3 + (s_x'^2 - S_X^2)g_4$$

$$\begin{aligned}
& + \frac{1}{2!} \left\{ (\bar{y} - \bar{Y})^2 g_{00} + (\bar{x} - \bar{X})^2 g_{11} + (\bar{x}' - \bar{X}')^2 g_{22} + (s_x^2 - S_x^2)^2 g_{33} + (s_x'^2 - S_x'^2)^2 g_{44} \right. \\
& + 2(\bar{y} - \bar{Y})(\bar{x} - \bar{X})g_{01} + 2(\bar{y} - \bar{Y})(\bar{x}' - \bar{X}')g_{02} + 2(\bar{y} - \bar{Y})(s_x^2 - S_x^2)g_{03} \\
& + 2(\bar{y} - \bar{Y})(s_x'^2 - S_x'^2)g_{04} + 2(\bar{x} - \bar{X})(\bar{x}' - \bar{X}')g_{12} + 2(\bar{x} - \bar{X})(s_x^2 - S_x^2)g_{13} \\
& + 2(\bar{x} - \bar{X})(s_x'^2 - S_x'^2)g_{14} + 2(\bar{x}' - \bar{X}')g_{23} + 2(\bar{x}' - \bar{X}')g_{24} \\
& \left. + 2(s_x^2 - S_x^2)(s_x'^2 - S_x'^2)g_{34} \right\} + \frac{1}{3!} \left\{ \frac{\partial}{\partial \bar{y}} (\bar{y} - \bar{Y}) + \frac{\partial}{\partial \bar{x}} (\bar{x} - \bar{X}) + \frac{\partial}{\partial \bar{x}'} (\bar{x}' - \bar{X}') \right. \\
& \left. + \frac{\partial}{\partial s_x^2} (s_x^2 - S_x^2) + \frac{\partial}{\partial s_x'^2} (s_x'^2 - S_x'^2) \right\}^3 g(\bar{y}^*, \bar{x}^*, \bar{x}'^*, s_x^{2*}, s_x'^{2*}) \quad (2.4)
\end{aligned}$$

where $g_0, g_1, g_2, g_3, g_4, g_{00}, g_{01}, g_{02}, g_{03}, g_{04}$ are already defined in equations (1.3) to (1.8) and the partial derivatives

$$\begin{aligned}
g_{11} &= \left(\frac{\partial^2}{\partial \bar{x}^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{22} &= \left(\frac{\partial^2}{\partial \bar{x}'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{33} &= \left(\frac{\partial^2}{\partial (s_x^2)^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{44} &= \left(\frac{\partial^2}{\partial (s_x'^2)^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{04} &= \left(\frac{\partial^2}{\partial \bar{y} \partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{12} &= \left(\frac{\partial^2}{\partial \bar{x} \partial \bar{x}'} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{13} &= \left(\frac{\partial^2}{\partial \bar{x} \partial s_x^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{14} &= \left(\frac{\partial^2}{\partial \bar{x} \partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{23} &= \left(\frac{\partial^2}{\partial \bar{x}' \partial s_x^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \\
g_{24} &= \left(\frac{\partial^2}{\partial \bar{x}' \partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T
\end{aligned}$$

$$g_{34} = \left(\frac{\partial^2}{\partial s_x^2 \partial s_x'^2} g(\bar{y}, \bar{x}, \bar{x}', s_x^2, s_x'^2) \right)_T \quad \text{and}$$

$$\bar{y}^* = \bar{Y} + h(\bar{y} - \bar{Y})$$

$$\bar{x}^* = \bar{X} + h(\bar{x} - \bar{X})$$

$$\bar{x}'^* = \bar{X}' + h(\bar{x}' - \bar{X}')$$

$$s_x^{2*} = S_X^2 + h(s_x^2 - S_X^2)$$

$$s_x'^{2*} = S_X'^2 + h(s_x'^2 - S_X'^2), \quad 0 < h < 1.$$

Now using the conditions given from (1.2) to (1.8) and (2.1) in (2.4), we have

$$\begin{aligned} t = & \bar{Y} + e_0 + e_1 g_1 - e_1' g_1 + e_2 g_3 - e_2' g_3 + \frac{1}{2!} \{ e_1^2 g_{11} + e_1'^2 g_{22} + e_2^2 g_{33} \\ & + e_2'^2 g_{44} + 2e_0 e_1 g_{01} - 2e_0 e_1' g_{01} + 2e_0 e_2 g_{03} - 2e_0 e_2' g_{03} + 2e_1 e_1' g_{12} \\ & + 2e_1 e_2 g_{13} + 2e_1 e_2' g_{14} + 2e_1' e_2 g_{23} + 2e_1' e_2' g_{24} + 2e_2 e_2' g_{34} \} \\ & + \frac{1}{3!} \left\{ e_0 \frac{\partial}{\partial \bar{y}} + e_1 \frac{\partial}{\partial \bar{x}} + e_1' \frac{\partial}{\partial \bar{x}'} + e_2 \frac{\partial}{\partial s_x^2} + e_2' \frac{\partial}{\partial s_x'^2} \right\}^3 g(\bar{y}^*, \bar{x}^*, \bar{x}'^*, s_x^{2*}, s_x'^{2*}) \quad (2.5) \end{aligned}$$

Now using (2.5) in (1.1), we have to the first degree of approximation

$$\begin{aligned} d_g = & \theta + e_3 - (\bar{Y} + e_0) \cdot \left[\bar{Y} + e_0 + e_1 g_1 - e_1' g_1 + e_2 g_3 - e_2' g_3 + \frac{1}{2!} \{ e_1^2 g_{11} + e_1'^2 g_{22} + e_2^2 g_{33} \right. \\ & \left. + e_2'^2 g_{44} + 2e_0 e_1 g_{01} - 2e_0 e_1' g_{01} + 2e_0 e_2 g_{03} - 2e_0 e_2' g_{03} + 2e_1 e_1' g_{12} \right. \\ & \left. + 2e_1 e_2 g_{13} + 2e_1 e_2' g_{14} + 2e_1' e_2 g_{23} + 2e_1' e_2' g_{24} + 2e_2 e_2' g_{34} \right] \\ d_g = & \frac{1}{N} \sum_{i=1}^N Y_i^2 + e_3 - \bar{Y}^2 - \bar{Y} e_0 - \bar{Y} e_1 g_1 + \bar{Y} e_1' g_1 - \bar{Y} e_2 g_3 + \bar{Y} e_2' g_3 - \bar{Y} e_0 - e_0^2 - e_0 e_1 g_1 \\ & + e_0 e_1' g_1 - e_0 e_2 g_3 + e_0 e_2' g_3 - \frac{\bar{Y}}{2!} \{ e_1^2 g_{11} + e_1'^2 g_{22} + e_2^2 g_{33} \\ & + e_2'^2 g_{44} + 2e_0 e_1 g_{01} - 2e_0 e_1' g_{01} + 2e_0 e_2 g_{03} - 2e_0 e_2' g_{03} + 2e_1 e_1' g_{12} \\ & + 2e_1 e_2 g_{13} + 2e_1 e_2' g_{14} + 2e_1' e_2 g_{23} + 2e_1' e_2' g_{24} + 2e_2 e_2' g_{34} \} \\ d_g - \sigma_Y^2 = & e_3 - 2\bar{Y} e_0 - \bar{Y} e_1 g_1 + \bar{Y} e_1' g_1 - \bar{Y} e_2 g_3 + \bar{Y} e_2' g_3 - e_0^2 - e_0 e_1 g_1 \\ & + e_0 e_1' g_1 - e_0 e_2 g_3 + e_0 e_2' g_3 - \frac{\bar{Y}}{2!} \{ e_1^2 g_{11} + e_1'^2 g_{22} + e_2^2 g_{33} \} \end{aligned}$$

$$\begin{aligned}
& + e_2'^2 g_{44} + 2e_0 e_1 g_{01} - 2e_0 e_1' g_{01} + 2e_0 e_2 g_{03} - 2e_0 e_2' g_{03} + 2e_1 e_1' g_{12} \\
& + 2e_1 e_2 g_{13} + 2e_1 e_2' g_{14} + 2e_1' e_2 g_{23} + 2e_1' e_2' g_{24} + 2e_2 e_2' g_{34} \} \quad (2.6)
\end{aligned}$$

Now taking expectation on both the sides of (2.6) and using values of the expectations given in (2.2) and (2.3), the bias in d_g to the first degree of approximation is given by

$$\begin{aligned}
\text{Bias in } d_g &= E(d_g) - \sigma_Y^2 \\
&= -\frac{\mu_{20}}{n} - \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{11} g_1 - \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{12} g_3 - \bar{Y} \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{11} g_{01} - \bar{Y} \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{12} g_{03} \\
&\quad - \frac{\bar{Y}}{2} \left\{ \frac{\mu_{02}}{n} g_{11} + \frac{\mu_{02}}{n'} g_{22} + \frac{\mu_{02}^2}{n} (\beta_2 - 1) g_{33} + \frac{\mu_{02}^2}{n'} (\beta_2 - 1) g_{44} + 2 \frac{\mu_{02}}{n'} g_{12} \right. \\
&\quad \left. + 2 \frac{\mu_{03}}{n} g_{13} + 2 \frac{\mu_{03}}{n'} g_{14} + 2 \frac{\mu_{03}}{n'} g_{23} + 2 \frac{\mu_{03}}{n'} g_{24} + 2 \frac{\mu_{02}^2}{n'} (\beta_2 - 1) g_{34} \right\} \quad (2.7)
\end{aligned}$$

Now squaring (2.6) on both the sides and then taking expectation, the mean square error to the first degree of approximation is given by

$$\begin{aligned}
\text{MSE}(d_g) &= E(d_g - \sigma_Y^2)^2 \\
&= E(e_3 - 2\bar{Y}e_0 - \bar{Y}e_1 g_1 + \bar{Y}e_1' g_1 - \bar{Y}e_2 g_3 + \bar{Y}e_2' g_3)^2 \\
&= E(e_3^2) + 4\bar{Y}^2 E(e_0^2) + \bar{Y}^2 g_1^2 E(e_1^2) + \bar{Y}^2 g_1^2 E(e_1'^2) + \bar{Y}^2 g_3^2 E(e_2^2) + \bar{Y}^2 g_3^2 E(e_2'^2) \\
&\quad - 4\bar{Y} E(e_0 e_3) - 2\bar{Y} g_1 E(e_1 e_3) + 2\bar{Y} g_1 E(e_1' e_3) - 2\bar{Y} g_3 E(e_2 e_3) + 2\bar{Y} g_3 E(e_2' e_3) \\
&\quad + 4\bar{Y}^2 g_1 E(e_0 e_1) - 4\bar{Y}^2 g_1 E(e_0 e_1') + 4\bar{Y}^2 g_3 E(e_0 e_2) - 4\bar{Y}^2 g_3 E(e_0 e_2') \\
&\quad - 2\bar{Y}^2 g_1^2 E(e_1 e_1') + 2\bar{Y}^2 g_1 g_3 E(e_1 e_2) - 2\bar{Y}^2 g_1 g_3 E(e_1 e_2') - 2\bar{Y}^2 g_1 g_3 E(e_1' e_2) \\
&\quad + 2\bar{Y}^2 g_1 g_3 E(e_1' e_2') - 2\bar{Y}^2 g_3^2 E(e_2 e_2')
\end{aligned}$$

using values of the expectations given in (2.2) and (2.3), we have

$$\begin{aligned}
\text{MSE}(d_g) &= \frac{1}{n} (\mu_{40} - \mu_{20}^2) + \bar{Y}^2 g_1^2 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{02} + \bar{Y}^2 g_3^2 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{02}^2 (\beta_2 - 1) \\
&\quad - 2\bar{Y} g_1 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{21} - 2\bar{Y} g_3 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{22} + 2\bar{Y} g_3 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{02} \mu_{20} \\
&\quad + 2\bar{Y}^2 g_1 g_3 \left(\frac{1}{n} - \frac{1}{n'}\right) \mu_{03}
\end{aligned}$$

$$\begin{aligned} \text{MSE}(d_g) = \text{MSE}(s_y^2) + \left(\frac{1}{n} - \frac{1}{n'}\right) \left\{ \bar{Y}^2 \mu_{02} g_1^2 - 2\bar{Y} \mu_{21} g_1 + \bar{Y}^2 \mu_{02}^2 (\beta_2 - 1) g_3^2 \right. \\ \left. + 2\bar{Y} \mu_{02} (\mu_{02} \mu_{20} - \mu_{22}) g_3 + 2\bar{Y}^2 \mu_{03} g_1 g_3 \right\} \quad (2.8) \end{aligned}$$

For minimizing (2.8) in two unknowns g_1 and g_3 , the normal equations after differentiating (2.8) partially with respect to g_1 and g_3 are

$$\bar{Y} \mu_{02} g_1 + \bar{Y} \mu_{03} g_3 - \mu_{21} = 0 \quad (2.9)$$

$$\bar{Y} \mu_{03} g_1 + \bar{Y} \mu_{02}^2 (\beta_2 - 1) g_3 + (\mu_{02} \mu_{20} - \mu_{22}) = 0 \quad (2.10)$$

Solving (2.9) and (2.10) for g_1 and g_3 , we get the minimizing optimum values to be

$$g_1^* = \frac{\mu_{21} \mu_{02}^2 (\beta_2 - 1) + \mu_{03} (\mu_{02} \mu_{20} - \mu_{22})}{\bar{Y} (\beta_2 - \beta_1 - 1) \mu_{02}^3} \quad \text{and} \quad (2.11)$$

$$g_3^* = -\frac{\mu_{21} \mu_{03} + \mu_{02} (\mu_{02} \mu_{20} - \mu_{22})}{\bar{Y} (\beta_2 - \beta_1 - 1) \mu_{02}^3} \quad (2.12)$$

The minimizing optimum values of g_1 and g_3 given in (2.11) and (2.12) which when substituted in (2.8) gives the minimum value of mean square error of the estimator d_g as

$$\begin{aligned} \text{MSE}(d_g)_{\min} = \text{mse}(s_y^2) - \left(\frac{1}{n} - \frac{1}{n'}\right) \frac{\mu_{21}^2}{\mu_{02}} \\ - \left(\frac{1}{n} - \frac{1}{n'}\right) \frac{\mu_{21}^2}{(\beta_2 - \beta_1 - 1) \mu_{02}^2} \left\{ \frac{\mu_{02} \mu_{20}}{\mu_{21}} - \frac{\mu_{22}}{\mu_{21}} + (\mu_{02})^{1/2} \gamma_1 \right\}^2 \quad (2.13) \end{aligned}$$

$$\text{where } \gamma_1 = \sqrt{\beta_1}, \quad \beta_1 = \frac{\mu_{03}^2}{\mu_{02}^3} \quad \text{and} \quad \beta_2 = \frac{\mu_{04}}{\mu_{02}^2}.$$

3. Comparison with the Estimators Available in the Literature

For comparing efficiency of the proposed generalized double sampling estimator, let us consider the following

- (i) **Usual Conventional unbiased Estimator of Population Variance in case of SRSWOR**

$$\hat{d}_1 = s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad \text{with } \text{MSE}(\hat{d}_1) = \frac{1}{n} (\mu_{40} - \mu_{20}^2) \quad (3.1)$$

from (3.1) and (2.14) it is clear that the proposed generalized double sampling estimator of population variance has mean square error lesser than the usual conventional unbiased estimator.

(ii) **Estimator of Population Variance given by Peeyush Misra and R. Karan Singh (2014)**

$$\hat{d}_2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y} \cdot f(\bar{y}, \bar{x}, \bar{x}')$$

$$\text{with } mse(\hat{d}_2) = \frac{1}{n} (\mu_{40} - \mu_{20}^2) - \left(\frac{1}{n} - \frac{1}{n'} \right) \frac{\mu_{21}^2}{\mu_{02}} \quad (3.2)$$

from (3.2) and (2.14) it is clear that the proposed generalized double sampling estimator of population variance has mean square error lesser than the mean square error of the estimator of variance given by Peeyush Misra and R. Karan Singh in 2014.

4. Empirical Study

For comparing efficiency of the proposed generalized class of estimator let us consider the data given in William G. Cochran (1977), Sampling Techniques, 3rd Edition, John Wiley and Sons, New York, dealing with Paralytic Polio Cases 'Placebo' (Y) group, Paralytic Polio Cases in not inoculated group (X), we have

$$n = 34$$

$$n' = 50 \text{ (say)}$$

$$\bar{Y} = 2.58$$

$$\bar{X} = 8370.6$$

$$\mu_{20} = 9.8894$$

$$\mu_{30} = 47.015235$$

$$\mu_{40} = 421.96088$$

$$\mu_{21} = 93.464705 \times 10^3$$

$$\mu_{11} = 19.34352945 \times 10^3$$

$$\mu_{02} = 7.1865882 \times 10^7$$

$$\mu_{03} = 1.4510955 \times 10^{12}$$

$$\mu_{04} = 4.5961952 \times 10^{16}$$

$$\mu_{12} = 3.443287 \times 10^8$$

$$\mu_{22} = 3.0156658 \times 10^9.$$

We have

$$mse(\hat{d}_1) = 9.534136695$$

$$mse(\hat{d}_2) = 8.390090538 \quad \text{and}$$

$$mse(d_g) = 7.1031288274.$$

Table 4.1: Percent Relative Efficiency (PRE) of the Proposed Estimator over the Estimators Described Above

PRE of the Proposed Estimator over the Estimators	PRE
PRE of the Proposed Estimator d_g over the Estimator \hat{d}_1	134.22
PRE of the Proposed Estimator d_g over the Estimator \hat{d}_2	118.11

5. Conclusion

The comparative study and empirical study of the proposed generalized double sampling estimator of population variance establishes its superiority in the sense of having minimum mean square error over some of the estimators available in the literature.

References

1. Cochran, W.G. (1977): Sampling Techniques, 3rd edition, John Wiley and Sons, New York.
2. Murthy, M. N. (1967): Sampling Theory and Methods, 1st edition, Statistical Publishing Society, Calcutta (India).
3. Peeyush Misra and R. Karan Singh (2014): Estimation of population variance using a generalized double sampling estimator, Sri Lankan Journal of Applied Statistics, Vol. 15, No. 3, 211- 220.
4. Singh, R. Karan (1982): Generalized double sampling estimators for the ratio and product of population parameters, Journal of Indian Statistical Association, 20, 39-40.
5. Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling theory of surveys with applications, 3rd revised edition, IOWA State University Press, Ames, U.S.A.

Converting weakness into the Strength: the Israel Story of Water

Ankit Saraswat

Abstract:

Three decades ago, Israel's water situation looked very critical. Water reserves were dangerously low. Rainfall was practically nonexistent. Fields and the farmers who managed them were both increasingly thirsty for fresh water. In the contrary, now, one would never know this country was once bone dry.

Water reservoirs are now full. Broad swaths of desert are green and lush with wheat, cotton and date fields. Amid acres of rolling sand, far from any freshwater source and under a beating, unforgiving sun, artificial lakes brim with fish. High-tech produce, including perfect cherry tomatoes, desert roses and plum-shaped, sugar-sweet bell peppers, burst from the earth and are swiftly collected for export.

Introduction & Objective

Israel, officially the State of Israel, is a country in West Asia, situated at the southeastern shore of the Mediterranean Sea and the northern shore of the Gulf of Aqaba in the Red Sea. It shares land borders with Lebanon to the north, Syria in the northeast, Jordan on the east, the Palestinian territories comprising the West Bank and Gaza Strip to the east and west, respectively, and Egypt to the southwest. It contains geographically diverse features within its relatively small area. Israel is well known to the world for transforming itself from a water threat country to a best water managing country. The current article is expected to throw some light on how the transformation was carried out by Israel.

The Israel Water Story:

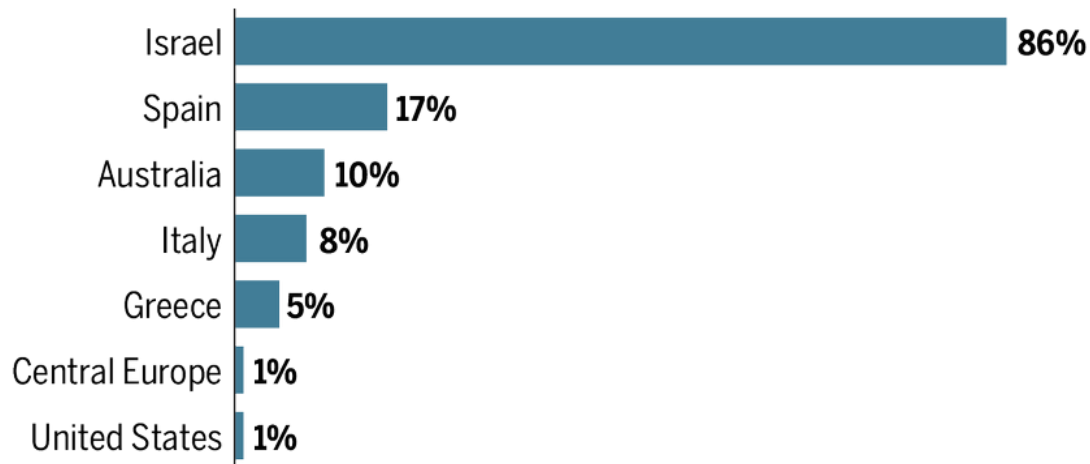
Israel has conquered its water problem. In one of the driest corners of the world, where rainfall sometimes amounts to less than 2 inches per year, there are sloshing swimming pools, manicured lawns and even an annual water fight. Studying Israel is important because Israel is probably the only place globally that managed, as a dry country, to truly solve its water issues.

Israel truly kicked its water problem in the 1980s, but water conservation and innovation efforts have been at the backbone of the State of Israel since before its establishment in 1948. Rain falls only in the winter there and in some parts of the country, there are years where it doesn't fall at all. Israel's national water authority, Mekorot, was created in 1937, well before the state itself was established. In its first 40 years, the company oversaw the completion of major nation-length pipelines that pumped water from the Jordan River and the Sea of Galilee down to the country's more-thirsty corners.

In Israel, so much water is recycled that the nation needs more reservoirs to hold it all. Stretches of desert have been turned green, thanks to salvaged water. Israel now purifies between 85 and 90 percent of its sewage water, more than anyone else in the world.

Effluent quantity for irrigation

Israel's use of effluent water — treated wastewater that is suitable for irrigation, but not for consumption — far exceeds the rest of the world.



Source: Israel Water Authority

Education is also crucial. Israelis are taught from a young age to treat water like gold. Schools have curriculum devoted to water conservation, with filtration systems and salvage stations in the schoolyard for hands-on science projects. At home, mothers watch over children brushing their teeth, not to make sure that they reach their back molars but that they keep the tap turned off while the brush is in their mouth. All toilets in Israel have a dual-flush system, with the option for a smaller, water-saving flush to be used for smaller jobs.

Drip irrigation — the Israeli-created farming method that waters crops through precise pinpoints in water sources, saving untold gallons of water is one of the key achievements. In recent years, however, as Israel has gained recognition as an incubator of startups, a slew of tiny, big-dreaming companies have joined the crowd to present technology to revolutionize the way water is handled.

New examples on the Israeli startup scene:

- Curapipe, which can discover microscopic leaks in municipal piping systems and patch them on the spot.
- RealiteQ, which peddles software to allow farmers to manage their water systems remotely.

- Aquanos, which has figured out how to use algae to treat wastewater and then harness the waste material for energy.

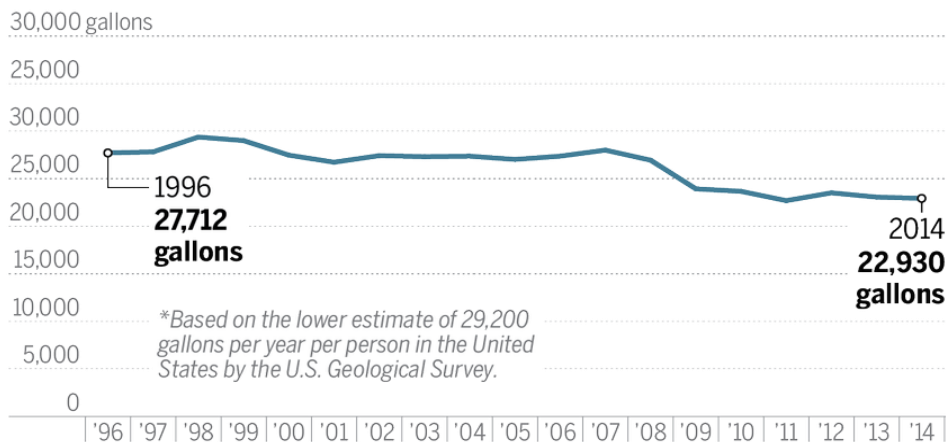
- Ayyeka, which offers cutting-edge water security.

A lot of the effort, however, is rooted in the fact that Israelis don't just treat water like gold, they also price it as such. Costs fluctuate, but an average Israeli can expect to pay about \$2 per cubic meter of water, meaning that a family that uses about 400 gallons of water a day will be charged about \$90 per month.

Domestic water consumption

From 1996 to 2014, Israel reduced its water consumption by 17 percent. It is estimated that on average, Israelis consumed nearly 23,000 gallons of water per capita in 2014 – 27 percent less than Americans.*

Israel's annual water consumption per capita:



Source: Israel Water Authority; U.S. Geological Survey

BETO ALVAREZ • U-T

Agriculture sits at the heart of Israel's economy, with the export of fresh fruits, vegetables and flowers, and the nation priding itself on its ability to meet nearly all of its food requirements from within. Some of the nation's top food exports, such as almonds and avocados, are major offenders when it comes to water consumption. To keep the fields lush and the desert in bloom, Israel has saved its fresh water for drinking and instead developed a sophisticated, multilayered system devoted to reclaiming sewage water for agriculture. One Israeli method for water conservation is the building of limans, a type of earthen construction in which desert valleys are flooded and the rainwater is then collected naturally. Israel now purifies between 85 and 90 percent of its sewage water, more than anyone else in the world. So much water is salvaged and recycled, while water for Israeli agriculture comes from the sewers, much of the water that flows from taps come from the sea.

Conclusion:

The quote, “Necessity is the mother of all invention” by Albert Einstein is most suitable to conclude the Israel Water Story. The efforts take by government and researchers are well supported by the people of the country, although the act was very much needed for the existence of the country itself. The education, awareness and the strict legislations all played important role. The concept seems as simple as a, b, c, Water Saved is Water Earned. This is the first commandment in success story of Israel. States of art water recycling techniques, measured irrigation technology, leak proof water transportation system are key policies in the transformation. If, we as a country can do 1% of what Israel has done, we won’t need to worry on the water conservation issue.

Data Sources:

- ✓ *Israel Water Authority*
- ✓ *Water Resource Institute*
- ✓ *U N Water Portal*

Managing CLOUD computing vulnerabilities with Managed services Technology

By Monisha

Cloud, Mobile and Web applications are growing in size and complexity. Despite their sophistication, these applications are designed to respond to simple HTTP requests. These requests can put applications and confidential information at risk as hackers can shield attacks with legal requests that pass through secured networks and intrusion detection systems. Once a malicious request interacts with a Cloud, Mobile or Web application, it can attack via vulnerabilities within the application and hence can become a threat to database.

Some of the top Cloud, Mobile and Web application vulnerabilities include:

- Cross Site Request Forgery
- Ineffective Session Termination
- Session ID Identification
- Application Path Disclosure
- Un-validated input
- Broken access control
- Broken authentication and session management
- Cross-site scripting (XSS) flaws
- Buffer overflows
- Injection flaws
- Improper error handling
- Insecure storage
- Insecure configuration management

Managed Service Technology for Cloud Security

A managed service solution that leverages a powerful technology platform allows companies to “jump start” their application security process without the overhead of installing software or hardware or dealing with implementation issues. This is particularly effective for companies, large and small with minimal in-house security expertise or resources. The managed service allows companies to have the vendor run the assessment for them, get the results in a professional report, and start working on remediation through their development process. This approach is far more cost effective than manual security assessment and penetration testing and companies can eventually transfer it back in-house once they have built the expertise.

Impact Analysis

The step in measuring level of risk is to determine the adverse impact resulting from a successful threat exercise of vulnerability.

The adverse impact of a security event can be described in terms of loss or degradation of any, or combination of any, of the following three security goals: integrity, availability, and confidentiality that can be describes qualitative categories as high, medium, low.

Participatory Behavioral Model Checking

Similarly, extension and revision is required for systems' model- checking properties. Model checking has been performed using the model of nondeterministic automata based on states determined by time-points. As an alternative to nondeterministic state machines, actor model theory provides the means to characterize all the possible computations of a concurrent system in terms of the Representation theorem. Recent improvements in paraconsistent logic provide the following advantages over previous versions. Almost all classical Boolean equivalences hold, including double negation elimination, excluded middle, and (for conjunction and disjunction) associatively.

Secure outsourcing of computation

Computation outsourcing is applied when an entity has to perform a task that requires computation power and resources that the entity cannot dispose and the delegation of this task to an external provider is the only option. Secure outsourcing refers to such an outsourcing in which security requirements, and sometimes privacy requirements can be met. The main challenge of a secure outsourcing is to delegate a computation to a set of service providers that are either distrusted or partially trusted without exposing and revealing either the input data or the computed output. Moreover, the client wants to be able to verify the correctness of the result in order to trust the new data that they possess.

The concept of computation outsourcing can be seen as the core of the cloud computing model. However, such a model of outsourcing, i.e. a distribution of computations under rather loose restrictions cannot be applied in all cases. Indeed, there is a plethora of applications for which there are strong security and privacy requirements and there are companies and research institutes that avoid at any cost a Cloud infrastructure due to security aspect. Several secure computation outsourcing techniques have been proposed and there are classic results for the problems of secure two-party and multi-party computation.

The function to be computed is first encrypted by an entity, called the constructor, with symmetric cryptography. Then, another party, called the evaluator, decrypts the function using the keys that correspond to the input data. The use of symmetric encryption algorithms endows efficiency in terms of implementation to the garbled circuits. However, this procedure is one-time-pad like. That means that the garbled circuits can be used only once and their size is

proportional to the size of the function to be computed. Several hardware implementations have been proposed to accelerate the procedure.

A hardware token which is tamper-proof against physical attacks but can perform arbitrary computations would enable the cloud remote client to perform confidential and verifiable computation on the cloud provider's site, given that the client trust the manufacturer of the token that it does not leak any information to the provider. We formulated some metric goals for each of the trusted cloud that identified objects.

The token based approach is reasonable because both, cryptographic coprocessors and standardized interfaces exist that can be used for such tokens. Of course, for trust reasons, the token vendor should not be the same as the cloud provider. However, the whole security-critical computation takes place in the token. Hence, such computation is not really high bandwidth channel outsourced to the cloud because the function is computed within the token. Some applications, however, require fast replies to queries which cannot be computed online within the tamper-proof token. For example, untrusted storage in personal health records or payroll databases may occur not very frequently, but need to be processed very fast while privacy of the data should be preserved.

Secure Data collection for identification of information security metrics

In this research aims at identifying SLA based information security metrics relevant in cloud computing. In this research investigates different Objects relevant in cloud policy models and each of them can be used as a component in the proposed framework. A role represents a specific function within an organization and can be seen as a set of actions or responsibilities associated with this function. In a Secure Data collection model, all grant authorizations deal with roles, rather than being granted directly to users. Users are then made members of Operations management, thereby acquiring the roles authorizations. User access to resources is controlled by roles; each user is authorized to play certain roles and, based on his own role he can perform accesses to the resources and operate them correspondingly.

Insider Access

Data processed or stored outside the confines of an organization, its firewall, and other security controls bring with it an inherent level of risk. The insider security threat is a well-known issue for most organizations and, despite the name, applies as well to outsourced cloud services. Insider threats go beyond those posed by current or former employees to include organizational affiliates, contractors, and other parties that have received access to an organization's networks, systems, and data to carry out or facilitate operations [36]. Incidents may involve various types of fraud, sabotage of information resources, and theft of information. Incidents may also be caused unintentionally. Moving data and applications to an external cloud computing environment expands the insider security risk not only to the service provider's staff, but also potentially among other customers using the service.

Server-Side Protection

Virtual servers and applications, much like their non-virtualized counterparts, need to be secured in IaaS clouds. Following organizational policies and procedures, hardening of the operating system and applications should occur to produce VM images for deployment. Care must also be taken to make adjustments for the virtualized environments in which the images run. For example, virtual firewalls can be used to isolate groups of VMs from other groups hosted, such as production systems from development systems or development systems from other cloud-resident systems. Carefully managing VM images is also important to avoid accidentally deploying images containing vulnerabilities.

Client-Side Protection

A successful defense against attacks requires both a secure client and a secure Website infrastructure. With emphasis typically placed on the latter, the former can be easily overlooked. Web browsers, a key element for many cloud computing services, and the various available plug-ins and extensions for them are notorious for their security problems. Moreover, many browser add-ons do not provide automatic updates, increasing the persistence of existing vulnerabilities.

Access Control

Besides authentication, the capability to adapt user privileges and maintain control over access to resources is also required, as part of identity management. Standards like the eXtensible Access Control Markup Language (XACML) can be employed to control access to cloud resources, instead of using a service provider's proprietary interface. XACML focuses on the mechanism for arriving at authorization decisions, which complements SAML's focus on the means for transferring authentication and authorization decisions between cooperating entities.

Conclusion

Data stored in the cloud typically resides in a shared environment collocated with data from other customers. Organizations moving sensitive and regulated data into the cloud, therefore, must account for the means by which access to the data is controlled and the data is kept secure.

Any progress must first occur in a particular domain in our work focuses on an important class of widely used applications that includes e-mail, personal financial management, social networks, and business tools such as word processors and spreadsheets. The following criteria define this class of applications:

Provide services to a large number of distinct end users, as opposed to bulk data processing or workflow management for a single entity;

Use a data model consisting mostly of sharable units, where all data objects have access control lists (ACLs) with one or more users; and

Developers could run the applications on a separate computing platform that encompasses the physical infrastructure, job scheduling, user authentication, and the base software environment, rather than implementing the platform themselves.

References

- K. D. Bowers, A. Juels, and A. Oprea, "Hail: A high-availability and integrity layer for cloud storage," in Proc. of CCS'09, 2009, pp. 187–198.
- T. Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in Proc. of ICDCS'06, 2006, pp. 12–12.
- M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A cooperative internet backup scheme," in Proc. of the 2003 USENIX Annual Technical Conference (General Track), 2003, pp. 29–41.
- M. Castro and B. Liskov, "Practical byzantine fault tolerance and proactive recovery," ACM Transaction on Computer Systems, vol. 20, no. 4, pp. 398–461, 2002.
- L. Carter and M. Wegman, "Universal hash functions," Journal of Computer and System Sciences, vol. 18, no. 2, pp. 143–154, 1979.
- J. Hendricks, G. Ganger, and M. Reiter, "Verifying distributed erasure-coded data," in Proc. of 26th ACM Symposium on Principles of Distributed Computing, 2007, pp. 139–146.
- J. S. Plank and Y. Ding, "Note: Correction to the 1997 tutorial on reed-solomon coding," University of Tennessee, Tech. Rep. CS-03-504, April 2003.
- C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for storage security in cloud computing," in Proc. of IEEE INFOCOM'10, San Diego, CA, USA, March 2010

Different techniques of Analysis of Data

By **ARUN PRATAP SINGH**

Research Scholar (MEWAR UNIVERSITY)

Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes. Business intelligence covers data analysis that relies heavily on aggregation, focusing on business information. In statistical applications, some people divide data analysis into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA focuses on discovering new features in the data and CDA on confirming or falsifying existing hypotheses. Predictive analytics focuses on application of statistical models for predictive forecasting or classification, while text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data. All are varieties of data analysis.

Process of Data Analysis

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.

Data analysis techniques allow researchers to review gathered data and make inferences or determination from the information. Most techniques focus on the application of quantitative techniques to review the data. A few of the more popular quantitative data analysis techniques include descriptive statistics, exploratory data analysis and confirmatory data analysis. The latter two involve the use of supporting or not supporting a predetermined hypothesis. Groups who may use these techniques include individual researchers, students, businesses, government agencies and actuaries, among other parties needing information and data.

Quantitative data analysis attempts to remove researcher bias from gathered data. Heavy use of statistics, probabilities or other mathematical techniques allows individuals to use standard

methods for interpreting data. When researchers attempt to use qualitative data analysis techniques — often based on the individual's personal background, preferences or basic research and reasoning principles — the gathered data may be misread or incorrectly interpreted. Therefore, mathematical techniques are less susceptible to these errors and often more accepted by other individuals or researchers.

Descriptive statistical analysis separate or summarize data into specific groups. Demographics are a common set of descriptive statistics. Researchers will gather information relating to a population's age, sex, household size, income, job type and other information. Another type of descriptive statistical analysis is the completion percentage of a quarterback when playing in a football game. If the quarterback completes six out of eight pass attempts, he has a 75 percent completion percentage. A flaw in this technique is the inability for the statistics to provide additional information, such as the length of each pass.

Exploratory data analysis techniques often involve the use of box plots, histograms, Pareto charts, scatter plots or a stem-and-leaf plot. The main purpose for this technique is to support a stated hypothesis from a researcher. For example, a researcher may desire to prove a hypothesis about the age range of owners who drive a specific type of car, such as a minivan. To test and support this hypothesis, the researcher will gather information and create a box plot to determine the number of owners within his stated range. The statistics will provide information to support or not support the hypothesis and show how many outliers are in the gathered data.

Confirmatory data analysis techniques are the opposite of exploratory techniques. In these tests, the researcher is looking to disprove the null hypothesis, which is a statement generally accepted as true by most individuals. To disprove the null hypothesis, the researcher will gather specific information related to the hypothesis and test the average, variance, p-values and confidence intervals. The confidence interval proven by the confirmatory data analysis will provide information as to how confident the researcher should be about whether the null hypothesis is true or false.

Data mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information -

information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

How does data mining work?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- **Classes:** Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- **Clusters:** Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- **Associations:** Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- **Sequential patterns:** Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.

- Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) . CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k > 1). Sometimes called the k-nearest neighbor technique.

Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

Technological infrastructure

Today, data mining applications are available on all size systems for mainframe, client/server, and PC platforms. System prices range from several thousand dollars for the smallest applications up to \$1 million a terabyte for the largest. Enterprise-wide applications generally range in size from 10 gigabytes to over 11 terabytes. NCR has the capacity to deliver applications exceeding 100 terabytes. There are two critical technological drivers:

Size of the database: the more data being processed and maintained, the more powerful the system required.

Query complexity: the more complex the queries and the greater the number of queries being processed, the more powerful the system required.

Relational database storage and management technology is adequate for many data mining applications less than 50 gigabytes. However, this infrastructure needs to be significantly enhanced to support larger applications. Some vendors have added extensive indexing capabilities to improve query performance. Others use new hardware architectures such as Massively Parallel Processors (MPP) to achieve order-of-magnitude improvements in query time. For example, MPP systems from NCR link hundreds of high-speed Pentium processors to achieve performance levels exceeding those of the largest supercomputers.

Conclusion

One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and glean a significant amount of information about individuals buying habits and preferences.

Another issue is that of data integrity. Clearly, data analysis can only be as good as the data that is being analyzed. A key implementation challenge is integrating conflicting or redundant data from different sources. For example, a bank may maintain credit cards accounts on several different databases. The addresses (or even the names) of a single cardholder may be different in each. Software must translate data from one system to another and select the address most recently entered.

A hotly debated technical issue is whether it is better to set up a relational database structure or a multidimensional one. In a relational structure, data is stored in tables, permitting ad hoc queries. In a multidimensional structure, on the other hand, sets of cubes are arranged in arrays, with subsets created according to category. While multidimensional structures facilitate multidimensional data mining, relational

structures thus far have performed better in client/server environments. And, with the explosion of the Internet, the world is becoming one big client/server environment.

Finally, there is the issue of cost. While system hardware costs have dropped dramatically within the past five years, data mining and data warehousing tend to be self-reinforcing. The more powerful the data mining queries, the greater the utility of the information being gleaned from the data, and the greater the pressure to increase the amount of data being collected and maintained, which increases the pressure for faster, more powerful data mining queries. This increases pressure for larger, faster systems, which are more expensive.

References

1. *"Data Mining Curriculum"*. ACM SIGKDD. 2006-04-30. Retrieved 2014-01-27.
2. Clifton, Christopher (2010). *"Encyclopædia Britannica: Definition of Data Mining"*. Retrieved 2010-12-09.
3. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). *"The Elements of Statistical Learning: Data Mining, Inference, and Prediction"*. Retrieved 2012-08-07.
4. ^ Jump up to:^a ^b ^c Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). *"From Data Mining to Knowledge Discovery in Databases"* (PDF). Retrieved 17 December 2008.
5. Han, Jiawei; Kamber, Micheline (2001). *Data mining: concepts and techniques*. Morgan Kaufmann. p. 5. ISBN 978-1-55860-489-6. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long
6. See e.g. OKAIRP 2005 Fall Conference, Arizona State University About.com: Datamining
7. Witten, Ian H.; Frank, Eibe; Hall, Mark A. (30 January 2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.). Elsevier. ISBN 978-0-12-374856-0.
8. Bouckaert, Remco R.; Frank, Eibe; Hall, Mark A.; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H. (2010). *"WEKA Experiences with a*

Java open-source project". Journal of Machine Learning Research 11: 2533–2541. the original title, "Practical machine learning", was changed ... The term "data mining" was [added] primarily for marketing reasons.

9. Mena, Jesús (2011). *Machine Learning Forensics for Law Enforcement, Security, and Intelligence*. Boca Raton, FL: CRC Press (Taylor & Francis Group). ISBN 978-1-4398-6069-4.
10. Piatetsky-Shapiro, Gregory; Parker, Gary (2011). "Lesson: Data Mining, and Knowledge Discovery: An Introduction". *Introduction to Data Mining*. KD Nuggets. Retrieved 30 August 2012.